



The Past and the Future of Mortality

Dissertation zur Erlangung des akademischen Grades eines Doktors der Wirtschaftswissenschaften
(Dr. rer. pol.) an der Fakultät für Mathematik und Wirtschaftswissenschaften der Universität Ulm

Vorgelegt von:

Martin Genz

Amtierender Dekan:

Professor Dr. Martin Müller

Vorsitzende des Promotionsausschusses:

Professor Dr. Sandra Ludwig

Gutachter:

Professor Dr. An Chen

apl. Professor Dr. Hans-Joachim Zwiesler

Tag der Promotion:

19. Juli 2019

Acknowledgments

I would like to thank everybody who supported me in my research and when writing my thesis.

First of all, I would like to offer my special thanks to my supervisor Prof. Dr. An Chen for her support and invaluable discussions and suggestions.

Furthermore, I would like to thank apl. Prof. Dr. Hans-Joachim Zwiesler for co-supervising, reviewing this thesis, and supporting with help and advice.

I am particularly grateful for the fruitful collaboration with my coauthors apl. Prof. Dr. Jochen Ruß, Dr. Matthias Börger, Jan Feifel, and Prof. Dr. Markus Pauly. Without your support this thesis would not have been possible.

I would also like to express my very great appreciation to the Institut für Finanz- und Aktuarwissenschaften (ifa) for giving me the opportunity to gain practical experience during my studies and to all my colleagues at the ifa for the great atmosphere. It is a pleasure for me being a part of this team. Also everybody at the Institut für Versicherungswissenschaften (IVW) at Ulm University deserves my thanks not least for valuable comments and discussions during our research workshops. I also thank the IVW for funding my attendance at many conferences.

I also appreciate the helpful suggestions and comments of numerous unnamed referees, discussants and participants at several conferences.

Last but not least, I am eternally grateful to my family, in the first place to my wife and my children, but also to my fast friends, for being there for me. They endowed me with their endless support and encouraged me to do this.

Ulm, Juli 2019

Martin Genz

Contents

Overview of Research Papers	vii
Co-Authorship	ix
Research Context and Summary of Research Papers	1
1 Field of Research	1
2 Motivation and Objectives	3
3 Summary of Research Papers	5
 Research Papers	
1 Extension, Compression, and Beyond - A Unique Classification System for Mortality Evolution Patterns	21
2 A Comprehensive Analysis of the Patterns of Worldwide Mortality Evolution	45
3 The Myth of Immortality: An Analysis of the Maximum Lifespan of US Females	75
4 The Future of Mortality: Mortality Forecasting by Extrapolation of Deaths Curve Evolution Patterns	99
Curriculum Vitae	125

Overview of Research Papers

Research papers included in this dissertation

1. Börger, M., Genz, M., and Ruß, J. (2018). Extension, Compression, and Beyond – A Unique Classification System for Mortality Evolution Patterns. *Demography*, 55(4): 1343-1361. DOI: 10.1007/s13524-018-0694-3.
2. Genz, M. (2017). A Comprehensive Analysis of the Patterns of Worldwide Mortality Evolution. *2017 Living to 100 Society of Actuaries International Symposium Monograph*. URL: <https://www.soa.org/essays-monographs/2017-living-to-100/2017-living-100-monograph-genz-paper.pdf>.
3. Feifel, J., Genz, M., and Pauly, M. (2018). The Myth of Immortality: An Analysis of the Maximum Lifespan of US Females. Working Paper.
4. Börger, M., Genz, M., and Ruß, J. (2019). The Future of Mortality: Mortality Forecasting by Extrapolation of Deaths Curve Evolution Patterns. Working Paper.

Co-Authorship

Matthias Börger

Matthias Börger is a consultant at the Institut für Finanz- und Aktuarwissenschaften (ifa). He received his doctoral degree from Ulm University in 2013.

Jan Feifel

Jan Feifel is a research fellow at the Institute of Statistics at the University of Ulm. He received his Master of Science degree from Ulm University in 2017.

Markus Pauly

Markus Pauly is a professor at TU Dortmund and formerly was a professor at Ulm University. He received his doctoral degree from Düsseldorf University in 2008 and his post-doctoral lecture qualification (Habilitation) from Düsseldorf University in 2013.

Jochen Ruß

Jochen Ruß is a managing partner at the Institut für Finanz- und Aktuarwissenschaften (ifa) and an adjunct professor at Ulm University. He received his doctoral degree from Ulm University in 1999 and his post-doctoral lecture qualification (Habilitation) from Ulm University in 2009.

Research Context and Summary of Research Papers

1 Field of Research

The explanation of realized mortality as well as the prediction of future mortality are key challenges in many different fields such as, for example, demography or actuarial sciences. This cumulative thesis contributes to the literature about these questions.

The distribution of human lifetimes has changed significantly during the last centuries. One indication of this is the steadily increasing mean of this distribution, i.e. the life expectancy at birth (see, e.g., Oeppen and Vaupel, 2002, who observe a linearly increasing trend in the world record life expectancy). However, the distribution of human lifetimes has evolved in a complex manner and it cannot be characterized by its mean only. Therefore a variety of different statistics has been proposed to measure these changes. For example, Kannisto (2001) suggests the modal length of life¹ as an alternative measure for the average human lifetime. Other statistics focus on the dispersion of the distribution of human lifetimes. For example, Wilmoth and Horiuchi (1999) give an overview of some of them. With the application of these new statistics also new terms – so-called *mortality scenarios* – have been created which are designed to describe different evolution patterns in the distribution of human lifetimes, for example “rectangularization of the survival curve” (Manton and Tolley, 1991), “compression of mortality” (Wilmoth, 2000), or “longevity extension” (Cheung et al., 2005). Most of the essays in the demographic literature on the evolution of the distribution of human lifetimes cover questions on the development of statistics to determine the presence or absence of particular mortality scenarios (see, e.g., Cheung et al., 2005; Kannisto, 2000, 2001; Wilmoth and Horiuchi, 1999) and on the application of existing frameworks to investigate realized mortality (see, e.g., Canudas-Romo, 2008; Cheung and Robine, 2007; Nusselder and Mackenbach, 1996; Ouellette and Bourbeau, 2011; Robine et al., 2008).

In addition to the aforementioned literature there is a stream of not necessarily demographic literature on the development of *mortality models*. Such mortality models are a valuable tool to

¹The modal length of life is defined as the age where the density function of the distribution of human lifetimes reaches its maximum.

describe the distribution of human lifetimes and the changes exhibited over time. One of the oldest mortality models of this kind is the “law of mortality” by Gompertz (1825). This model has been modified and extended several times. Pollard (1987) gives a comprehensive review of mortality models which were developed up until the 1980s. Most of these mortality models are also used to forecast future mortality. During the most recent decades a new class of mortality models has emerged – the so-called stochastic mortality models. For example, the mortality models by Lee and Carter (1992) or Cairns et al. (2006) are widely used in many applications today. Overviews of more recent mortality models are given by Booth and Tickle (2008) or Cairns et al. (2008).

The evolution of the distribution of human lifetimes in the highest age range is of special interest to the demographic literature. Essentially two different and contradictory cases for how the shape of the distribution of lifetimes evolves in the highest age range have been made: Either it has a fixed right endpoint (see, e.g., Dong et al., 2016; Fries, 1980; Gavrilov and Gavrilova, 2011), which implies that there is a maximum attainable age. Or the probability of death flattens in the highest age range (see, e.g., Barbi et al., 2018), which implies that yet observed record ages at death can be outperformed in theory.² Often such claims are based on the application of mortality models to mortality data for the highest age range. However, the question on the existence of a finite limit to human lifespan is often already implied by the model’s design.³ Instead of using more or less sophisticated mortality models, more advanced methodical approaches are needed to decide on this question and this is a typical question of extreme value theory. One of the first attempts to employ methods from extreme value theory to estimate the maximum human lifespan can be found in Aarsen and de Haan (1994). More recently similar methods have been applied to data of many different countries (see, e.g., Gbari et al., 2017; Hanayama and Sibuya, 2015; Rootzén and Zholud, 2017; Weon and Je, 2009).

The research papers included in this thesis contribute to the demographic literature on the classification of changes in the distribution of lifetimes by developing a unique classification framework for mortality evolution patterns, provide new methodical approaches for the estimation of the maximum possible human lifespan, and combine a demographic understanding of realized mortality evolution patterns with a new mortality model, which can be used to forecast future mortality. In this sense this thesis links the past with the future of mortality.

²If a maximum attainable age exists, it might increase over time. This sometimes makes the differentiation between these two concepts difficult in practice. For a given population in a given year, however, the existence of a maximum attainable age implies a probability of zero for outliving this age. In contrast, if this age does not exist (which is the second case), the probability for survival to any (even infinitely high) age might be very small but always is greater than zero.

³For example, in every application of the Gompertz law of mortality the force of mortality reaches the value of one at some age. Thus, this model does not even hypothetically allow for an infinite human lifespan.

2 Motivation and Objectives

Assumptions about future mortality are necessary for example for social security systems, pension funds, and life insurers and it is not sufficient to know people's future average lifetime, i.e. forecasts of the life expectancy, only. To have an option on how much uncertainty is involved in their planning and reserving an estimate of the future distribution of human lifetimes is of particular interest. Usually mortality models are used to project future mortality, but the choice and the calibration of such models is challenging in practice. To meet this challenge the realized changes of the distribution of lifetimes should be analyzed. Only if we understand these past trends, assumptions about future mortality usually can be derived.

There is a wide range of literature which covers the question of how realized mortality has changed over time and how these changes can be classified. As mentioned above, many different mortality scenarios have been developed to explain these patterns. Most of them can be easily illustrated by changes in the deaths curve or the survival curve.⁴ However, questions related to the definition and differentiation of such mortality scenarios have hardly been discussed. Hence often mortality scenarios are imprecise or only vaguely defined, which makes it difficult to test the presence or the absence of a particular mortality scenario in practice. To this end, often single demographic statistics are used. For example, Wilmoth and Horiuchi (1999) suggest the inter-quartile range (*IQR*)⁵ as a single measure for “compression of mortality”. However, the evolution of the deaths curve is too complex for a single statistic being sufficient to characterize it. Therefore, such approaches are debatable. The matter is further complicated as the mortality scenarios are not mutually exclusive, as is sometimes assumed in the literature. For example, Canudas-Romo (2008) claims that the presence of “shifting mortality” rules out the presence of “compression of mortality” and thus does not allow for the coexistence of two or more mortality scenarios – although this is what can be observed in practice. Last but not least, effects of the chosen age range frequently are neglected, but it is possible to observe different mortality scenarios in different age ranges (see, e.g., Myers and Manton, 1984). For example, pension funds and life insurers should know the mortality scenario of their retirees and they might get a wrong picture if they analyze the trends in the deaths curve's evolution on the complete age range. These issues lead to the first research question:

Research Question 1: How can we describe and measure the changes the deaths curve exhibits over time and how can these changes be classified? How can we define mortality scenarios, so that this definition is unique, intuitive but precise, captures any observed mortality evolution, allows for the coexistence of mortality scenarios, and is

⁴The deaths curve gives the number of deaths per age while the survival curve gives the number of survivors for each age. Thus, these two curves directly correspond to each other.

⁵The inter-quartile range is defined as the distance between the 75% and the 25% quantile of the distribution of lifetimes.

applicable to different age ranges? How can we estimate the presence or absence of a particular mortality scenario?

Once a clear definition of mortality scenarios is made, a further, natural question to ask is what mortality scenarios have prevailed in the history in different countries, sexes, and age ranges, and how these scenarios can be compared. Such questions for example are relevant whenever (sub-)populations are jointly modeled for the purposes of the unisex calculation of life insurers or for the application of multi-population mortality models, but they are hardly covered in literature. This leads to the second research question:

Research Question 2: Which mortality scenario prevails in which region of the world and what are the differences in the mortality evolutions between males and females or between different age ranges? How can detected trends in the deaths curve's evolution be compared between different (sub-)populations?

The question how to estimate demographic statistics in practice is not always clear and sometimes preparation of the input data is needed.⁶ In particular, the estimation of demographic statistics which capture the right tail of the deaths curve is challenging because in this extreme age range mortality data is typically sparse and often censored. Thus, for example, the estimation of the right endpoint of the deaths curve's support is difficult. Often this statistic is estimated using mortality models (see, e.g., Barbi et al., 2018; Gavrilov and Gavrilova, 2011) but as mentioned before this approach is debatable as it depends on the design of the respective mortality model. The question of how the right endpoint of the distribution of human lifetimes can be estimated in the presence of sparse and censored mortality data is a typical question of extreme value theory. If, however, the underlying mortality data is right censored, methods of classical extreme value theory do not yield reasonable estimates. Moreover, if the data is (too) sparse, the sample size might become too small to obtain statistically significant results. However, in the field of censored extreme value theory there are methods which allow estimating the right endpoint of a distribution's support in presence of random right censoring in principle (see Einmahl et al., 2008) but the underlying data has to be chosen carefully: For example, Dong et al. (2016) found "evidence to the limit of human lifespan", but they only used death counts for their analysis. This completely ignores people who are still alive – and are probably even older than the oldest dying person. This leads to the third research question:

Research Question 3: In the presence of sparse and censored old-age mortality data, how can we determine the existence of a limit to human lifespan without application of mortality models with predefined assumptions on the existence of a maximum attainable age? In case of existence, how can it be estimated? How can we select proper mortality data to examine these questions?

⁶For example, Ouellette and Bourbeau (2011) smooth their input data prior to the estimation of demographic statistics, which in practice often is necessary.

Given the mortality scenario and the evolution of the corresponding demographic statistics of a particular population, assumptions on future mortality can be derived, for example by extrapolating the most recent trends in the respective statistics. However, this only yields some characteristics of a potential future deaths curve, whereas in most practical applications, forecasts of complete deaths curves are needed. In this regard it is desirable that these forecasts smoothly continue recently observed trends in demographic statistics and the deaths curve's changes in particular. For most mortality models this is not true as their parameters usually lack any demographic interpretation. Hence such forecasts should be based on an extrapolation of statistics with a demographic interpretation. Furthermore, this would allow for including expert's opinions to the forecasts of future mortality. This leads to the fourth research question:

Research Question 4: How can we make use of an understanding of recent trends in the deaths curve's changes when forecasting future mortality?

The four research papers included in this thesis examine the aforementioned research questions. In the next section we provide a detailed summary of these papers.

3 Summary of Research Papers

Research Paper 1: Extension, Compression, and Beyond – A Unique Classification System for Mortality Evolution Patterns

In the first paper we discuss the definition of mortality scenarios and their detection on observed mortality data. We commence with a review of previous approaches for the classification of changes in the distribution of human lifetimes and illustrate important shortcomings inherent in these approaches. Furthermore, we develop a classification framework for mortality evolution patterns which is based on the deaths curve and consists of four characteristics of this curve. Questions related to the practical implementation of the framework are discussed as well as an example application. This article is joint work with Matthias Börger and Jochen Ruß and has been published in *Demography*. It has been presented at the 2nd Annual International Conference on Demography and Population in Athens, Greece (2015), the 11th International Longevity Risk and Capital Market Solutions Conference in Lyon, France (2015), the IAALS Colloquium in Barcelona, Spain (2017), and at the 31st International Congress of Actuaries (ICA) in Berlin, Germany (2018). Furthermore, the paper has been awarded the IAALS Best Paper Award at the ICA in 2018.

As described above, there is a variety of mostly demographic literature on the question of how the distribution of lifetimes changes over time and how these changes can be classified. We

discuss four major shortcomings of these approaches: First of all, most mortality scenarios are defined imprecisely. An obvious example of this is *rectangularization* – a mortality scenario which is defined by its theoretical (and hence unreachable) final state in which the survival curve is perfectly rectangular. The survival curve can converge to this final state on different routes, which makes this scenario definition ambiguous. Secondly, some statistics used to detect mortality scenarios are misleading or insufficient. In particular, a single statistic can never indicate every change of the deaths curve. Thus (probably significant) changes of the deaths curve might remain undetected, if only one statistic is used. The same holds for the combination of several statistics which take into account only parts of the age range, e.g. old-age mortality. Thirdly, mortality scenarios can coexist and they are not mutually exclusive. The evolution of the deaths curve typically is not only driven by a single process but there are several different and overlapping processes which cause the observed changes. For example, the deaths curve can exhibit right-shifting mortality and compression at the same time: Right-shifting mortality can intuitively be interpreted as a horizontal shift of the complete deaths curve to the right. On the other hand, compression can intuitively be interpreted as a vertical deformation of the deaths curve: There must be an age range in which more and more people are dying whenever compression prevails. Hence a coexistence of these two concepts can intuitively be interpreted as a “diagonal deformation” of the deaths curve. This can be observed in practice.⁷ Finally, the effect which the choice of the age range has on the estimation of certain statistics – and thus on the assessment of mortality scenarios – often is ignored. It is possible to observe different mortality scenarios on different age ranges for the same population in one time period. For example, if the number of people dying in the younger age range decreases over time, this number of people must die in the older age range. Consequently, we observe compression toward higher ages. However, the probability of death in the higher age range might have stayed constant over time for all respective ages. Thus we cannot observe compression within the higher age range in this example and we have two different mortality scenarios in two different age ranges. In the paper all these shortcomings are illustrated with graphical examples to make the respective points clearer.

We develop a classification framework for mortality evolution patterns which uniquely assigns a mortality scenario to any change of the deaths curve. This framework combines four concepts for the change of the deaths curve over time: “shifting mortality” (see Canudas-Romo, 2008), “longevity extension” (see Rossi et al., 2013), “compression of mortality” (see Myers and Manton, 1984), and “concentration of mortality” (see Kannisto, 2001). These concepts are considered simultaneously so that in this framework a mortality scenario is always a four-dimensional vector. We use one demographic statistic for each component of this vector to identify changes of the deaths curve in the respective dimension:

⁷This example also illustrates why a single statistic often is not sufficient to detect this kind of evolution of the deaths curve: Most demographic statistics either focus on horizontal or vertical deformations. Hence such “diagonal” deformations usually cannot be captured by a single statistic.

- Shifting mortality is linked to the age, where the largest portion of the population under consideration dies, i.e. the modal age at death M . A significant increase in M is called *right-shifting mortality*, while a significant decrease in this statistic is called *left-shifting mortality*.
- Extension is linked to the length of the deaths curve's support measured by its upper bound which we call UB . A significant increase in UB is called *extension*, while a significant decrease in this statistic is called *contraction*.
- Compression is linked to the distance between the observed deaths curve and a perfectly decompressed deaths curve. The latter corresponds to a uniform distribution. The statistic we use for this measures the size of the area between these two curves. This is what we call the degree of inequality DoI . A significant increase in DoI is called *compression*, while a significant decrease in this statistic is called *decompression*.
- Concentration is linked to the record number of deaths in a single age, i.e. to the number of deaths in the modal age at death $d(M)$. A significant increase in $d(M)$ is called *concentration*, while a significant decrease in this statistic is called *diffusion*.

Of course each of these statistics can remain unaltered over time. In this case the respective component is referred to as *neutral*.

This classification framework eliminates the aforementioned shortcomings of previous approaches: Each mortality scenario is defined precisely and uniquely. The used statistics are easily interpretable and intuitive. Any significant change of the deaths curve is detected by at least one of the four statistics. Owing to the design of the framework a mixed scenario prevails whenever at least two components exhibit a significant change. Not least, the classification framework is applicable to any age range which includes the deaths curve's right tail.

For an application of the classification framework we suggest estimators for each of the four statistics. This yields a time series for each statistic, but these time series are typically noisy and have extreme outliers. Therefore we need methods to determine trends and trend changes within these time series. To this end, we first eliminate the outliers from each time series. After that we fit piecewise linear trends with or without jumps to the time series.⁸ For each part of the time series with constant linear trend we use a set of statistical tests to determine if the trend is significantly decreasing, increasing, or constant. This provides a four-dimensional mortality scenario for each calendar year.

In the final section of the paper we exemplarily apply the classification framework to mortality data for US females from 1933 to 2013. By means of this application we show that the framework eliminates the aforementioned shortcomings of previous approaches.

⁸Such jumps can be caused for example by changes in data-processing methods or extreme historical events.

In summary, in this paper we analyze existing definitions for mortality scenarios as well as previous approaches for the assessment of these scenarios and discuss their shortcomings; based on that we develop a new classification framework for mortality evolution patterns which precisely defines mortality scenarios and uniquely assigns a mortality scenario to any significant change of the deaths curve. Hence this paper answers the first part of the research questions.

Research Paper 2: A Comprehensive Analysis of the Patterns of Worldwide Mortality Evolution

In the second paper we apply the classification framework from the first paper to a comprehensive data set with mortality data from different countries, sexes, and age ranges, and go on to compare the obtained mortality scenarios. This paper has been accepted for a presentation at the 2017 Living to 100 Society of Actuaries International Symposium, Orlando, FL (USA). Subsequently the paper has been published online in the *2017 Living to 100 Monograph*.

The choice of the underlying mortality data is crucial for an application of a classification framework for mortality evolution patterns. This matter is particularly complicated when the study focuses on the comparison between different populations, because then mortality data of similar quality and quantity is needed for each population. The Human Mortality Database (HMD, 2015) provides standardized mortality data of high quality for many different countries all over the world. In this study we use sex-specific mortality data (i.e. death counts and exposure-to-risk) from the HMD of 34 different countries.⁹ As this data is right censored we first determine the force of mortality from the death counts and exposure-to-risk data and extrapolate it to the highest age ranges for every population and period using a P-Spline approach.¹⁰

For each population we determine a time series of deaths curves for females and males and for the age ranges starting at age 0 and 60, respectively. Moreover, we estimate the four statistics of the classification framework for each deaths curve and determine the respective directions of their trends (i.e. piecewise increasing, decreasing, or constant). This yields 136 time series of four-dimensional mortality scenarios and accordingly 544 time series of one dimensional statistics' evolutions. To compare these time series we proceed as follows: The directions of the trends of each time series are illustrated with a colored time bar at a glance. For each statistic we arrange these time bars according to regional clusters to better account for geographical proximity in the comparison of the trends. For the sake of clarity, the mortality evolution

⁹When we accessed the data the HMD offered mortality data for 42 different populations. However, we exclude populations where the HMD alerts the user to the lower quality of the input data, as well as populations with less than 40 years of data history. This leads to 34 different countries.

¹⁰For some countries the HMD also offers uncensored input data. However, our study has a focus on the comparison of different countries and thus we need structurally homogeneous mortality data for all countries. This is not true for the (partially uncensored) input data but for the edited data. The latter, however, is censored.

patterns are first analyzed for one combination of sex and age range as a reference, and thereafter we compare them to the mortality evolution patterns of the other combinations of sex and age range. Moreover, we develop a statistic to determine the degree to which the trends of two respective time series of mortality scenarios coincide, which we call the relative similarity RS . This statistic can be scaled which allows to determine the relative similarity of the time series of complete four-dimensional mortality scenarios between populations, sexes, and age ranges.

For the discussion of the results, first of all we focus on the mortality evolution of males and the complete age range. We find increasing trends in all four statistics in almost all populations for the most recent decades. These trends are sometimes interrupted by considerable supra-regional patterns, for example, a plateau in the almost global trend of right-shifting mortality in the 1960s. This can also be detected with respect to compression. The mortality scenarios for males from the Eastern European populations and the rest of the world considerably differ in all four statistics. By comparing the evolution of mortality scenarios between males and females we find significant differences between sexes. In particular with respect to right- and left-shifting mortality these differences become evident, whereas with respect to compression or decompression they are immaterial. The disparity between sexes is higher for some Eastern European populations (with few exceptions like Estonia and Bulgaria) and lower for mainly Central and Western European populations (with few exceptions like Austria). The same essentially holds for the difference between the chosen age ranges but these differences are more distinct for males than for females.

In summary, in this paper we analyze the evolution of the mortality scenarios of many different populations almost all over the world and highlight differences in the evolutions of female and male mortality and two different age ranges, respectively. In addition we address questions on how to compare time series of four-dimensional mortality scenarios. Hence this paper answers the second part of the research questions.

Research Paper 3: The Myth of Immortality: An Analysis of the Maximum Lifespan of US Females

In the third paper we discuss the estimation of the upper bound of the deaths curve's support, which is one of the statistics of the previously described classification framework. When faced with sparse and censored mortality data in the highest age ranges, special methods are required from the extreme value theory which we describe in this paper. Moreover, by combination of two data sets we can exploit the respective advantages of each data set to improve the obtained estimates. This paper is joint work with Jan Feifel and Markus Pauly.

Old-age mortality data is typically sparse and often censored. We describe two examples for this claim: On the one hand, each entry of the *International Database on Longevity* (IDL, 2015) relates to one individual dying in the age range beyond 110. For each entry the IDL gives the exact date of birth and the exact age at death (in days) and each entry is validated separately. However, for example for US females the IDL contains 309 entries only. Even though it is regarded to fit the profile of the female US population well, the absolute number of death counts in the IDL probably underestimates the absolute number of US females dying in this age range (see Maier et al., 2010). On the other hand, each entry of the *Human Mortality Database* (HMD, 2015) gives the number of people dying or being alive per age and calendar year. The HMD is regarded to be complete, i.e. it contains mortality and survival data of complete populations. Unfortunately, for the highest age range the HMD uses extrapolation techniques and the single-age data is given only until the age 109. Thus, the HMD is censored and possibly biased in this age range. In summary, we have two databases each with an advantage and a disadvantage: The IDL data is sparse but exact and thus of high quality; the HMD data is right-censored but comprehensive and thus of high quantity.

In the paper we apply different methods from the extreme value theory (EVT) to the data for US females from each of the aforementioned databases separately. This means that first we apply *classical* EVT methods to the uncensored but sparse IDL data for US females in order to determine the existence and – in case of existence – to estimate the value of the upper bound of the deaths curve’s support. As the IDL data for US females has a small sample size, the results of this analysis are not statistically significant: At a confidence level of 5% we can neither prove nor disprove the existence of a finite limit to the lifespan of US females. Furthermore, we apply methods of the *censored* EVT to the comprehensive but censored HMD data for US females to estimate UB .¹¹ In this case we find statistically significant evidence for the existence of a finite limit to the lifespan for US females, but the estimates are very close to the age 110 where the HMD data is censored. This indicates that additional information on deaths beyond the age 110 would improve the estimates of the right endpoint.

This additional information can be gathered by including the information of the IDL into the data set of the HMD. This is possible since we assume that the population recorded in the IDL is a sub-population of the population recorded in the HMD, at least for US females. In the paper we describe in detail how we proceed in combining these two data sets while avoiding double counts in the combined data set (CDS). Consequently, the CDS contains deterministic death counts at the age range beyond 110 and by including survival data given by the HMD it is randomly right censored. Therefore, methods of the censored EVT (e.g. the methods described

¹¹As the HMD also provides survival data, we can construct a randomly right censored database from the HMD alone. This is based on the fact that for each survivor to a particular age we know his minimum age at death (which is the age she survived to) but do not know the exact age at death. Therefore methods of the censored EVT are applicable in principle. In the part of the paper where we apply these methods to the combined data set, we discuss the applicability of this methods in more detail.

by Einmahl et al., 2008) become applicable in principle. In the paper we describe these methods in detail. One important condition for the application of the methods of Einmahl et al. (2008) is that the observations in the sample are independently and identically distributed (iid). As the CDS contains mortality data of more than two decades (from 1980 to 2003) and we know that the deaths curve typically changes over time this assumption is debatable. However, the proportion of the IDL data in each calendar year would be too small if we dissected the CDS into single calendar years. We found a compromise by dissecting the CDS in overlapping time windows of nine adjacent calendar years, i.e. from 1980–1989, from 1981–1990, and so on. This yields the possibility to estimate the right endpoint of the deaths curve’s support on a sufficiently large partition of the CDS (with respect to IDL data) and to check the resulting estimates for trends. However, the sample size of the HMD data is very large and so are the sample sizes of each window of the CDS. This would cause extreme computing times in the practical implementation and therefore we follow a sub-sampling approach to improve the performance of the estimation. An additional cross-validation procedure which we apply to the sub-samples further stabilizes the estimators we use.

We determine the existence and – in case of existence – estimate the value of the right endpoint of the deaths curve’s support on 16 overlapping nine-year moving windows for two different estimators. On a significance level of 5% we obtain the existence of a finite limit to the maximum possible lifespan of US females for each time window. The estimates for this limit are always greater than 120 years and the corresponding confidence intervals range between 117 and 131. A trend in the evolution of these estimates cannot be observed. The figures obtained are considerably higher than those we obtain by an exclusive analysis of the HMD data and thus also more far away from the censoring age of the HMD. This underlines that the construction of the CDS significantly improves our results. Furthermore, in the paper we discuss the obtained results with respect to other results from the literature though the direct comparison of the results is difficult as different methods and/or different sources of data are used. It turns out that our results are consistent to other results from literature.

To summarize, in this paper we describe a method to estimate the limit to human lifespan for a particular population which includes the assessment of its existence in the presence of sparse and censored mortality data; this method does not need any predefined assumptions of mortality models and is strictly based on the information given by the mortality data. Hence this paper answers the third part of the research questions.

Research Paper 4: The Future of Mortality: Mortality Forecasting by Extrapolation of Deaths Curve Evolution Patterns

In the fourth and final paper in this thesis we examine how the findings on the past evolution of the deaths curve can be transferred to make forecasts on future mortality. To this end we develop a new best estimate mortality model which is based on an extrapolation of the four statistics of the classification framework we presented in the first paper. We introduce the model theoretically, discuss questions on its practical implementation, and we demonstrate its benefits with several example applications. This paper is joint work with Matthias Börger and Jochen Ruß.

In scientific literature, various different mortality models are discussed. Many of these models are not only used to describe realized mortality but also to make forecasts on future mortality. However, the parameters of most of these models lack a clear demographic interpretation. Thus extrapolating these parameters to the future might lead to forecasts which are not plausible in a demographic sense. In the paper we provide an example for this claim: Using the mortality models by Lee and Carter (1992) and Cairns et al. (2006) we forecast deaths curves until the year 2070. We estimate the four statistics M , UB , DoI , and $d(M)$ on the observed deaths curves, where we calibrated the mortality models, as well as on the forecast deaths curves we obtained from these mortality models. In this example there are considerable trend changes or even jumps in the time series of three of these statistics right at the transition from the calibration to the forecasting period. This implies that the forecasts of deaths curves with these two well-established mortality models are not plausible from a demographic perspective in the sense that the forecasts for the immediate future are not consistent with the development in the immediate past. Also demographers make forecasts on future mortality based on their particular expertise. They, however, often focus on single aspects of the complete mortality curve and typically do not give a comprehensive picture of future mortality. This, in turn, is a feature of mortality models. We have developed a mortality model which closes this gap.

This new mortality model builds on the deaths curve, which can be seen as the density function of the distribution of lifetimes. Starting at the space of all density functions on the support of the deaths curve, we define requirements on the shape of the deaths curve to find a subset of this space with reasonably shaped deaths curves. In a second step we further restrict this set and we select all deaths curves which fit a prediction of the four statistics M , UB , DoI , and $d(M)$. This leads to a subset of the set of reasonably shaped deaths curves. If this final set is not empty, all deaths curves therein are regarded to be valid forecasts of our mortality model. In the paper we discuss some criteria which can be used to select a particular deaths curve as a unique forecast of the future deaths curve. If the set of deaths curves fitting the predictions of the statistics is empty, either the shape requirements are too restrictive or the predictions of the

statistics are not plausible.

Furthermore, we describe how we implement this model. To this end, we choose a deaths curve representation with B-Splines and describe an algorithm which yields a deaths curve forecast of our model. This algorithm starts at a given deaths curve – for example in the last year where the deaths curve can be observed – and alters the components of the B-Spline representation (i.e. the position and the weights of the splines) such that a reasonably shaped deaths curve is obtained which fits the forecasts of the four statistics.

Finally, we apply the mortality model to mortality data for Swiss females on the age range starting at age 60. We determine the most recent mortality scenario, which in this example, is the vector (*right-shifting mortality, extension, compression, concentration*) and we determine future deaths curves for different mortality scenarios: To illustrate that the mortality model produces demographically reasonable forecasts, we first extrapolate the most recent trends in the four aforementioned statistics from 2014 to 2070 (which we call the “base scenario”) and determine respective deaths curves for each future calendar year. We analyze if these forecasts reasonably extrapolate the observed deaths curve’s changes, the changes to the probability of death, and other demographic statistics such as, for example, the remaining period life expectancy. The results show that the recently observed trends are smoothly continued until 2070 by this model. Furthermore, we analyze four so-called “pure scenarios” in which only one of the four components continues the trend of the base scenario while the others remain constant. It becomes apparent that in our example only the pure right-shifting mortality scenario can be extrapolated until 2070 with reasonable shape, while this is not true for the other pure scenarios. This again underlines that pure scenarios can be observed for a certain time period, but in the long term only mixed scenarios can prevail. We analyze the remaining period life expectancy, which increases in the base scenario, for each pure scenario and find that the process of right-shifting mortality has a stronger impact on the increase of this statistic than all other components. To illustrate the flexibility of the model with respect to any expert’s assessment on future mortality scenarios we also forecast deaths curves for a “stress scenario” where the trend of right-shifting mortality and extension is intensified. For this example we compare the cohort life expectancy of the base scenario and the stress scenario. Our comparison shows that the cohort life expectancy is more than 5% higher in the stress scenario than in the base scenario.

Such analyses may be interesting, for example, for annuity providers to test their assumptions on future mortality for sensitivity to future changes of the deaths curve. This is facilitated by the model building on statistics with a clear demographic interpretation. Moreover, as the structure of this mortality model and its forecasting approach are fundamentally different from other mortality models it is a valuable addition to the toolkit of mortality projection models.

In summary, in this paper we develop a mortality model which is based on the extrapolation of four statistics with a clear demographic interpretation and thus forecasts deaths curves

which are consistent with recently observed demographic trends; also expert opinions to future demographic trends can easily be included. Hence this paper answers the fourth part of the research questions.

References

- Aarsen, K. and de Haan, L. (1994). On the maximal life span of humans. *Mathematical Population Studies* 4(4): 259–281. DOI: 10.1080/08898489409525379.
- Barbi, E., Lagona, F., Marsili, M., Vaupel, J.W., and Wachter, K.W. (2018). The plateau of human mortality: Demography of longevity pioneers. *Science* 360: 1459–1461. DOI: 10.1126/science.aat3119.
- Booth, H. and Tickle, L. (2008). Mortality Modelling and Forecasting: A Review of Methods. *Annals of Actuarial Science* 3(1-2): 3–43. DOI: 10.1017/S1748499500000440.
- Cairns, A.J.G., Blake, D., and Dowd, K. (2006). A Two-Factor Model for Stochastic Mortality with Parameter Uncertainty: Theory and Calibration. *The Journal of Risk and Insurance* 73(4): 687–718. DOI: 10.1111/j.1539-6975.2006.00195.x.
- Cairns, A.J.G., Blake, D., and Dowd, K. (2008). Modelling and Management of Mortality Risk: A Review. *Scandinavian Actuarial Journal* 2008(2-3): 79–113. DOI: 10.1080/03461230802173608.
- Canudas-Romo, V. (2008). The modal age at death and the shifting mortality hypothesis. *Demographic Research* 19(30): 1179–1204. DOI: 10.4054/DemRes.2008.19.30.
- Cheung, S.L.K. and Robine, J.-M. (2007). Increase in common longevity and the compression of mortality: The case of Japan. *Population Studies* 61(1): 85–97. DOI: 10.1080/00324720601103833.
- Cheung, S.L.K., Robine, J.-M., Tu, E.J.-C., and Caselli, G. (2005). Three dimensions of the survival curve: Horizontalization, verticalization, and longevity extension. *Demography* 42: 243–258. DOI: 10.1353/dem.2005.0012.
- Dong, X., Milholland, B., and Vijg, J. (2016). Evidence for a limit to human lifespan. *Nature* 538(7624): 257–259.
- Einmahl, J.H.J., Fils-Villetard, A., and Guillou, A. (2008). Statistics of extremes under random censoring. *Bernoulli* 14(1): 207–227.
- Fries, J.F. (1980). Aging, natural death, and the compression of morbidity. *New England Journal of Medicine* 303(3): 130–135.
- Gavrilov, L.A. and Gavrilova, N.S. (2011). Mortality Measurement at Advanced Ages: A Study of the Social Security Administration Death Master File. *North American Actuarial Journal* 15(3): 432–447. DOI: 10.1080/10920277.2011.10597629.
- Gbari, S., Poulain, M., Dal, L., and Denuit, M. (2017). Extreme Value Analysis of Mortality at the Oldest Ages: A Case Study Based on Individual Ages at Death. *North American Actuarial Journal* 21(3): 397–416.

- Gompertz, B. (1825). On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Mode of Determining the Value of Life Contingencies. *Philosophical Transactions of the Royal Society of London* 115: 513–585.
- Hanayama, N. and Sibuya, M. (2015). Estimating the Upper Limit of Lifetime Probability Distribution, Based on Data of Japanese Centenarians. *The Journals of Gerontology: Series A* 71(8): 1014–1021. DOI: 10.1093/gerona/glv113.
- HMD (2015). *Human Mortality Database*. University of California, Berkeley, and Max Planck Institute for Demographic Research. URL: www.mortality.org.
- IDL (2015). *International Database on Longevity*. Max Planck Institute for Demographic Research. URL: www.supercentenarians.org.
- Kannisto, V. (2000). Measuring the compression of mortality. *Demographic Research* 3(6). DOI: 10.4054/DemRes.2000.3.6.
- Kannisto, V. (2001). Mode and Dispersion of the Length of Life. *Population: An English Selection* 13(1): 159–171. URL: <http://www.jstor.org/stable/3030264>.
- Lee, R.D. and Carter, L. (1992). Modelling and Forecasting U.S. Mortality. *Journal of the American Statistical Association* 87(419): 659–671. DOI: 10.1080/01621459.1992.10475265.
- Maier, H., Gampe, J., Jeune, B., Robine, J.-M., and Vaupel, J.W., eds. (2010). *Supercentenarians. Demographic Research Monographs (A series of the Max Planck Institute for Demographic Research)*. Berlin, Heidelberg: Springer.
- Manton, K.G. and Tolley, D. (1991). Rectangularization of the Survival Curve: Implications of an Ill-Posed Question. *Journal of Aging and Health* 3: 172–193. DOI: 10.1177/089826439100300204.
- Myers, G.C. and Manton, K.G. (1984). Compression of mortality: Myth or reality? *Gerontologist* 24: 346–353. DOI: 10.1093/geront/24.4.346.
- Nusselder, W.J. and Mackenbach, J.P. (1996). Rectangularization of the Survival Curve in the Netherlands, 1950–1992. *The Gerontologist* 36(6): 773–782. DOI: 10.1093/geront/36.6.773.
- Oeppen, J. and Vaupel, J.W. (2002). Broken Limits to Life Expectancy. *Science* 296(5570): 1029–1031. DOI: 10.1126/science.1069675.
- Ouellette, N. and Bourbeau, R. (2011). *Recent adult mortality trends in Canada, the United States and other low mortality countries*. Paper presented at the Living to 100 Society of Actuaries International Symposium, Orlando, FL. URL: <https://www.soa.org/essays-monographs/2011-living-to-100/mono-li11-4b-ouellette.pdf>.
- Pollard, J.H. (1987). Projection of Age-Specific Mortality Rates. *Population Bulletin of the United Nations* 21/22: 55–69.
- Robine, J.-M., Cheung, S.L.K., Horiuchi, S., and Thatcher, A. R. (2008). *Is there a limit to the compression of mortality?* Paper presented at the Living to 100 and Beyond Society of Actuaries Symposium Orlando, FL. URL: <https://www.soa.org/essays-monographs/2008-living-to-100/mono-li08-03-cheung.pdf>.
- Rootzén, H. and Zholud, D. (2017). Human life is unlimited – but short. *Extremes* 20(4): 713–728. DOI: 10.1007/s10687-017-0305-5.

- Rossi, I. A., Rousson, V., and Paccaud, F. (2013). The contribution of rectangularization to the secular increase of life expectancy: An empirical study. *International Journal of Epidemiology* 42(1): 250–258. DOI: 10.1093/ije/dys219.
- Weon, B.M. and Je, J.H. (2009). Theoretical estimation of maximum human lifespan. *Biogerontology* 10(1): 65–71.
- Wilmoth, J.R. (2000). Demography of longevity: Past, present, and future trends. *Experimental Gerontology* 35: 1111–1129. DOI: 10.1016/S0531-5565(00)00194-7.
- Wilmoth, J.R. and Horiuchi, S. (1999). Rectangularization Revisited: Variability of Age at Death Within Human Populations. *Demography* 36(4): 475–495. DOI: 10.2307/2648085.

Research Papers

1 Extension, Compression, and Beyond - A Unique Classification System for Mortality Evolution Patterns

Source:

Börger, M., Genz, M., and Ruß, J. (2018). Extension, Compression, and Beyond - A Unique Classification System for Mortality Evolution Patterns. *Demography* Vol: 55(4) 1343–1361.

DOI: 10.1007/s13524-018-0694-3

Reprinted with permission from Springer Nature Customer Service Centre GmbH:

Springer Demography, *Extension, Compression, and Beyond - A Unique Classification System for Mortality Evolution Patterns*

by Matthias Börger, Martin Genz, and Jochen Ruß.

© 2018.

Extension, Compression, and Beyond

– A Unique Classification System for Mortality Evolution Patterns

Matthias Börger*

Martin Genz[†]Jochen Ruß[‡]

Abstract

A variety of literature addresses the question of how the age distribution of deaths changes over time as life expectancy increases. However, corresponding terms such as *extension*, *compression*, or *rectangularization* are sometimes defined only vaguely, and statistics used to detect certain scenarios can be misleading. The matter is further complicated because mixed scenarios can prevail, and the considered age range can have an impact on observed mortality patterns. In this article, we establish a unique classification framework for realized mortality scenarios that allows for the detection of both pure and mixed scenarios. Our framework determines whether changes of the death curve over time show elements of extension or contraction; compression or decompression; left- or right-shifting mortality; and concentration or diffusion. The framework not only can test the presence of a particular scenario but also can assign a unique scenario to any observed mortality evolution. Furthermore, it can detect different mortality scenarios for different age ranges in the same population. We also present a methodology for the implementation of our classification framework and apply it to mortality data for U.S. females.

Keywords: Mortality scenario classification, Rectangularization, Shifting mortality, Extension, Compression

*Institut für Finanz- und Aktuarwissenschaften (ifa), Lise-Meitner-Straße 14, 89081 Ulm, Germany, email: m.boerger@ifa-ulm.de

[†]Institut für Finanz- und Aktuarwissenschaften (ifa) and Institut für Versicherungswissenschaften, Universität Ulm, Lise-Meitner-Straße 14, 89081 Ulm, Germany, email: m.genz@ifa-ulm.de (corresponding author)

[‡]Institut für Finanz- und Aktuarwissenschaften (ifa) and Institut für Versicherungswissenschaften, Universität Ulm, Lise-Meitner-Straße 14, 89081 Ulm, Germany, email: j.russ@ifa-ulm.de

1 Introduction

Mortality evolutions – that is, realized changes in mortality rates – have been analyzed extensively in the last decades. These analyses typically examine changes in the distribution of lifetimes and hence go far beyond determining trends in the evolution of life expectancy. In this sense, changes in aggregated statistics such as life expectancy are simply a consequence of the underlying change of the age distribution of deaths.

A wide range of literature addresses the question of how realized mortality has changed over time and how patterns of past developments – which we also call *mortality scenarios* – can be described and classified. In this context, different terms have been created, for example rectangularization, compression, extension, expansion, and shifting mortality. These terms have been helpful in the analysis of historical mortality evolution patterns. Their definitions are, however, mostly intuitive, which can lead to ambiguity. For instance, Fries (1980) defined *rectangularization* as the convergence of the survival curve to a theoretical but not completely reachable final state, where everybody dies at the same age. Many authors have adopted this definition (see, e.g., Cheung et al., 2005; Kannisto, 2000; Manton and Tolley, 1991). However, as we show in the following section, this definition can be misleading. Similarly intuitive but difficult to verify from observed mortality patterns is the definition of compression in Debón et al. (2011) as a ”state in which mortality from exogenous causes is eliminated and the remaining variability in the age at death is caused by genetic factors.” Thus, a precise and feasible definition for each mortality scenario is necessary to test its occurrence in practice.

Furthermore, different authors have defined certain scenarios in different ways. In contrast to the aforementioned intuitive definition in Debón et al. (2011), many authors have used certain statistics of the deaths curve – that is, the age distribution of deaths – to define compression. According to Kannisto (2001), (old-age) compression can be observed if the modal age at death M (i.e., the age with the largest number of deaths) increases and $SD(M+)$ (i.e., the standard deviation of the distribution of deaths above M) simultaneously decreases. Other authors have (at least implicitly) applied this definition (e.g., Cheung and Robine, 2007; Ouellette and Bourbeau, 2011). Wilmoth and Horiuchi (1999), on the other hand, identified compression by a shrinking interquartile range (IQR) – that is, the length of the age range between the 25th and the 75th percentile of the distribution of deaths. Analogously, Kannisto (2000) used the so called $C\alpha$ -statistics – the shortest age range in which $\alpha\%$ of all deaths occur. Thatcher et al. (2010) observed compression if the slope parameter in a logistic mortality model increased with time. We show in the following section that different definitions of compression do not always yield the same results.

Scenario definitions can also be critical when they rely on observations for a rather small age range only. For instance, when analyzing the evolutions of M and $SD(M+)$, one completely ignores the mortality evolution for all ages less than M . As we show in the following section, if M increases and $SD(M+)$ decreases, compression need not be present for the whole age range

under consideration.

The distinction between different scenarios is also not always clear. For instance, Wilmoth (2000) stated that rectangularization should be "best thought of as 'compression of mortality'." For Myers and Manton (1984), compression and rectangularization also seem to be the same scenario. Others, like Nusselder and Mackenbach (1996), see rectangularization as a special case of compression in which the life expectancy increases with time. A similar issue exists for definitions of extension, expansion, and shifting mortality. For example, Debón et al. (2011) used the terms *expansion* and *shifting mortality* but did not explain the differences between them. Others have defined the three terms differently. For example, Wilmoth and Horiuchi (1999) used the term *expansion* if the force of mortality decreased faster for older ages than for younger ages. Bongaarts (2005), on the other hand, explained the scenario of shifting mortality as a result of "delays in the timing of deaths"; that is, the force of mortality curve exhibits simply a shift in age. Cheung et al. (2005) used the term *longevity extension* for a scenario in which longevity beyond the modal age at death increases.

Sometimes, one scenario is defined by the absence of some other scenario. For instance, Canudas-Romo (2008) regarded shifting mortality as a scenario in which "the compression of mortality has stopped." Obviously, such a definition implies that these scenarios – namely, shifting mortality and compression – are mutually exclusive, which rules out mixed scenarios by definition. As we show later, elements of different mortality scenarios can often be present at the same time. Therefore, analyses that focus solely on testing for one particular scenario – for example, compression – can never provide a comprehensive insight into the mortality evolution.

In this article, we address these issues and establish a unique classification framework for mortality scenarios. The framework is based on observed changes in the deaths curve for the age range under consideration. We build on existing concepts such as shifting mortality, extension, and compression, and then combine these concepts into a framework that particularly allows for the detection of mixed scenarios of mortality change. We provide precise definitions of scenarios and show how they can be identified.

Furthermore, our framework is applicable to any age range from some starting age to the age at which everybody has died. Thus, the age range can be chosen such that it best suits the question at hand. We show that different scenarios might prevail for different age ranges and that our framework can identify this. For instance, sometimes scenarios can be observed in which more and more deaths are shifted from younger to older ages but deaths become more and more evenly spread at the older ages. Such a scenario might be thought of as compression on the age range starting at 0 but quite the opposite on the age range starting at 60 (see the following section). We also provide a possible methodology for implementing our framework and show its practical applicability in an example.

2 Typical Issues with Scenario Definitions and Statistics

In this section, we identify and discuss some shortcomings of existing approaches for the classification of mortality scenarios. These shortcomings motivate a need for a new classification framework, as developed in the following section.

2.1 Imprecise Mortality Scenario Definitions

Mortality scenarios describe patterns in the evolution of mortality over time, this is a process of change. However, in the literature, we find several imprecise mortality scenario definitions. One example is the attempt to define the process of change solely by some (only theoretical and hence unreachable) final stage. This is the case when *rectangularization* is defined as a process in which the survival curve approaches a rectangular form. However, a rectangular form can be reached via different routes.

This is illustrated by the left panel of Fig. 1, which shows a hypothetical albeit not unrealistic evolution of deaths curves $d(x)$ over time.¹ Assume that at some point in time, mortality in a population follows the curve labeled State 1. At some later point, it follows State 2 and so on until it reaches State 5. Without using any formal definition, one would intuitively conclude that some scenario of compression takes place between States 3 and 5. Between States 1 and 3, however, a scenario that is somewhat the opposite of compression can be observed.

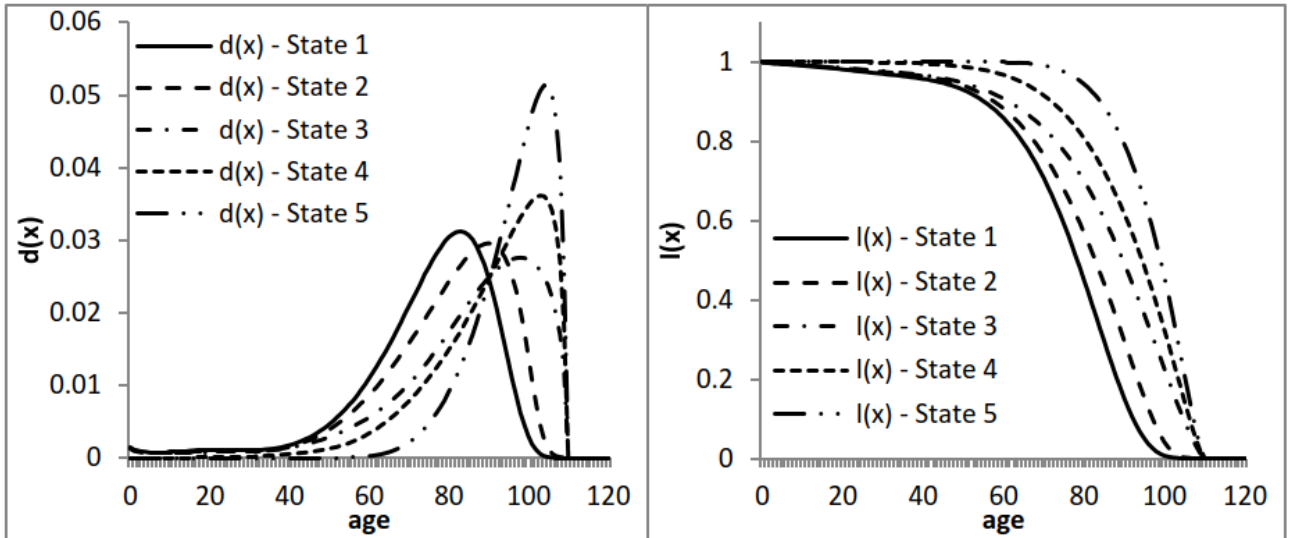


Figure 1: Mortality evolution in a hypothetical example. Left: deaths curves; right: survival curves

If, however, one looks at the corresponding survival curves $l(x)$ (right panel of Fig. 1), one might intuitively conclude that with every step, the shape becomes more rectangular. Therefore,

¹All deaths curves in this article are scaled such that the areas underneath the curves each integrate to 1. Thus, the corresponding survival curves start with a radix of 1. Also note that all examples in the second and third section of this article are based on hypothetical illustrative curves that are, however, reasonable given that overall mortality improves and life expectancy increases.

one might identify the whole transition from State 1 to State 5 as rectangularization, which is sometimes seen as a special case of compression. This clearly contradicts the observation that between States 1 and 3, the opposite of compression prevails.

Thus, the definition of a mortality scenario by some theoretical final state that is being approached will not always lead to a correct result.

2.2 Misleading or Insufficient Statistics

Of course, a reduction of complexity by looking at some key statistics of deaths or survival curves rather than at the whole curves is desirable. On the other hand, this approach always leads to a loss of information. Therefore, one should very carefully identify statistics that preserve the part of the information that is of interest. Unfortunately, for some statistics that are frequently being used to describe patterns of mortality changes, this is not the case (at least if they are not analyzed together with additional statistics). In this subsection, we will explain this point. Returning to Fig. 1, observe that the modal age at death M increases from state to state starting with 83 years in State 1 and reaching 104 years in State 5. At the same time, $SD(M+)$ decreases from state to state starting at 7.62 in State 1 and ending at 2.71 in State 5. Following, for example, Robine et al. (2008), this would mean that compression prevails throughout the process and, in particular, also between States 1 and 3, which is inconsistent with the intuition from looking at the left panel of Fig. 1.

Sometimes, different statistics designed to measure the same thing can lead to contradicting results. For example, compression is often defined by a reduction of the IQR and/or a $C\alpha$ -statistic (see Kannisto, 2000; Wilmoth and Horiuchi, 1999). Figure 2 shows two scenarios of mortality evolution in which the structures of the mortality distributions changed considerably from State 1 to State 2, with clear characteristics of mortality improvement and compression. However, IQR remains unchanged in the left panel of the figure, whereas $C50$ remains at the same value in the right panel. Thus, neither IQR nor $C50$ alone is always able to identify compression.

Such issues can always occur when changes of the entire deaths curve are identified using statistics that take into account only parts or certain points of the deaths curve.

2.3 Ignoring Mixed Scenarios

Next, we show that it may not be appropriate to define a certain mortality scenario as the opposite of some other scenario or, more generally, that mixed scenarios should be allowed for. Hence, more than one dimension is required to get a full picture of a mortality scenario.

A classical example is the relationship between shifting mortality² (or extension) and compression.

²As noted in the Introduction, the terms *expansion*, *extension*, and *shifting mortality* coexist in the literature. We consider *expansion* and *extension* to be the same, and use the term *extension* for that. We consider *shifting mortality* to be a different phenomenon, as explained in the next section.

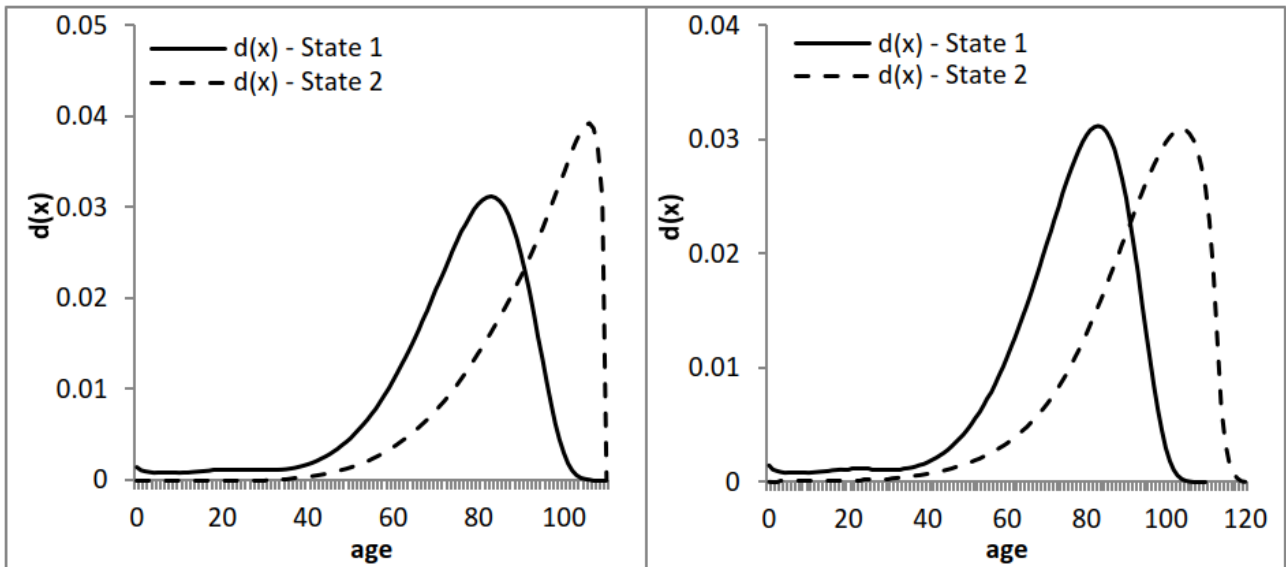


Figure 2: Mortality evolutions with compression. Left: unchanged IQR ; right: unchanged $C50$

The left panel of Fig. 3 shows a mixed scenario in which (in the transition from State 1 to State 2) shifting mortality and compression seem to coexist. Therefore, identification of one scenario should not rule out the other. Analogously, in the right panel of Fig. 3, neither shifting mortality nor compression can be observed. Thus, rejection of one scenario does not imply that the other scenario prevails. Thus, clearly it is not suitable to consider compression and shifting mortality as disjoint categories. This again shows the need for a more sophisticated classification system that combines different concepts of compression, shifting mortality, and so forth in the form of mixed scenarios.

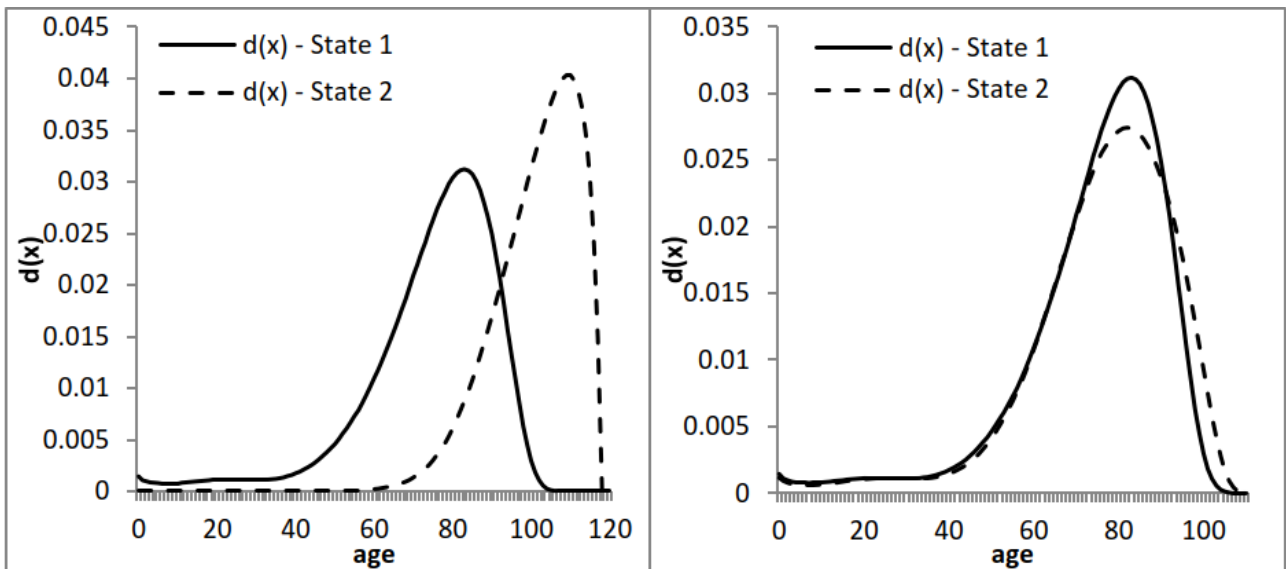


Figure 3: Two hypothetical examples. Left: shifting mortality and compression coexist; right: neither shifting mortality nor compression exists

2.4 Effect of Age Range

Different types of mortality evolution can occur in different age ranges. Myers and Manton (1984) compared the survival curve starting at age 0 with the survival curve starting at age 65 for U.S. females and males between 1962 and 1979. They observed a clear tendency toward rectangularization for the entire age range but not at the older ages. If one is interested primarily in a certain age range (e.g., old-age mortality), one should therefore consider only the corresponding part of the mortality curve.

However, when restricting the age range, undesired effects may occur whenever statistics are being used that depend on the number of people being alive at the beginning of the considered age range: for example, $d(M)$, the number of deaths at age M . Assume that one is interested in the age range starting at age 65. If between two points in time, younger-age mortality decreased, then more people would reach age 65. Even if older-age mortality did not change at all, $d(M)$ would increase (with M remaining unchanged), suggesting a change in old-age mortality. And if a change in old-age mortality actually occurred, the change in $d(M)$ would be affected by both the change in old-age mortality of interest and a change in younger-age mortality not of interest. These undesired effects can be eliminated by normalizing the population sizes such that at all considered points in time, the number of people alive at the beginning of the considered age range is the same (e.g., $l(65) = 1$).

The left panel of Fig. 4 shows some mortality evolution over the entire age range. Here, clearly compression toward higher ages can be observed. If one is interested only in the age range 65+, one might intuitively look at the respective age range of the left panel of Fig. 4 (i.e., without normalizing), which displays signs of compression. However, in the normalized curves (right panel of Fig. 4), the deaths curve of State 2 looks less dense than for State 1, which is an indication against compression.

3 A New Classification Framework for Mortality Scenarios

In the previous section, we identified shortcomings of existing approaches for the classification of mortality evolutions. We now propose a new framework in which unique mortality scenarios are defined based on observable changes in the shape of the deaths curve. In this section, we introduce the intellectual concept of the framework. In the next section, we describe a methodology that can be applied to estimate the statistics used in our framework and to identify trends and trend changes in these statistics.

Our framework combines and uniquely defines four concepts for the change of mortality over time that are well known from the literature: (1) shifting mortality (see, e.g., Canudas-Romo, 2008), (2) longevity extension (see, e.g., Rossi et al., 2013), (3) compression of mortality (see,

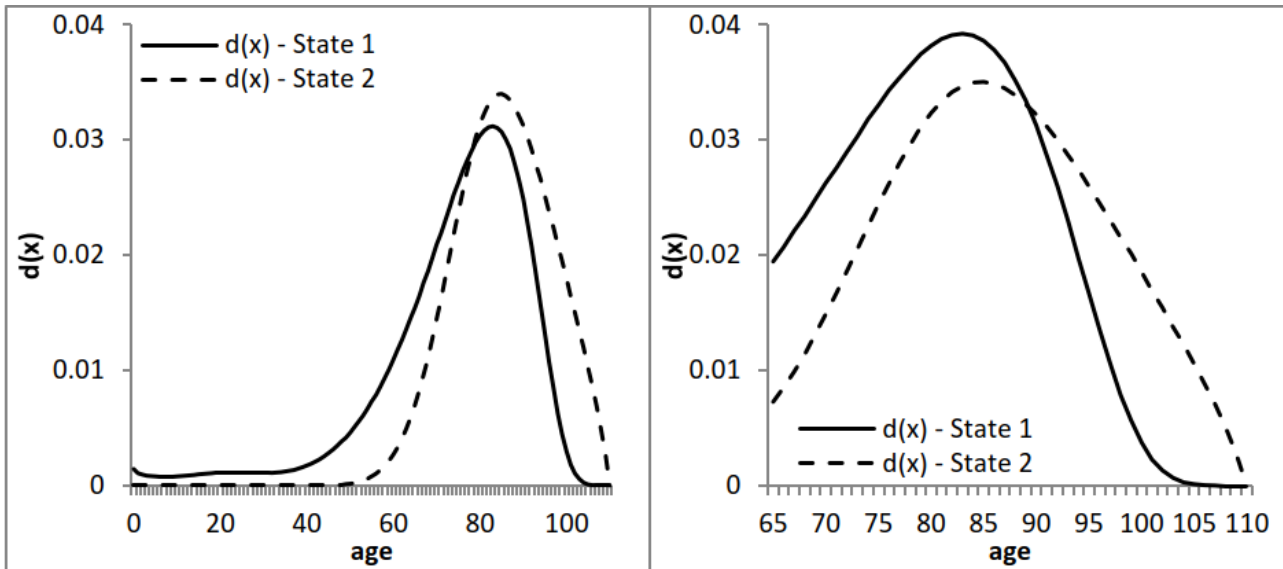


Figure 4: Mortality evolution with increasing number of survivors to age 65. Left: complete age range; right: starting at age 65 with normalized $l(65)$

e.g., Myers and Manton, 1984), and (4) concentration of mortality (see, e.g., Kannisto, 2001). As we will show, only a combined look at all four dimensions – which automatically allows us to consider both pure and mixed scenarios – gives a full picture of the considered mortality evolutions.

Our classification framework can be applied for any age range that includes the right tail of the deaths curve. Depending on the question at hand, the age range could start, for example, at 0, some juvenile age, or retirement age. In particular, it is possible that the classification framework identifies different mortality scenarios for different age ranges (see the upcoming section on the application of our classification framework for an example).

For any given age range, we use four key characteristics of the deaths curve, each corresponding to one of the aforementioned concepts. Significant changes in one or several characteristics over time mean that the deaths curve has changed. Conversely, if these four characteristics remain unaltered, changes in a deaths curve are regarded as immaterial. We will show that these four characteristics are sufficient to distinguish between a great variety of deaths curves and to uniquely classify mortality scenarios. The four characteristics are as follows:

1. The position of a deaths curve's peak is measured by the modal age at death M and describes general shifts in the distribution of deaths. Because the shape of a deaths curve typically changes over time, a pure shift of the entire deaths curve will rarely occur, and therefore we consider its center M as a reference point. An increase in M indicates *right-shifting mortality*, and a decrease in M implies *left-shifting mortality*. In this section, we assume that the modal age at death can be determined uniquely.³

³The peak might not be unique in only rather theoretical scenarios – for example, because of multiple peaks of the same height or a plateau. In such a case, one might use a suitable alternative to M or modify the framework to include additional statistics.

2. The support of a deaths curve is determined by its upper bound, which we refer to as UB .⁴ We denote the respective changes of UB as *extension* (if UB increases over time) and *contraction* (if UB decreases over time). Estimating UB in practice involves some ambiguity; see the following section for more details.
3. The degree of inequality in the distribution of deaths, which we denote by DoI , is the least obvious of the four key characteristics. However, Fig. 5 shows two deaths curves that are significantly different, although the other three statistics of our framework coincide. Therefore, an additional statistic related to the shape of the curve is required. The deaths curve of State 2 is almost 0 up to age 50, while State 1 shows a somewhat more balanced distribution of deaths over all ages. DoI is designed to pick up such differences by measuring the equality/inequality of the distribution of deaths over the whole age range. Intuitively, a low value of DoI indicates that deaths are rather equally distributed over the whole age range considered and vice versa. We use the terms *compression/decompression* if DoI increases/decreases; see the following section for more details.

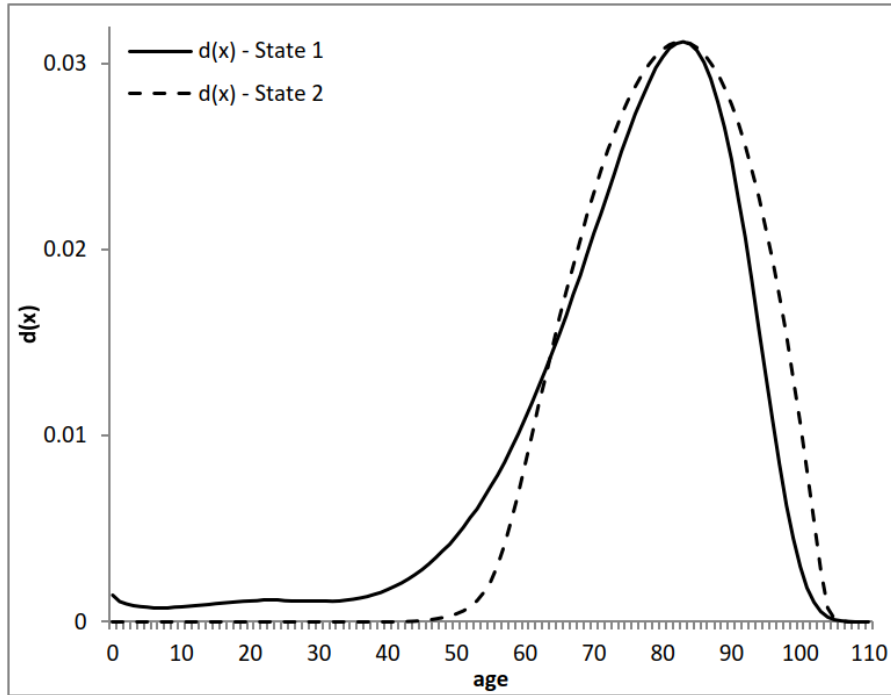


Figure 5: Mortality evolution with constant M , $d(M)$, and UB , but changing DoI

4. The height of the peak of a deaths curve is given by $d(M)$. This component addresses the evolution of a deaths curve at and close to its center, M . An increase in $d(M)$ is referred to

⁴In theory, UB can exist only if the probability of death reaches 1 for some age. If the probability of death remains below 1 for all ages, any age could be reached in principle. Research by several authors (see, e.g., Gampe, 2010) has indicated that probabilities of death typically flatten out at very old ages, possibly somewhere near 0.5. Thus, the population surviving up to such ages would get halved every year; but if the initial population was large enough, there would be a few survivors up to any age. Therefore, one could argue that UB does not exist in theory, which is, however, irrelevant for our application.

as concentration and indicates that the distribution of deaths becomes more concentrated around M . The counterpart to concentration is what we refer to as diffusion, and it is observed if $d(M)$ decreases. Similar to DoI , $d(M)$ can also be seen as an indicator for the equality/inequality of the distribution of deaths. A large $d(M)$ implies that many deaths are concentrated at and around M . However, $d(M)$ is a more local measure for a small region around M , whereas DoI measures the equality/inequality of the distribution of deaths over the whole age range.

Of course, each of the four components can remain unchanged over time. In this case, the respective component is referred to as *neutral*. Thus, every component can attain three states.⁵ Two of the four aforementioned statistics (UB and M) primarily determine the position of the deaths curve, and the other two ($d(M)$ and DoI) primarily describe its shape. We believe that these four characteristics provide a good trade-off between granularity and complexity. The four components are summarized in Table 1.

In principle, any combination of the three states for each component is possible, implying that we can classify both pure and mixed scenarios, which was one of the requirements we outlined earlier. In a pure scenario, only one component of the scenario vector is different from neutral. For instance, the vector (*neutral*, *extension*, *neutral*, *neutral*) denotes a pure extension scenario. On the other hand, a vector such as (*neutral*, *extension*, *compression*, *neutral*) describes a mixed scenario, which contains elements of both extension and compression. In total, $3^4 = 81$ mortality scenarios are possible, which might seem unfeasible at first glance. However, many scenarios will hardly be observed in practice – for example, (*left-shifting mortality*, *extension*, *compression*, *diffusion*). Those scenarios are nevertheless part of our classification framework to make sure that there are no unclassifiable evolutions and that classifications are unique.

4 Methodology for the Implementation of the Classification Framework

The application of the classification framework introduced in the preceding section involves two main steps. First, the four statistics need to be estimated from deaths curves for each year in the observation period. A reasonable estimator for each of the statistics is proposed in the following subsection. Thereafter, trends in the resulting time series need to be analyzed to determine the prevailing states in each of the four scenario components, as we address in a later subsection. Obviously, various different estimators and methods could be used in both steps, and thus the specific estimators and methods described in this section are only one possible implementation.

⁵If a distinction between different intensities of increase or decrease is desired, more than three states can be considered or additional information about the slope of the respective trend line (see the section on methodology) can be added.

Scenario Component	Attainable States	Criterion (in terms of deaths curve characteristic and statistic to be computed)
1	Right-shifting mortality	Peak shifts to the right; M increases
	Left-shifting mortality	Peak shifts to the left; M decreases
	Neutral	Peak does not move; M constant
2	Extension	Support is prolonged; UB increases
	Contraction	Support shrinks; UB decreases
	Neutral	Support remains unchanged; UB constant
3	Compression	Distribution of deaths less balanced; DoI increases
	Decompression	Distribution of deaths more balanced; DoI decreases
	Neutral	Distribution of deaths equally balanced, DoI constant
4	Concentration	More deaths at/around M ; $d(M)$ increases
	Diffusion	Less deaths at/around M ; $d(M)$ decreases
	Neutral	Number of deaths at/around M unchanged; $d(M)$ constant

Table 1: Scenario components, attainable states, and criteria

4.1 Estimation of Statistics

We now explain how we calculate the four statistics from the deaths curve in any given year. Both, raw or smoothed deaths curves can be used in principle. In our application later in the article, we explain why we prefer using smoothed data.

For the *position of a deaths curve's peak* measured by M , we use the following estimator by Kannisto (2001):

$$M = x_{d_{max}} + \frac{d(x_{d_{max}}) - d(x_{d_{max}} - 1)}{(d(x_{d_{max}}) - d(x_{d_{max}} - 1)) + (d(x_{d_{max}}) - d(x_{d_{max}} + 1))}, \quad (1)$$

where $x_{d_{max}}$ is the age for which the largest number of deaths is observed. As a byproduct, the *height of a deaths curve's peak* ($d(M)$), can then be estimated by the number of deaths at age $x_{d_{max}}$:

$$d(M) = d(x_{d_{max}}). \quad (2)$$

For the *upper bound of a deaths curve's support* (UB), we use the age at the α percentile of the distribution of deaths, x_α , plus an estimate for the remaining life expectancy at that age. Thus, the estimator for UB is

$$UB = x_\alpha + e_{x_\alpha}. \quad (3)$$

This approach builds on Rossi et al. (2013), who proposed using the 90th percentile of the distribution of deaths as an approximation for the highest attainable ages. We prefer our combined estimator because it is considerably less biased. In our application, we set $\alpha = 99\%$. For the populations we analyzed, this choice provides a reasonable compromise between cutting

off only a small part of the distribution of deaths and stability in the statistic's evolution over time. For smaller (sub)populations, however, smaller values for α might be more appropriate. The statistic measuring the *degree of inequality* (DoI) in the distribution of deaths needs to take into account the whole age range. Therefore, statistics such as $SD(M+)$, IQR , or $C\alpha$ – which, as explained earlier, are commonly used to measure compression – are not feasible. An intuitive alternative is the area between the actual deaths curve and a hypothetical flat deaths curve $d_{flat}(x)$ as illustrated in Fig. 6. Using discrete data, this area can be approximated by summing the absolute differences in the numbers of deaths between the two deaths curves. Thus, we estimate DoI as

$$DoI = c \cdot \sum_{x=x_0}^{\lfloor UB \rfloor} |d(x) - d_{flat}(x)| = c \cdot \sum_{x=x_0}^{\lfloor UB \rfloor} \left| d(x) - \frac{l_{x_0}}{(UB - x_0 + 1)} \right|, \quad (4)$$

where x_0 is the starting age of the deaths curve, and $c = \frac{\lfloor UB \rfloor - x_0 + 1}{2 \cdot l_{x_0} (\lfloor UB \rfloor - x_0)}$ is a scaling factor such that DoI assumes its maximum value of 1 if all people die at the same age. The minimum value of DoI is 0 if deaths are uniformly distributed over all ages – that is, if $d(x) = d_{flat}(x)$ holds for all x .

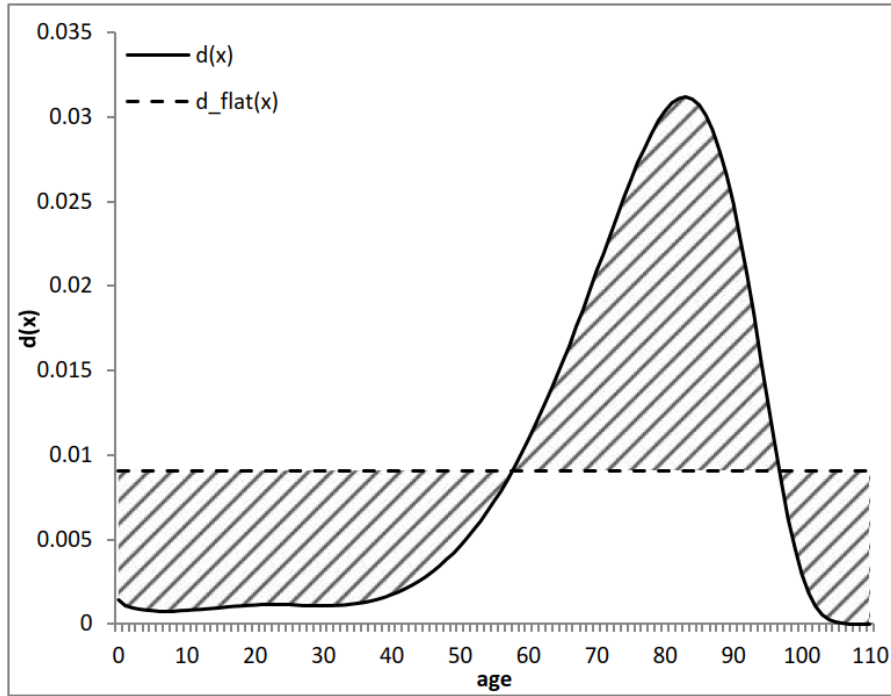


Figure 6: DoI as the area between observed deaths curve $d(x)$ and hypothetical flat deaths curve $d_{flat}(x)$

Note that the dependence of DoI on UB is uncritical in our framework given that we are interested only in changes of DoI over time. A potential misestimation/bias of UB would affect DoI in the same way for each point in time. Further, changes in UB over time do not automatically imply changes in DoI . For instance, if UB increases while the deaths curve's

shape does not change materially, the slight changes of $d(x)$, $d_{flat}(x)$, and the scaling factor c would basically cancel each other.

As mentioned earlier, alternative estimators could be used for the four statistics. In particular, an extensive literature explores measuring UB , which is sometimes referred to as maximum lifespan (see, e.g., Finch and Pike, 1996) or finite lifespan (see Fries, 1980). Alternative estimators for UB can be found in, for example, Cheung and Robine (2007), Fries (1980), and Wilmoth (1997). As alternative measures for DoI , one could consider the variance in the number of deaths, the Gini index, as proposed by Debón et al. (2011); or entropy, as originally proposed by Demetrius (1974) and adopted by Keyfitz (1985) and Wilmoth and Horiuchi (1999). These statistics also consider the whole age range as required. However, the Gini index and entropy are defined on the survival curve, which makes them less intuitive in our deaths curve-based framework.

4.2 Determination of Prevailing States

After the four statistics are estimated for each year in the observation period (see an example of the resulting time series in Fig. 7), the trends prevailing at each point in time need to be determined. We now introduce a possible methodology that we found to be suitable for all data sets we analyzed. However, a different methodology or modifications of our methodology – for example, with respect to the significance levels in the different tests – could be used and might be advisable for certain applications.

Elimination of Outliers

Potential outliers should be eliminated because they are irrelevant with respect to long-term trends and can significantly blur the trend analysis. Such outliers are typically caused by extreme events, such as the Spanish flu pandemic. To detect whether a data point is an outlier, we fit a linear regression to the 10 adjacent data points. The sample variance of the residuals (assumed to be normally distributed) can then be used to derive a 99 % prediction interval for the data point under consideration. If the data point lies outside the prediction interval, it is considered an outlier.

Determination of Trends, Trend Changes, and Jumps

To determine trends in the four statistics, we fit piecewise linear trends to the respective time series. Most of the time, mortality evolves rather steadily over time, and hence the piecewise linear trends should connect continuously. However, jumps can occur in case of extreme changes – such as the fall of the Soviet Union or a world war – or changes in data processing methods. Thus, at every data point of a time series under consideration, the previous linear trend can persist, a new trend can commence starting at the end point of the previous trend (change in slope), or a new trend can commence at some other level (jump and change in slope). The

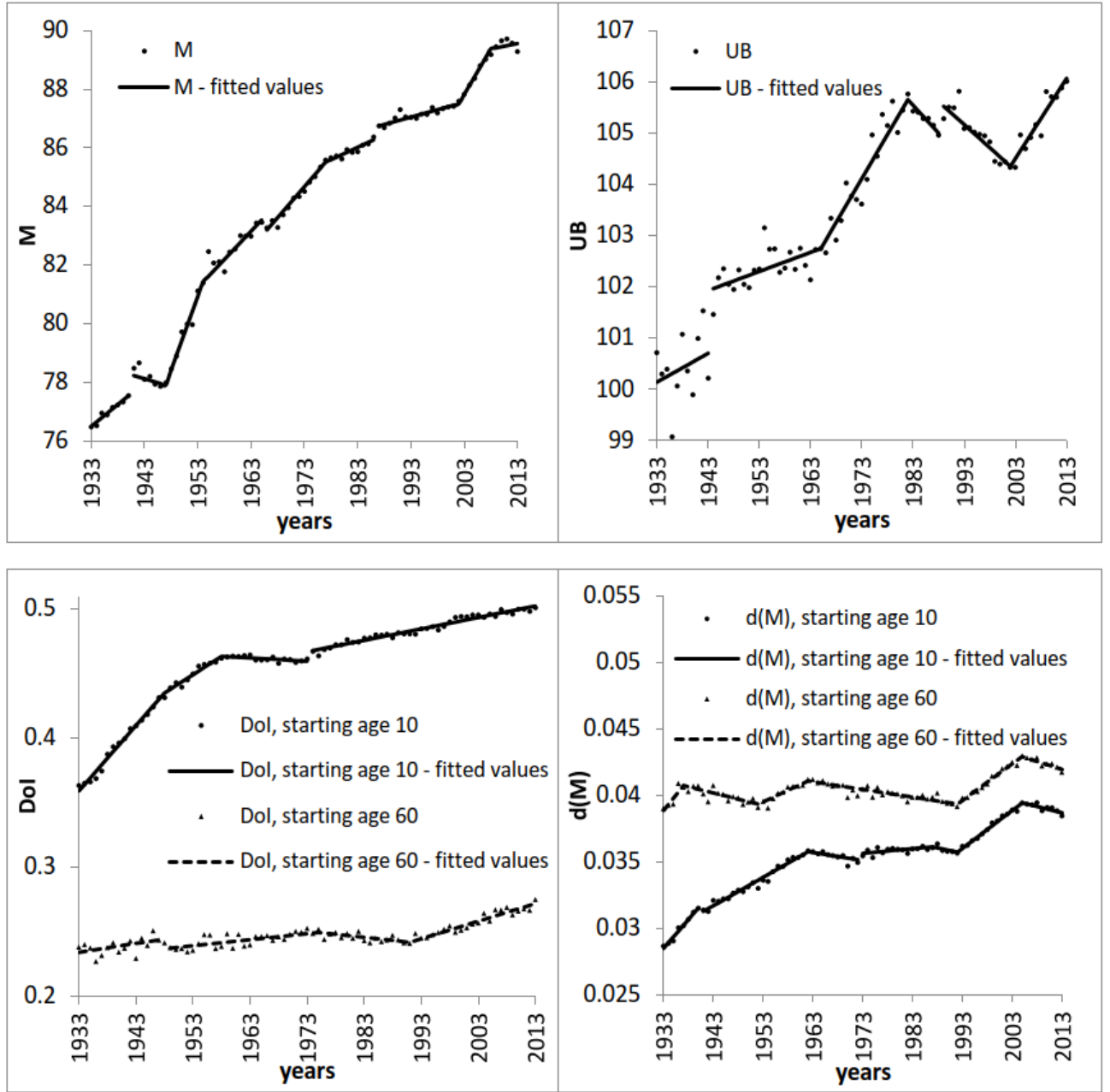


Figure 7: Development of the four components of our classification framework for U.S. females from 1947 to 2013. Upper-left panel: M ; upper-right panel: UB ; lower-left panel: DoI ; lower-right panel: $d(M)$

following methodology first determines which of the three possibilities is the most likely for each data point and then analyzes how many changes in slope and jumps are most suitable to describe the structural patterns in the entire time series and where they should occur.

To identify *candidate* data points for trend changes – that is, changes in slope with or without jumps – we first perform a preliminary analysis. We carry out three fits for every possible combination of three data points:⁶ (1) a straight regression line to the data from the first to the third data point, (2) a continuous regression line to the data from the first to the third

⁶If the time series has k data points, we consider all $k \times (k - 1) \times (k - 2)/6$ possible triples.

data point with a change in slope at the second data point, and (3) two straight regression lines (to the data from the first to the second and from the second to the third data point, respectively) that allow for a jump at the second data point. A set of Chow tests (see Chow, 1960) is used to determine which trend evolution is most likely for the second data point, under the assumption that adjacent trend changes are located at the first and third data point or that these data points are the first or last data points of the entire time series. In the first Chow test (significance level of 1%), the null hypothesis of one persistent trend – that is, no change at the second data point – is tested against a continuous change in slope. The result of the test (the new null hypothesis) is then tested against a jump in a second Chow test. The results of the Chow tests usually depend on the choice for the first and third data points. Thus, whether a data point is a candidate for a trend change (and if so, of which kind) depends on the position of the neighboring trend changes.

After the preliminary analysis, we use the following main algorithm to identify the number and locations of trend changes that result in an optimal fit:⁷

- *Step 1:* We commence by fitting a straight regression line to the entire time series. This is the case of no trend change (i.e., the number of possible trend changes n is 0).
- *Step 2:* The number of possible trend changes is increased from n to $n + 1$.
- *Step 3:* We determine the sample variances of the residuals from the fit with n trend changes. They will be required as variance estimators in Step 5. The sample variances are to be computed separately for each period with constant trend. We use a regime-switch argument here to justify that the variance can change when the trend changes and thereby allow for heteroscedasticity as can be observed in Fig. 7, for example.
- *Step 4:* Building on the preliminary analysis, we determine all feasible combinations of $n + 1$ candidate data points for trend changes. The preliminary analysis also indicates for each candidate data point whether the trend change would be a change in slope with or without jumps. If there is no feasible combination, the fit with n trend changes is the overall optimal fit, and the algorithm terminates.
- *Step 5:* For each feasible combination of trend changes from Step 4, we fit a piecewise linear trend curve to the data (and allow for discontinuities only when the type of the potential trend change is a change with jump). To account for heteroscedasticity, we use the sample variances from Step 3 as weights.
- *Step 6:* The optimal trend change positions (and thus also the trend change types) for $n + 1$ trend changes are determined by comparing the fits from Step 5 by the Akaike information criterion (AIC) (Akaike, 1973). The number of parameters is two (initial

⁷The presentation of the algorithm aims for a clear presentation of and distinction between the steps involved and does not pay attention to computing efficiency.

intercept and slope) plus $n + 1$ for the trend change positions plus $n + 1$ for the changes in slope plus one for every jump (i.e., the new intercepts after the jumps).

- *Step 7:* Finally, we compare the optimal fits with n and with $n + 1$ trend changes to assess the contribution of the additional trend change to the time series representation. To this end, we use another Chow test (again with significance level 1%). Because the original test by Chow considers only one trend change versus none, we use an extended version of the test. The test statistic remains unchanged, but the number of parameters increases (one for each trend change position, each intercept, and each slope). Note that for $n \geq 2$, we can also account for heteroscedasticity in this test by applying variance estimates from the optimal fit with $n - 1$ trend changes as weights. The null hypothesis in the Chow test is the case of n trend changes. Thus, the additional trend change is accepted only if it significantly improves the fit, which is in line with our intention of determining long-term trends. If the null hypothesis stands, the time series can be adequately described by n trend changes, and the algorithm terminates. If the additional trend change significantly improves the fit, we return to Step 2.

Testing for Increasing, Decreasing, or Neutral Statistics

Finally, we have to determine whether the resulting trend curve (see the lines in Fig. 7) should be considered increasing, decreasing, or neutral in the context of our framework. For each period with constant trend, we use an F test with a significance level of 10% to analyze whether the slope of the trend is significantly different from 0. If the slope is not significantly different from 0, the state neutral is assigned. Otherwise, we consider the statistic as increasing (decreasing) if the slope is positive (negative) during the corresponding period. This definition implies that the state neutral is assigned not only if the slope is clearly close to 0 but also if the uncertainty in the underlying data is too large to identify a significant trend.

5 Application of the Classification Framework

In this section, we apply our classification framework to the mortality evolution of females in the United States.⁸ We derive log mortality rates $\ln(m(x, t))$ for ages 0 to 109 from the deaths and exposure data in the Human Mortality Database (HMD, 2015) for years 1933–2013. For each calendar year, these log mortality rates are then smoothed and extrapolated using P splines, allowing us to derive normalized and smoothed deaths curves. We prefer this approach over using the raw deaths curves from the HMD for several reasons: (1) potential disturbing effects

⁸We also applied the framework to several other populations, such as Sweden, Japan, and West Germany. In all cases, the framework yielded reasonable and informative results. For the sake of brevity, however, we show the results for only one population. We chose U.S. females for illustration because the variety of different observed scenarios was the largest. See Genz (2017) for an application of our framework to a larger number of countries and a comparison of the respective mortality patterns.

resulting from birth cohorts of different sizes are eliminated; (2) random effects in the data, which might lead to double peaks in the deaths curve, are significantly reduced; (3) the potential effect of age misspecifications in the raw data, particularly with respect to estimating UB , is reduced; and (4) the time series for the four statistics exhibit fewer random fluctuations and are thus easier to analyze.

We consider deaths curves covering different age ranges as discussed earlier. The curves $d_{10}(x, t)$ start with a fixed radix at age 10 and thus exclude effects from infant mortality, whereas the $d_{60}(x, t)$ curves allow for an analysis of mortality at typical retirement ages. Figure 7 displays the four components of our classification framework for both starting ages along with the respective piecewise linear trend lines.⁹

By definition, the curves for the modal age at death M coincide for both starting ages. From a theoretical perspective, the same holds for UB . However, the chosen estimator yields slight differences for the different starting ages. Because the two sets of data points would be difficult to distinguish, and the resulting scenarios for this component are the same for both starting ages, we display UB only for starting age 10.

From Fig. 7, we note that our framework identifies several trend changes for each of the statistics and both starting ages. Such trend changes can mean that (1) the direction of the trend changes (e.g., from increasing to neutral or decreasing), or (2) only the intensity of the trend (i.e., the slope of the trend line) changes significantly while its direction remains unchanged. For example, the first two trend changes in M are changes in the direction of the trend, whereas the subsequent trend changes (except the last one) concern only the intensity of the increase (i.e., the pace of the right shift in mortality). Thus, a trend change does not inevitably lead to a change in the scenario vector. Moreover, as mentioned earlier, trends can change with or without a jump in the absolute level of the statistic. In our example, such jumps occur for all statistics except $d(M)_{60}$.

The direction of each trend as well as the position of the trend changes and their types – that is, a change in slope with or without an upward/downward jump – are summarized in Fig. 8. This representation allows for an easy visual assessment of the scenario vector at each point in time. For instance, in the year 2010, for both starting ages, the scenario vector is $(0, +, +, -)$: that is, the scenario is neutral with respect to shifting mortality and exhibits extension, compression, and diffusion at the same time.

By comparing Fig. 7 with Fig. 8, we find some periods with seemingly increasing (decreasing) trends in Fig. 7 but a classification as neutral in Fig. 8. One such example is the first trend for UB . Here, the underlying data has a relatively strong variance, and therefore the seemingly increasing trend is not significant, as explained earlier.

The results of our analysis particularly underline the need for combining different concepts of

⁹We also considered the starting ages 0 (i.e., the complete age range) and 30 in order to exclude effects of young adult's mortality, such as accidents. The observed scenarios for starting ages 0, 10, and 30 are quite similar.

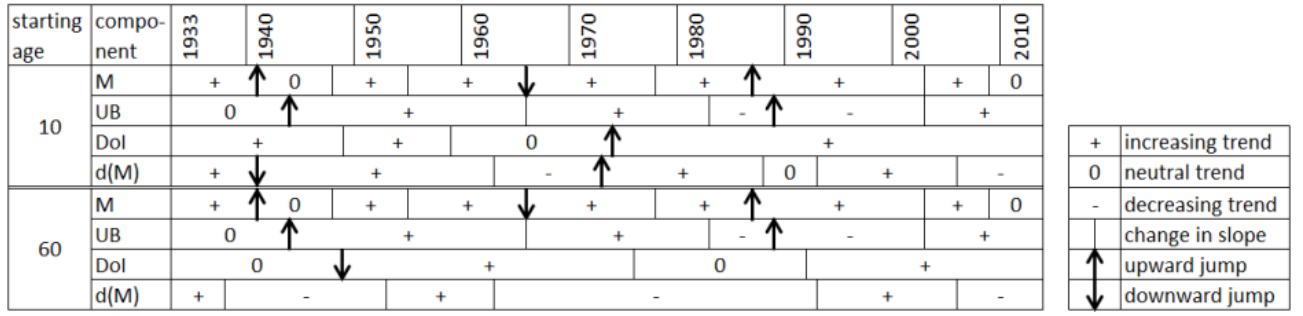


Figure 8: Time bars of mortality evolution for U.S. females, each statistic, and both starting ages

mortality change in one framework given that we observe mixed scenarios over almost the whole observation period. There are even periods in which all four indicators change – for example, between 1973 and 1982 for starting age 10. During this period, we simultaneously observe right-shifting mortality and extension (i.e., both the mode and the upper bound of the deaths curve move to the right) combined with a compression of the whole curve and an increase of the concentration around the mode. In contrast, pure scenarios seem to be very rare: only for starting age 60 and years 1941 to 1948 do we find a scenario of pure diffusion.

Furthermore, the results show that each of our four components is relevant in the sense that no component can be explained by the others. For instance, as one would expect, M and UB increase over the observation period in general: we observe right-shifting mortality and extension. However, particularly for UB , we observe the opposite trend (i.e., contraction) for some periods (e.g., the 1990s), and thus these two statistics do not move in the same direction throughout the entire observation period. This also holds for DoI and $d(M)$, although they also frequently follow the same trend. For example, after 2006, $d(M)$ decreases for both starting ages, while DoI increases for both starting ages: we observe diffusion and compression at the same time. The results also highlight the importance of choosing a suitable age range. For both DoI and $d(M)$, we find several periods in which the trends differ by starting age. For instance, between 1975 and 1990, we observe compression for starting age 10 but decompression for starting age 60.

6 Conclusion

In this article, we explain why many existing approaches to classify patterns of mortality evolution have four major shortcomings: (1) mortality scenario definitions are often imprecise and intuitive rather than rigorous; (2) some frequently used statistics are not always sufficient to identify the respective scenarios; (3) mixed scenarios are usually not accounted for; and (4) the effect of the considered age range is often ignored.

We propose a new framework for classifying patterns of mortality evolution. Our approach is based on changes of the deaths curve and uses four statistics that should be considered

simultaneously. Each mortality scenario then consists of four components: (1) the deaths curve can exhibit a right-shift or a left-shift or be neutral in that respect; (2) the deaths curve can exhibit extension or contraction or be neutral in that respect; (3) the deaths curve can exhibit compression or decompression or be neutral in that respect; or (4) the deaths curve can exhibit concentration or diffusion or be neutral in that respect. This approach overcomes the shortcomings of previous approaches: each mortality evolution is uniquely and precisely classified; by considering all four components simultaneously, mixed scenarios are automatically detected; and the framework is applicable to different age ranges.

For some of the statistics used, the estimation is not straightforward. Beyond an introduction of the intellectual concept of the framework, we therefore also introduce a methodology that can be used to estimate the statistics and determine trends and trend changes in the data. We apply our approach to data for U.S. females, illustrating that the structure of the change in mortality can be quickly assessed and well understood. We further demonstrate empirically that none of the four components can be explained by the other three and that results can significantly differ for different age ranges.

The purpose of our framework is a classification of realized mortality evolutions. In this sense, it is purely descriptive: it does not provide explanations for observed trends and trend changes. It seems obvious that any research that intends to provide such explanations or seeks to explore a link between determinants of mortality and observed patterns of mortality change needs as a prerequisite a common understanding of which pattern of mortality change has been observed in which situation. Our methodology can provide this and hence serves as a basis for such research. In particular, the detected trend changes can be an indication when and how demographic changes have occurred. Similarly, by applying our framework to different populations, time and structure of differences in their demographic evolutions can be detected, which again can serve as a basis for research on the causes.

If a mortality model is to be calibrated to historical data, our framework can also be used to identify suitable time spans (e.g., without major trend breaks). Further, the framework can be applied for testing whether existing mortality projections are consistent with observed trends in the most recent history.

References

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In: *2nd international symposium on information theory, Tsahkadsor, Armenia, USSR, September 2-8, 1971*. Ed. by B.N. Petrov and F. Csáki. Budapest, Hungary: Akadémiai Kiadó: 267–281.
- Bongaarts, J. (2005). Long-range trends in adult mortality: Models and projection methods. *Demography* 42(1): 23–49. DOI: 10.1353/dem.2005.0003.

- Canudas-Romo, V. (2008). The modal age at death and the shifting mortality hypothesis. *Demographic Research* 19(30): 1179–1204. DOI: 10.4054/DemRes.2008.19.30.
- Cheung, S.L.K. and Robine, J.-M. (2007). Increase in common longevity and the compression of mortality: The case of Japan. *Population Studies* 61(1): 85–97. DOI: 10.1080/00324720601103833.
- Cheung, S.L.K., Robine, J.-M., Tu, E.J.-C., and Caselli, G. (2005). Three dimensions of the survival curve: Horizontalization, verticalization, and longevity extension. *Demography* 42: 243–258. DOI: 10.1353/dem.2005.0012.
- Chow, G.C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica* 28(3): 591–605. DOI: 10.2307/1910133.
- Debón, A., Martínez-Ruiz, F., and Montes, F. (2011). *Temporal evolution of some mortality indicators: Application to Spanish data*. Paper presented at the 2011 Living to 100 Society of Actuaries International Symposium, Orlando, FL. URL: <https://www.soa.org/essays-monographs/2011-living-to-100/mono-li11-1b-debon.pdf>.
- Demetrius, L. (1974). Demographic parameters and natural selection. *Proceedings of the National Academy of Sciences* 71(12): 4645–4647. DOI: 10.1073/pnas.71.12.4645.
- Finch, C.E. and Pike, M.C. (1996). Maximum Life Span Predictions From the Gompertz Mortality Model. *The Journals of Gerontology, Series A: Biological Sciences & Medical Sciences* 51A(3): B183–B194. DOI: 10.1093/gerona/51A.3.B183.
- Fries, J.F. (1980). Aging, natural death, and the compression of morbidity. *New England Journal of Medicine* 303(3): 130–135.
- Gampe, J. (2010). Human Mortality Beyond Age 110. In: *Supercentenarians. Demographic Research Monographs (A series of the Max Planck Institute for Demographic Research)*. Ed. by H. Maier, J. Gampe, B. Jeune, Robine J.-M., and J.W. Vaupel. Berlin, Heidelberg: Springer: 219–230. DOI: 10.1007/978-3-642-11520-2_13.
- Genz, M. (2017). *A Comprehensive Analysis of the Patterns of Worldwide Mortality Evolution*. Paper presented at the 2017 Living to 100 Society of Actuaries International Symposium, Orlando, FL. URL: <https://www.soa.org/essays-monographs/2017-living-to-100/2017-living-100-monograph-genz-paper.pdf>.
- HMD (2015). *Human Mortality Database*. University of California, Berkeley, and Max Planck Institute for Demographic Research. URL: www.mortality.org.
- Kannisto, V. (2000). Measuring the compression of mortality. *Demographic Research* 3(6). DOI: 10.4054/DemRes.2000.3.6.
- Kannisto, V. (2001). Mode and Dispersion of the Length of Life. *Population: An English Selection* 13(1): 159–171. URL: <http://www.jstor.org/stable/3030264>.
- Keyfitz, N. (1985). *Applied Mathematical Demography*. 2nd ed. New York, NY: Springer Verlag.
- Manton, K.G. and Tolley, D. (1991). Rectangularization of the Survival Curve: Implications of an Ill-Posed Question. *Journal of Aging and Health* 3: 172–193. DOI: 10.1177/089826439100300204.

- Myers, G.C. and Manton, K.G. (1984). Compression of mortality: Myth or reality? *Gerontologist* 24: 346–353. DOI: 10.1093/geront/24.4.346.
- Nusselder, W.J. and Mackenbach, J.P. (1996). Rectangularization of the Survival Curve in the Netherlands, 1950–19921. *The Gerontologist* 36(6): 773–782. DOI: 10.1093/geront/36.6.773.
- Ouellette, N. and Bourbeau, R. (2011). *Recent adult mortality trends in Canada, the United States and other low mortality countries*. Paper presented at the Living to 100 Society of Actuaries International Symposium, Orlando, FL. URL: <https://www.soa.org/essays-monographs/2011-living-to-100/mono-li11-4b-ouellette.pdf>.
- Robine, J.-M., Cheung, S.L.K., Horiuchi, S., and Thatcher, A. R. (2008). *Is there a limit to the compression of mortality?* Paper presented at the Living to 100 and Beyond Society of Actuaries Symposium Orlando, FL. URL: <https://www.soa.org/essays-monographs/2008-living-to-100/mono-li08-03-cheung.pdf>.
- Rossi, I. A., Rousson, V., and Paccaud, F. (2013). The contribution of rectangularization to the secular increase of life expectancy: An empirical study. *International Journal of Epidemiology* 42(1): 250–258. DOI: 10.1093/ije/dys219.
- Thatcher, A.R., Cheung, S.L.K., Horiuchi, S., and Robine, J.-M. (2010). The compression of deaths above the mode. *Demographic Research* 22(17): 505–538. DOI: 10.4054/DemRes.2010.22.17.
- Wilmoth, J.R. (1997). In search of limits. In: *Between Zeus and the salmon: The biodemography of longevity*. Ed. by K.W. Wachter and C.E. Finch. Washington, DC: National Academies Press: 38–64.
- Wilmoth, J.R. (2000). Demography of longevity: Past, present, and future trends. *Experimental Gerontology* 35: 1111–1129. DOI: 10.1016/S0531-5565(00)00194-7.
- Wilmoth, J.R. and Horiuchi, S. (1999). Rectangularization Revisited: Variability of Age at Death Within Human Populations. *Demography* 36(4): 475–495. DOI: 10.2307/2648085.

2 A Comprehensive Analysis of the Patterns of Worldwide Mortality Evolution

Source:

Genz, M. (2017). *A Comprehensive Analysis of the Patterns of Worldwide Mortality Evolution*. Paper presented at the 2017 Living to 100 Society of Actuaries International Symposium, Orlando, FL.

URL: <https://www.soa.org/essays-monographs/2017-living-to-100/2017-living-100-monograph-genz-paper.pdf>

Copyright © 2017 by the Society of Actuaries, Schaumburg, Illinois. Reprinted with permission.

A Comprehensive Analysis of the Patterns of Worldwide Mortality Evolution

Martin Genz*

Abstract

A variety of literature deals with the question how the age distribution of deaths develops over time, and many different notions have been established for certain scenarios. In Börger et al. (2016), a classification framework has been developed that allows for a unique classification of mortality evolution patterns. In particular, the framework assigns a unique scenario to any possible mortality evolution. In contrast to many other classification approaches, this approach allows for so-called mixed scenarios, such as a combination of elements of compression and shifting mortality. Thus, it provides a more comprehensive picture of historical and potential future mortality evolution patterns.

In the present paper, we briefly summarize this classification framework and discuss issues in its practical application. Then we apply the framework to mortality data for different countries all over the world. This yields a complete picture of historical mortality evolution patterns in those countries and adds to existing analyses where only certain aspects of mortality evolution patterns have been considered (e.g., a test for one scenario like compression) for only one or a few countries. We then discuss similarities and differences in the historical mortality evolution patterns between different populations. We also apply the framework to different age ranges, since sometimes different scenarios can be observed for different age ranges, even within one population.

*Institut für Finanz- und Aktuarwissenschaften (ifa) and Institut für Versicherungswissenschaften, Universität Ulm, Lise-Meitner-Straße 14, 89081 Ulm, Germany, email: m.genz@ifa-ulm.de

1 Introduction

The age-specific structure of mortality changes probably all over the world. For instance, life expectancy has been increasing in most countries over the last decades. This, however, is only a particular symptom of the underlying change of the deaths curve (i.e., the age distribution at death). Thus, the question of how the structure of mortality changes in detail goes beyond the change in life expectancy: What are the main drivers of this evolution? Which age ranges are mostly affected by the change of the mortality structure? Moreover, potential dependencies and differences between the evolution of the structure of mortality for different populations often need to be analyzed in order to get an idea of supra-regional or even global trends in the deaths curve's evolution.

Many recent publications deal with the evolution of the mortality structure of single populations, including Nusselder and Mackenbach (1996) for the Netherlands, Cheung et al. (2009) for Switzerland, and Debón et al. (2011) for Spain. Of course, such an analysis usually is motivated by the objective of the research, but the results may be strongly affected by country-specific circumstances (e.g., the country's social structure or health care system). Moreover, supra-regional or even global trends in the deaths curve's evolution cannot be detected, and it is not possible to separate them from population-specific effects. That is why the analysis of mortality structures should be set in a context of coherent populations. In contrast, some authors analyze trends in mortality evolutions of different populations (e.g., Canudas-Romo, 2008; Edwards and Tuljapurkar, 2005; Kannisto, 2000; Robine et al., 2008; Thatcher et al., 2010). However, often such analyses are mere demonstrations of the framework presented in the particular articles, so the focus is less on the comparison of the evolution of mortality structures between different countries. Only a few authors focus on the comparison of the trends in mortality evolution, but they typically restrict their analyses to specific aspects of the mortality structure. For example, Ouellette and Bourbeau (2011) explore trends in the adult mortality structure of 10 different countries. They inspect the evolution of the two statistics, the modal age at death M and the standard deviation above the modal age at death, which they call $SD(M+)$. These statistics were also suggested by, for example, Kannisto (2001). In a recent paper, Börger et al. (2016) observed that these two statistics are often not sufficient, particularly if the focus is on the change of the complete mortality structure, rather than on certain characteristics of its evolution. Also, Edwards (2011) has an exclusive focus on a certain characteristic of the deaths curve's evolution – in particular on the inequality of the age distribution of deaths. To this end, he analyzes the mortality data of 180 different populations. However, the inequality or dispersion of the distribution is, of course, just one aspect in which the structure of mortality can change. For instance, life expectancy might increase over time while the dispersion of the age distribution of deaths stays constant or even decreases. Finally, Viner et al. (2011) use data of 50 different populations, which are clustered in terms of amount of income. However, the focus of this research was specifically on "mortality trends in children and young people."

In summary, there is only little research on the differences in the development of the mortality structure between different populations. Authors covering that topic mainly focus on specific age ranges or particular characteristics of the mortality structure.

The present paper aims to fill this gap and provides a comparison of the past trends in the change of the (full) deaths curve of 34 different populations. To this end, we use a unique classification system for mortality evolution patterns that has been introduced by Börger et al. (2016). Based on the deaths curve, it can be applied to different age ranges and makes use of four statistics: The modal age at death M and the upper bound of the deaths curve's support UB measure changes in the position of a deaths curve over time, while the degree of inequality DoI and the number of deaths at the modal age at death $d(M)$ display changes in the deaths curve's shape. The framework not only offers a classification of the trends in the deaths curve's evolution of one country, but also allows for a comparison of such trends between different populations (e.g., between countries) or subpopulations (e.g., between females and males, different age groups, or different socioeconomic or ethnic groups) at a glance.

For our analysis, we use sex-specific data and compute the age distribution at death for the starting ages 0 and 60. This provides insights into the trends in the mortality structure for the complete age range, as well as for the mortality structure of female and male retirees.

The remainder of this paper is organized as follows: In Section 2, we briefly describe the data and outline how we prepare our input data prior to the analysis. In Section 3, we discuss the classification framework we use. Section 4 covers some issues in the practical application of the described framework, such as the smoothing of time series of single statistics, mathematical methods for the detection of trends, and methods for analyzing the results. In Section 5, we present and discuss our results. Finally, Section 6 concludes.

2 Data

To the best of our knowledge, there is no single, publicly available database that includes mortality data for every population in the world, has a sufficiently long history, and has a data quality that allows for an immediate comparison of data from different populations. Thus, we have to find a trade-off between completeness, availability, quality, and standardization of our input data. We use data from the Human Mortality Database (HMD, 2015), since it provides mortality data that is highly standardized, of good quality, and easily available. Unfortunately, the HMD provides mortality data for only 38 countries. For some of these countries, the HMD offers data for subpopulations,¹ which leads to a total of 42 different populations. Besides these shortcomings in terms of completeness, we have decided to eliminate those calendar years in the history of each population's data where the HMD alerts the user to the lower quality of the (input) data. Moreover, for the discussion of the results in Section 5, we only focus on

¹These subpopulations are East and West Germany for Germany, England and Wales, Scotland and Northern Ireland for the United Kingdom, and non-Maori and Maori populations for New Zealand.

rather recent trends. Thus, we use only data after 1920 – i.e., roughly one century back – for each population, even if the HMD offers data for the time prior to 1920.² Finally, our analysis requires input data for each population with a sufficiently long history, since we aim to detect long term trends in the mortality evolution, rather than short-term fluctuation of mortality evolution patterns. We have therefore decided to include only populations with a data history of at least 40 years. This leaves us with 34 populations, a large portion of which are European (see Section 4 for a list). For each population, we use sex-specific data, meaning we consider females and males separately.

We use deaths and exposure-to-risk data each with discrete calendar year and age subscript to calculate the discrete log force of mortality $\log(m_{x,t})$ for every age x between 0 and 109 and every calendar year t where we have data. Thereafter, we smooth and extrapolate these $\log(m_{x,t})$ in the age direction using P-splines. Given these smoothed log force of mortality curves for every calendar year, we calculate the deaths curves, which we scale such that they integrate to the value of 1 (i.e., we chose the radix 1 for the calculation of the corresponding survival curve). Hence, these curves can be interpreted as the density functions of the distribution of the age at death of a particular population. We calculate the curves for all populations, both sexes, and every calendar year. Moreover, we compute deaths curves for starting ages 0 and 60, which allows for a separate analysis of the trends of the deaths curve for the entire age range on the one hand, and the age range of retirement on the other hand. Note that the deaths curve with starting age 60 is the density function of the age distribution at death conditional on survival to age 60.

3 The Classification Framework

Recently, Börger et al. (2016) developed a unique classification framework for mortality evolution patterns, and they describe the advantages of this framework compared with previous approaches. For instance, the framework gives a clear definition for each scenario that might (or might not) prevail in the evolution of the deaths curve. Moreover, it allows for so-called mixed scenarios (e.g., compression and extension at the same time) and is able to measure any significant change in the deaths curve's position and shape over time. As we use this framework here, we give a short overview in this section.

The basic idea of this framework is to focus on the deaths curve and analyze its changes over time. To this end, the framework consists of four components:

The first component is the modal age at death, M . It is defined as the position of the deaths curve's peak. Whenever this peak moves to the left or to the right (i.e., M is decreasing or increasing), this is called *left-* or *right-shifting mortality*.

²The choice of the year 1920 is reasonable, since especially in the 1910s we have several historic events that blur the long-term trends of the mortality evolution in many countries (for example, World War I and the Spanish flu).

The second component is the upper bound of the deaths curve's support, UB , which is defined as the age where the survival curve reaches 0.³ An increase or decrease of UB means that the support of the deaths curve extends or contracts. Consequently, UB indicates *extension* and *contraction*. These first two components, M and UB , together describe the evolution of the deaths curve's position over time.

The third component of the framework is the so-called degree of inequality DoI , which was developed in Börger et al. (2016) to measure *compression* and *decompression*. Following their definition, compression is a process in which the deaths curve becomes more unequal over time. A perfectly equal deaths curve over all ages in this respect would be a uniform distribution of the age at death on the age interval $[0, UB]$. Whenever the difference between this hypothetical and the realized deaths curve increases or decreases, this is called compression or decompression, respectively.

Finally the fourth component is the number of deaths at the modal age at death $d(M)$, which describes the relative importance of the age M compared with the remaining ages. Whenever $d(M)$ increases or decreases, this is called *concentration* or *diffusion*, respectively. The latter two components, DoI and $d(M)$, together describe the evolution of the deaths curve's shape over time.

Besides increasing and decreasing trends in these four statistics, the evolution of a statistic can, of course, be neutral, which means that no (significant) change has occurred over time in the concerned component. Thus, every possible evolution of a deaths curve between two points in time can be expressed as a four-dimensional vector, where each component denotes the trend of one statistic. For example, if we observe an increase in the first and third statistic while the second and fourth do not change significantly, this vector would be (*right-shift*, *neutral*, *compression*, *neutral*). Since we have four statistics and three attainable states of each component of this "scenario vector", we obtain 81 different theoretically possible scenarios. Note that some of these scenarios might not be relevant in practice. However, the complete set of the 81 scenarios guarantees that a scenario can be assigned to every possible mortality evolution. Also note that the framework only detects qualitative rather than quantitative changes. Thus, we do not answer the question of the pace of, say, right-shifting mortality, but answer the question if we observe it at all.

For the estimation of the statistics, we follow the methodology described by Börger et al. (2016). Note that, taking a theoretical perspective, both M and UB are independent from the starting age of the deaths curve (as long as the starting age is smaller than M), although in practice, the chosen estimators induce slight, nonsignificant differences between starting ages. That is why, in what follows, we only display and discuss M_0 and UB_0 (i.e., the time series for starting age 0) in order to reduce the number of charts and simplify the presentation of the results.

³We admit that in a theoretical, continuous setting, this age may not exist. However, we use data of finite populations and in a discrete setting. Thus, there is an age where the oldest member of a population dies.

4 Issues in Practical Application of the Framework

In this section, we cover issues that arise after the four statistics described in Section 3 have been calculated. Figure 1 exemplarily shows the evolution of M for French males between 1920 and 2013. We do not smooth the input data by calendar years, so we observe considerable random fluctuations in this figure. These random fluctuations can blur the long-term trends in the time series, so we have to find methods to clearly decide when and where the statistic is increasing, neutral, or decreasing, respectively.

In this respect, we search for periods (with specified limits) on which the time series follow a linear trend. As a result, we get the direction of the respective trend (increasing, neutral, and decreasing) and the position of the trend changes (i.e., the limits of these periods). Trend changes can have two different qualities: either only the direction of the trend changes (we observe a continuous trend change within the time series) or both the direction of the trend and the level of the time series change at the same time (we observe a jump within the time series). Such jumps may be caused by historical events (for example, World War II or the fall of the Soviet Union) but also may be a result of changes in data-processing methods.

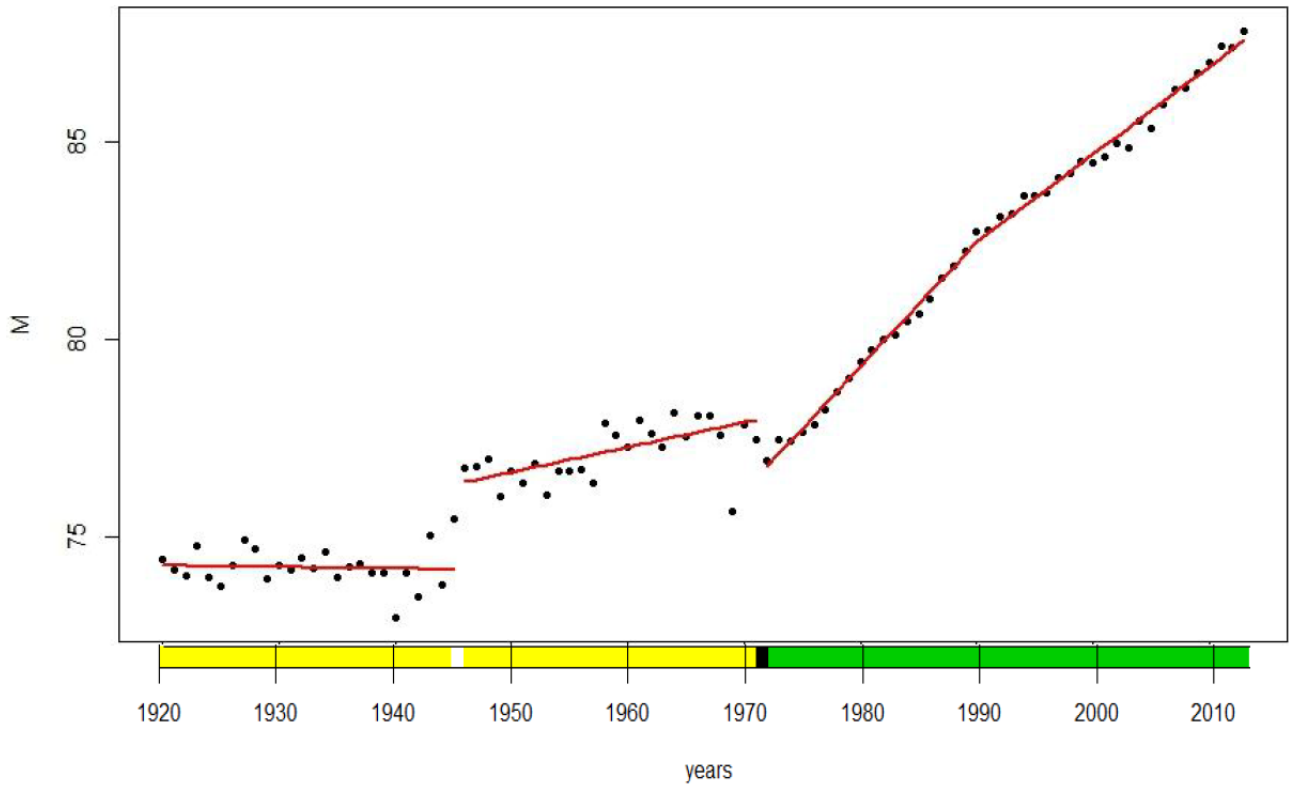


Figure 1: Evolution of Modal Age at Death (M) for French Males, 1920-2013

Note: The red line shows the fitted polygonal function. On the time bar, colors identify trends: red = decreasing trend (not present in the example), yellow = neutral trend, green = increasing trend, white = upward jump, black = downward jump.

We take a three-step approach proposed by Börger et al. (2016) to find the trends and trend changes of the time series and refer to their paper for more details. Here we only give the basic

idea: In the first step, we identify and eliminate outliers. For example, in Figure 1, we can detect a considerable downward outlier in the year 1969, which has to be eliminated. In the second step, we identify time periods within which the time series follows constant linear trends. This particularly includes a determination of the optimal number of trend changes. As a result, we have determined a fit of the time series (see red line in Figure 1). In the third step, for each period from the second step, we test whether a given trend is significantly different from 0. This gives us the desired direction of the trends (i.e., increasing, neutral, and decreasing). The results of this test for M for French males are illustrated with a colored time bar in Figure 1. To facilitate the analysis of the large number of results below, we will display time bar plots that allow for an easy and intuitive comparison of the trends of each statistic between different populations. To this end, we simultaneously display colored time bars (as in Figure 1) for several populations in order to compare the trends and trend changes between the respective populations at a glance. Green intervals on such time bars indicate increasing trends, yellow and red intervals display neutral and decreasing trends, respectively. As mentioned above, we observe both continuous trend changes and jumps. In the time bar plots, the color changes only if the direction of the trend changes (e.g., from increasing to decreasing). If, however, the direction remains unchanged and only the slope changes, this cannot be detected in such time bar plots. For example in Figure 1, the continuous trend change around 1990 is not shown in the time bar, as the direction of the trend does not change. In contrast, downward and upward jumps are always illustrated with a white or a black one-year trend, respectively, even if the direction of the trend does not change (see, for example, the upward jump in 1946 in Figure 1). For our analysis, we construct these plots separately by sex and starting age. Since neighboring countries frequently have similar trends in the mortality evolution, we sort the time bars by regional clusters:

- **Northwestern Europe:** Sweden, Norway, Iceland, Finland, Denmark, Scotland, England and Wales
- **Central Europe:** Netherlands, Belgium, Luxemburg, West Germany, Austria and Switzerland
- **Southwestern Europe:** France, Spain, Portugal and Italy
- **Eastern Europe:** East Germany, Poland, Czech Republic, Slovakia, Estonia, Latvia, Hungary, Bulgaria, Belarus, Ukraine and Russia
- **North America:** United States and Canada
- **Asia-Pacific area:** Japan, Australia and New Zealand (non-Maori and Maori)

Beyond displaying the trends in the four statistics over time for each population, we also analyze similarities and differences across populations. For this, we have developed a figure that we

call relative similarity (RS). Since we have only three potential values for each trend process at each point in time, we use a rather simple approach: Let $T^{(\cdot)} = \{t_0^{(\cdot)}, \dots, t_{n(\cdot)}^{(\cdot)}\}$ be the time range of a considered time series, and let $(u_t)_{t \in T^u}$ and $(v_t)_{t \in T^v}$ be two trend processes. Then $u_{t_i^u} \in \{\text{decreasing, neutral, increasing}\}$ and $v_{t_j^v} \in \{\text{decreasing, neutral, increasing}\}$, respectively, with $i \in \{0, \dots, n_u\}$ and $j \in \{0, \dots, n_v\}$. Moreover, the number of common data points of these two trend processes is given by $N(u, v) = \min(t_{n_u}^u, t_{n_v}^v) - \max(t_0^u, t_0^v) + 1$. For these data points, we determine the relative similarity as

$$RS(u, v) = \frac{1}{N(u, v)} \sum_{\tau \in T^u \cap T^v} \delta_\tau,$$

where

$$\delta_\tau = \begin{cases} -1 & \text{if } (u_\tau = \text{decreasing and } v_\tau = \text{increasing}) \text{ or } (u_\tau = \text{increasing and } v_\tau = \text{decreasing}), \\ 0 & \text{if } (u_\tau = \text{neutral and } v_\tau \neq \text{neutral}) \text{ or } (u_\tau \neq \text{neutral and } v_\tau = \text{neutral}), \text{ and} \\ 1 & \text{if } u_\tau = v_\tau. \end{cases}$$

Note that the value of RS is always in the interval $[-1, 1]$, where -1 means perfect dissimilarity and 1 means perfect similarity. This concept not only allows us to compare single statistics between different populations; we can also calculate the relative similarity of two populations – for example, for females with starting age 0 in total. This can be done by computing the average of the relative similarities for the four statistics.

5 Results

In this section, we present and analyze our results. We first analyze the trends (and trend changes) in the mortality evolution of 34 populations for males with starting age 0 (i.e., the complete age range) as a reference. We focus on the identification of supra-regional patterns, and thereby we identify most similarities or differences in the trends of the deaths curve's evolutions of the considered populations. In the second part of this section, we also analyze the trends in the mortality evolution of females and the trends in old-age mortality (i.e., with the starting age 60). For the sake of brevity, however, we then only highlight differences compared with the findings for males with starting age 0.

5.1 Reference Trends in the Mortality Evolution

In this subsection, we discuss the time bar plots for each statistic in detail before we summarize these findings in order to provide a complete picture of the recent mortality evolution of males with starting age 0. Since we have data for only few countries for the early years after 1920, searching for supra-regional patterns is difficult. The first year where we have data for each

population is 1971. Hence, we mainly focus on rather recent trends when we discuss the results. Figure 2 displays the time bars for the modal age at death M . Prior to 1960, we find hardly any supra-regional patterns in the trends of M . The only observation worth noting in this period is an upward jump in the middle of the 1940s for many European populations, but also for Canada. This is probably caused by World War II. During the 1960s, however, we observe a relatively short period of left-shifting mortality (i.e., a decrease in M), particularly for most northern and central European populations. However, for the majority of the populations outside eastern Europe (and few other exceptions), we observe an inversion of this trend almost at the same time around 1970 and an increase in the modal age at death afterward until the end of the observation period.

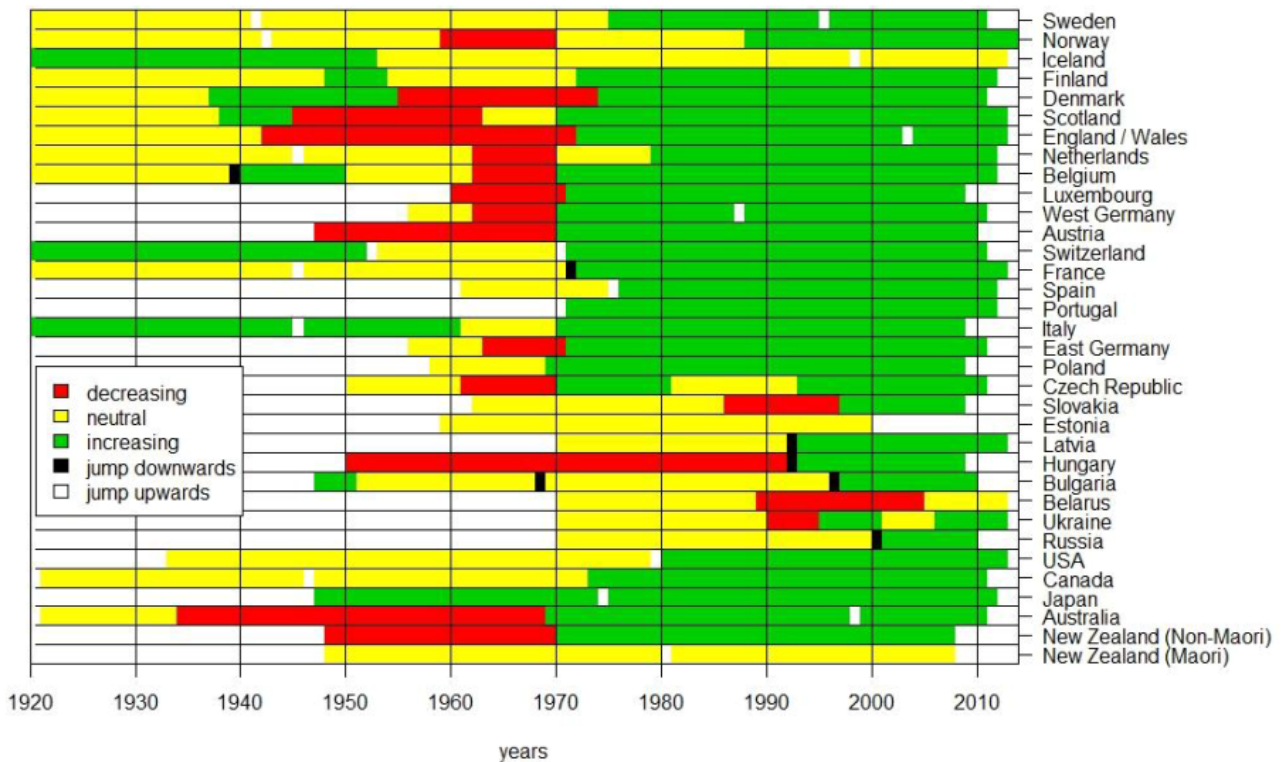


Figure 2: Trends in the Evolution of Modal Age of Death (M): Males, Starting Age 0

In contrast, the modal age at death in most eastern European populations (except East Germany and Poland) does not follow these trends. In particular, after 1970, we observe only neutral trends or even left-shifting mortality at least until the early 1990s. It seems that the fall of the Soviet Union for these countries had a certain impact on the evolution of the deaths curve, because since then, we observe right shifts for many eastern European populations. However, for the easternmost European populations, this trend change is less clear than for the other eastern European populations, and thus we observe a certain degree of heterogeneity within the eastern European cluster after 1990.

Figure 3 shows the time bars for the upper bound of the deaths curve's support UB . Also here, we can find apparent differences between eastern European populations and the other

populations. Around 2000 at the latest, however, the direction of the trends of UB for all eastern European populations assimilates to the direction of the trends for the other populations. Moreover, we also find considerable differences within the eastern European cluster: On the one hand, we observe contraction for the easternmost populations until the late 1990s. On the other hand, we find only neutral periods (neither extension nor contraction) for the other eastern European populations before 1990. Thus, the long-term trends within this cluster are not homogeneous prior to 1990.

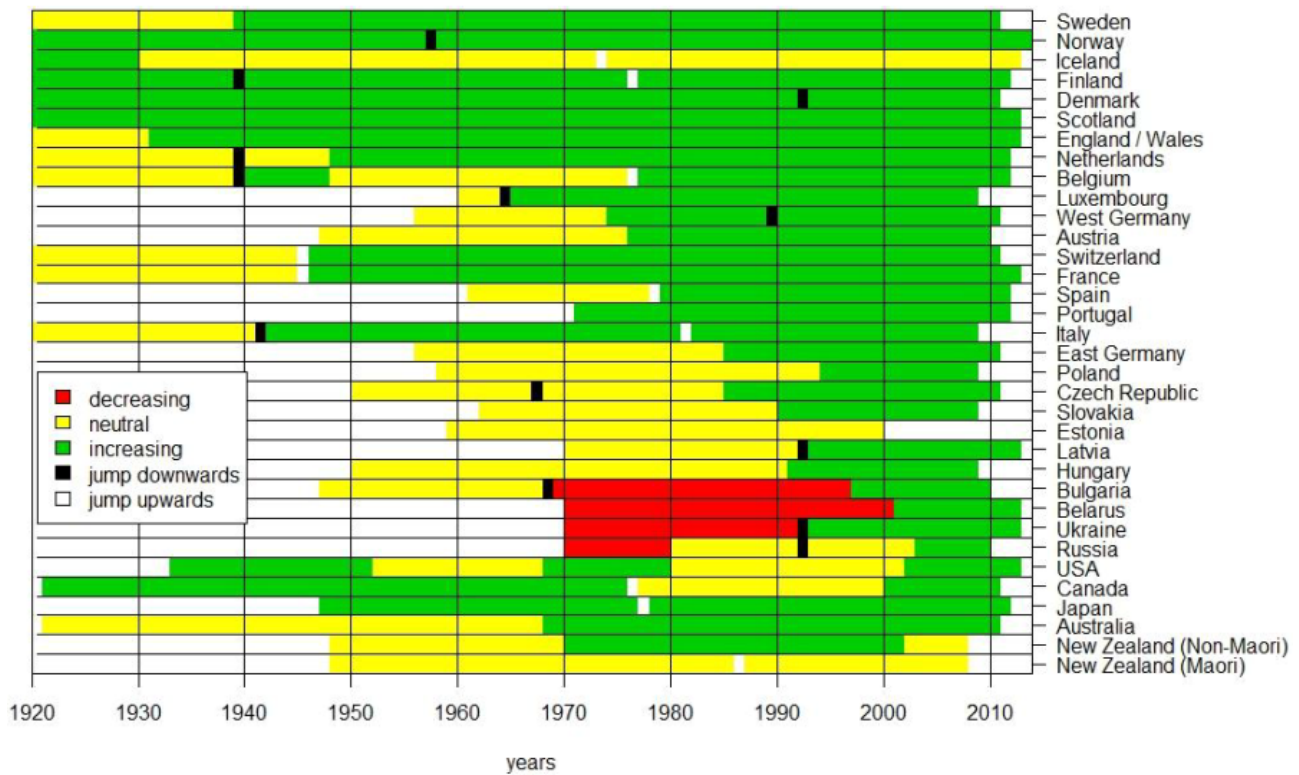


Figure 3: Trends in the Evolution of the Upper Bound of the Deaths Curve's Support (UB): Males, Starting Age 0

Apart from the observations for the eastern European cluster, there are hardly any supra-regional trend changes for UB , especially during the first few decades. However, for example, almost all northern European populations (except Iceland, which also shows different patterns for M) experience extension over almost the entire observation period. This, however, could be misleading, since there are single downward jumps in these time series, which interrupt the overall increase in UB . Thus, if we would not admit for jumps in the time series (e.g., if we would fit a straight regression line to the time series during this period), the observed long-term trend might be not be significantly increasing any more. Also for most central European and Asia-Pacific populations, we observe long-term extension at least after the 1970s. In the United States and Canada, we have a two-decade period during the 1980s and 1990s where we observe a neutral trend. Thus, the North American cluster is homogeneous in this respect but shows special trends compared with the other populations.

Figure 4 shows the trends in the evolution of the degree of inequality DoI_0 for males. Until the middle of the 1950s, we mostly observe compression with few exceptions, in particular around the time of World War II. For the subsequent about two decades, there are several in parts even opposing trends (e.g., in the northern European cluster). After that, however, we find compression for almost all populations except eastern Europe until the end of the observation period.

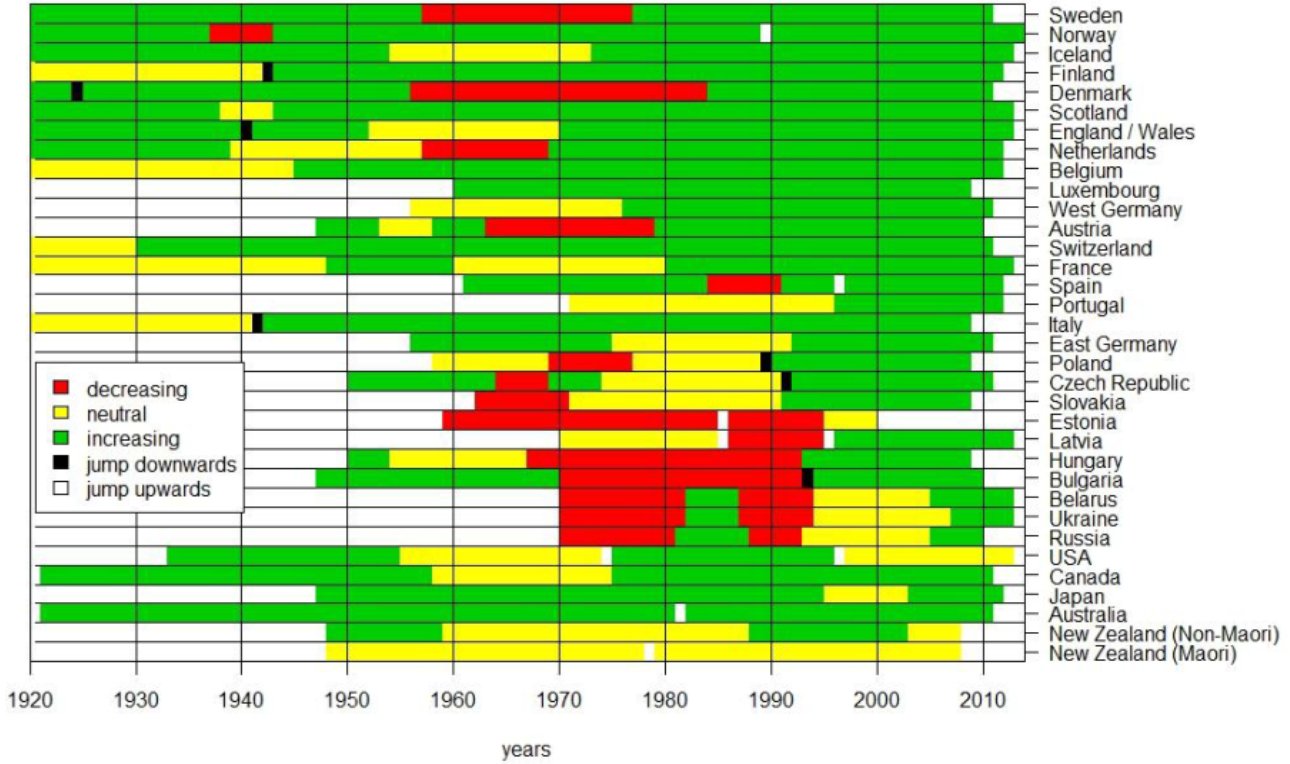


Figure 4: Trends in the Evolution of Degree of Inequality (DoI_0): Males, Starting Age 0

Also with respect to the trends in DoI_0 , the eastern European and the other populations have different trends, and we observe a similar heterogeneity in the direction of the trends, as for UB (decrease in the easternmost populations and neutral in the other populations until about 1990). For a few years during the 1980s, we observe an increasing trend for Belarus, Ukraine and Russia. Almost at the same time, we find an upward jump in the trends of Estonia and Latvia. Thus, for those populations, we have a short but significant period of compression (or an upward jump, respectively), even prior to the fall of the Soviet Union. For the easternmost European populations as well as outside Europe, we observe several neutral periods during the most recent decades. In particular, for the United States, we find almost two decades of neutral trend, which may compensate for the upward jump in the middle of the 1990s.

Finally, Figure 5 displays the time bars for the numbers of deaths at the modal age at death $d(M)_0$ for males. For most populations, the trends in $d(M)_0$ and DoI_0 are very similar. However, there are some exceptions. For example, in the United States, DoI_0 did not change significantly during the 2000s, but we observe diffusion (decreasing $d(M)_0$) during that period. Moreover, for

Austria, we observe decompression (decreasing DoI_0) during the 1960s and 1970s while there is concentration (increasing $d(M)_0$) at the same time. Also, in Austria and West Germany, we find a neutral period in the trend of $d(M)_0$ during parts of the 1980s and 1990s, which has no counterpart in the trends of DoI_0 . In Japan, we observe diffusion during the 1990s and the first half of the 2000s, while we only observe a shorter neutral period for DoI_0 . Such examples show that DoI and $d(M)$ are not fully correlated and that these statistics indicate different phenomena.

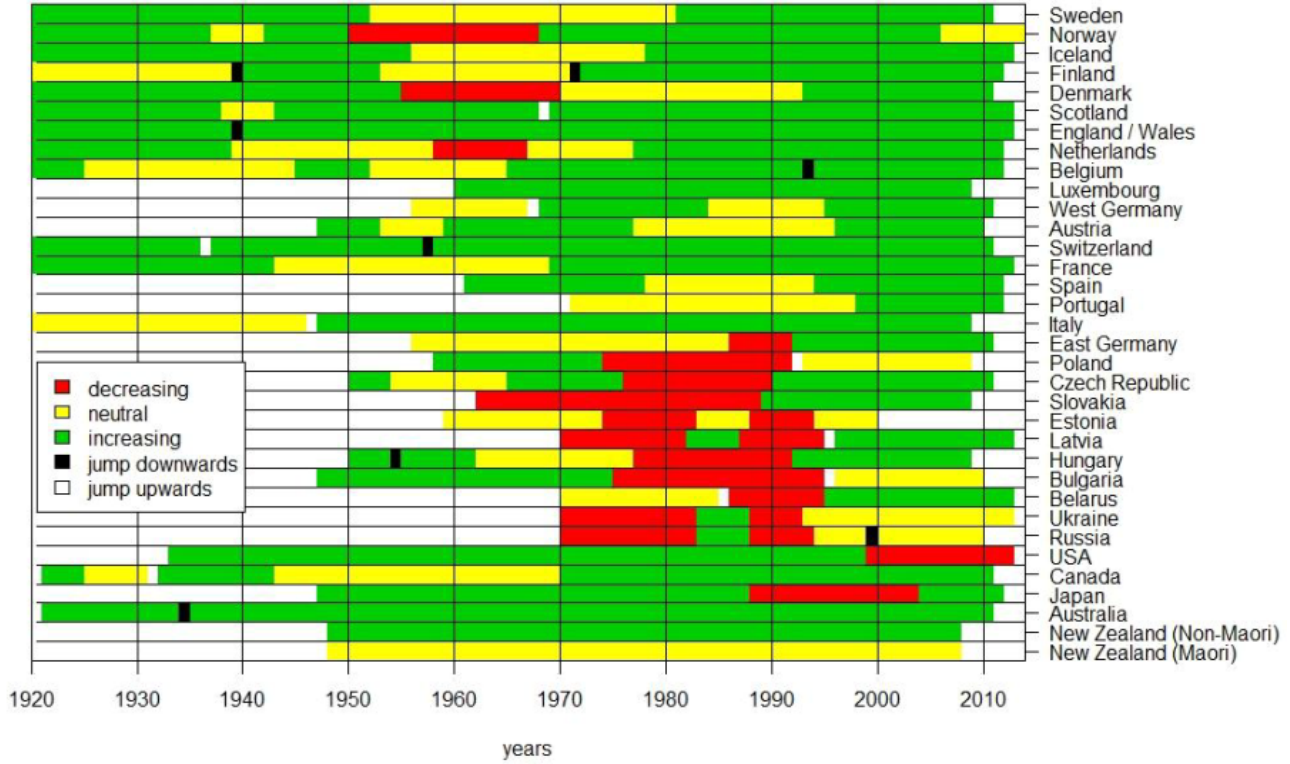


Figure 5: Trends in the Evolution of the Number of Deaths at Modal Age of Death ($d(M)_0$): Males, Starting Age 0

Figure 6 shows the relative similarity (RS) between each pair of the 34 populations (averaged over all four statistics) and provides a summary of the similarities and differences of the trends in the deaths curve's changes. Where there is a higher RS between two populations, the corresponding cell in this figure is brighter. Hence, in this figure, areas with white/yellow shades point to clusters with higher relative similarity, whereas areas with more reddish shades indicate larger differences. At first glance, we can detect bright areas in every corner of the figure. This means we have high relative similarities among the northwestern, central and southwestern European clusters, as well as the North American and Asia-Pacific clusters. Within these areas, we find some horizontal and vertical patterns. These patterns identify potential outliers regarding trends in the evolution of the deaths curve in their respective neighborhood. We observe such vertical and horizontal patterns, for example, for Iceland and the Maori population of New Zealand.

As expected, the relative similarity between the eastern European populations and the other populations is comparatively low, which again points to the difference in the trends between the eastern European cluster and the other populations. Though within this cluster, the relative similarity appears to be comparatively high, we also detect outliers here.

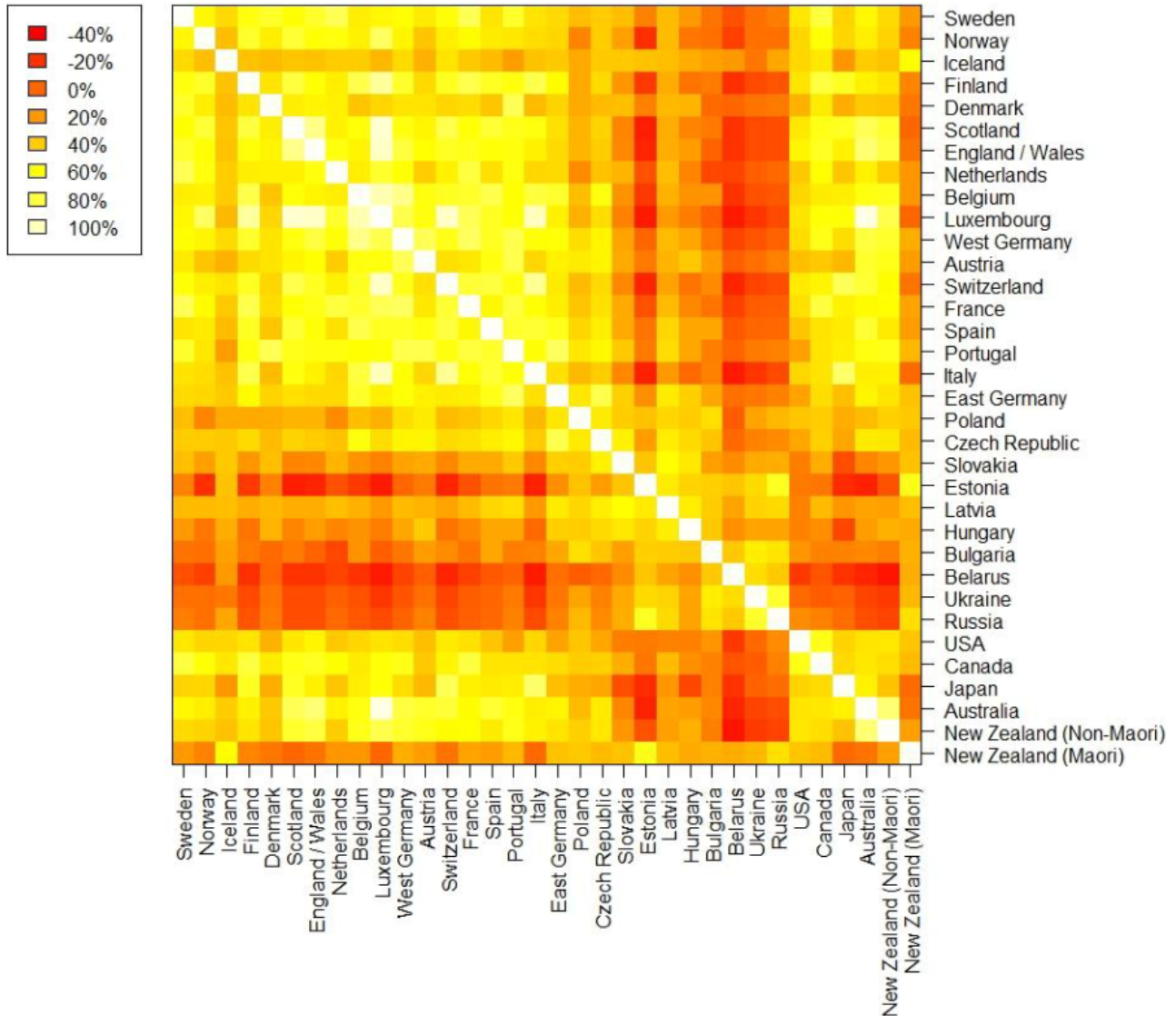


Figure 6: Average Relative Similarities for all Populations: Males, Starting Age 0

For the most recent decades, we can summarize three major observations. First, toward the end of the observation period for most populations, we observe an increase for the statistics measuring the position of the deaths curve (i.e., M and UB). This means that during that time period, the deaths curves shifted to the right, and at the same time, the deaths curves' support extended. Some eastern European populations adopted these trends in the 1990s, which leads to an almost global scenario of right-shifting mortality and extension at the same time. For the majority of the other populations, this trend started even in the 1970s and thus is rather long-term. The general trend toward increasing statistics can also be observed for the statistics measuring changes in the shape of the deaths curve. However, here we observe more exceptions,

both quantitatively and qualitatively, in particular in the eastern European cluster and outside Europe. In summary, the global reference scenario – derived from the trends we observed for males and apart from the exceptions mentioned above – for the most recent decades is (*right shift, extension, compression, concentration*).

Second, around 1970 (plus or minus about a decade), we observe a considerable accumulation of trend changes. For many populations, we observe a long-term increasing trend in many statistics after that. Especially for M , we observe a trend change around 1970 for many populations at the same time.

Finally, the trends in each of the four statistics for most eastern European populations experience a change shortly after 1990. Prior to this trend change, the trends of the eastern European and the other populations differ in general. Compared with the other populations, the eastern European cluster appears to be rather homogeneous, although the trends in the easternmost European populations and the other eastern European populations often are different.

5.2 Comparisons of the Trends in the Mortality Evolution

In this subsection, we analyze the mortality evolution of females with starting age 0 and the evolution of old-age mortality (i.e., starting age 60) for both sexes, and we compare the results with the results from Subsection 5.1. These analyses address two questions in particular: Are there significant differences in the trends of mortality evolution between females and males? And can we detect significant differences between the trends in the mortality evolution for the complete age range versus old-age mortality?

Females vs. Males

Figure 7 displays the time bars for the modal age at death M for females. At first glance, the patterns in the trends and trend changes for females are quite different from those of males. We observe much more long-term right-shifting mortality, there is no left-shifting mortality during the 1960s, and we cannot find the almost global trend change around 1970 that we observed for males.

However, though we do not observe a decrease in M during the 1960s for females, we can observe downward jumps for some populations at the end of the 1960s (e.g., England and Wales and Austria). This effect might be connected to the left shift for males at that time. Further, we can find a difference in the trends of some eastern European populations and the other populations, although this is only true for the easternmost populations. Moreover, the right shift during the most recent decades for most populations seems to be a unisex phenomenon. Thus, there are some similarities between the trends of females and males, but in general, the sex-related differences in the trends in M are considerable.

The trends in UB for females are shown in Figure 8. The differences between the trends of UB for females and males are immaterial for some populations (e.g., the southwestern European

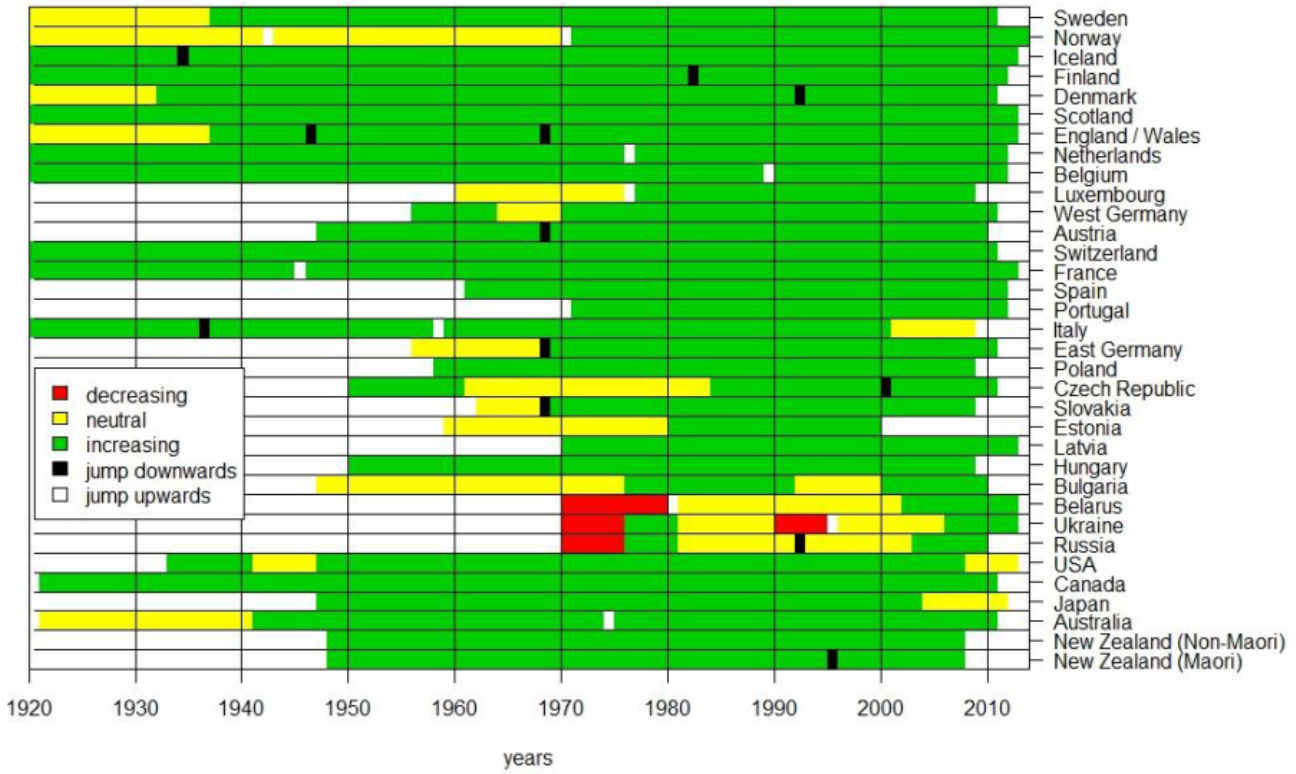


Figure 7: Trends in the Evolution of Modal Age at Death (M): Females, Starting Age 0

cluster). The absence of sex-specific differences in the trend of UB for these populations might point to a decreasing importance of the individual's sex for the mortality structure with increasing age. The difference between the trends for eastern European populations and most of the other populations is also apparent for females. Also, we observe a very long-term increase for females, which starts even earlier for some female populations than for the corresponding male populations (e.g., in Belgium, Austria and the non-Maori population of New Zealand). Some populations seem to follow this long-term extension, but these trends are interrupted by downward jumps (e.g., in Norway, Luxembourg and West Germany).

Figure 9 shows the time bars for the degrees of inequality (DoI_0) for females. The differences between females and males here have the same quality as the difference between sexes for UB . For some populations (e.g., Denmark), the sex-specific differences appear to be immaterial; for others (e.g., Austria), we observe a very long-term increase (i.e., compression) for females, where we observed periods of decrease (i.e. decompression) for males. Also, the sex-specific trends in DoI_0 are completely different in certain single populations (e.g., the Maori population of New Zealand). All in all, we can state two things: First, the supra-regional patterns we identified in the trends of DoI_0 for males seem to be unisex phenomena. Second, however, when looking at single countries, we sometimes find considerable differences between sexes.

The time bars for the trend in the numbers of deaths at the modal age at death $d(M)_0$ for females are displayed in Figure 10. As for the other statistics, we also observe more long-term increases in $d(M)_0$ (i.e., concentration) than for males. In particular, for most central, southwestern and

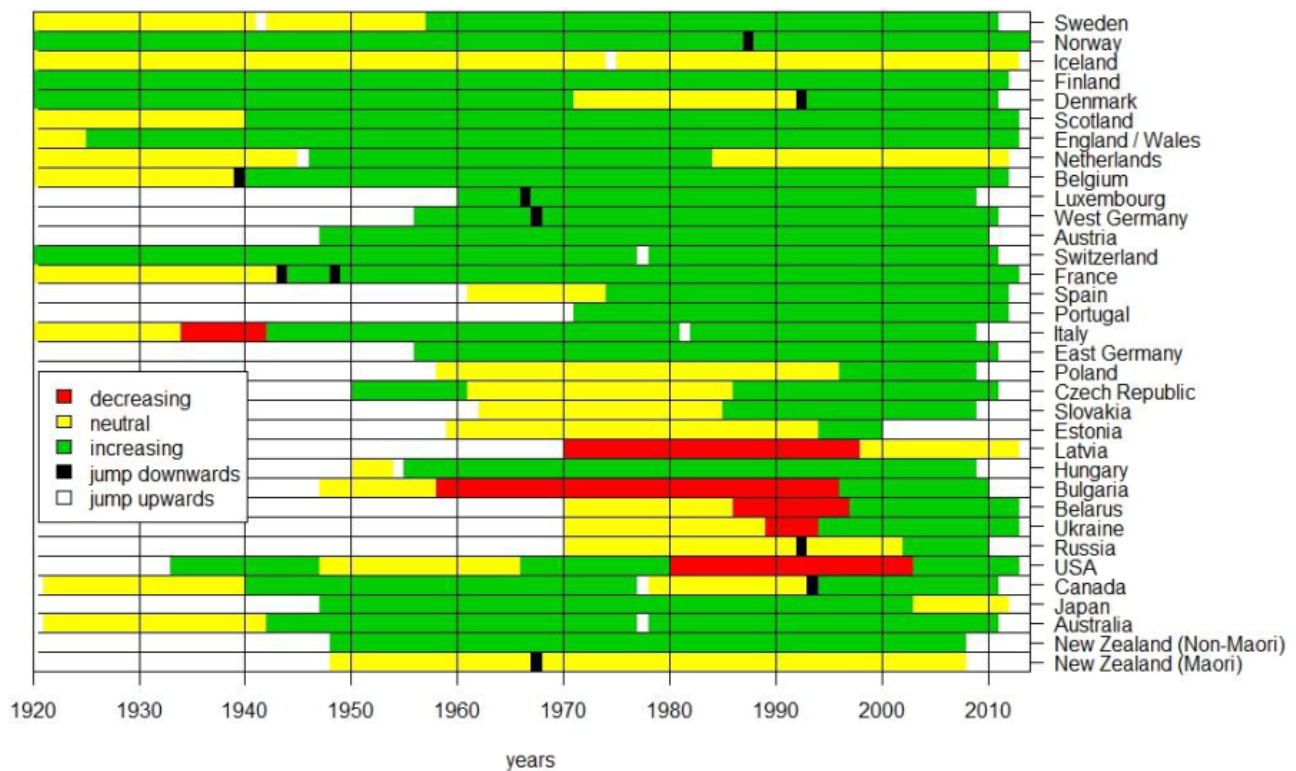


Figure 8: Trends in the Evolution of the Upper Bound of the Deaths Curve's Support (UB): Females, Starting Age 0

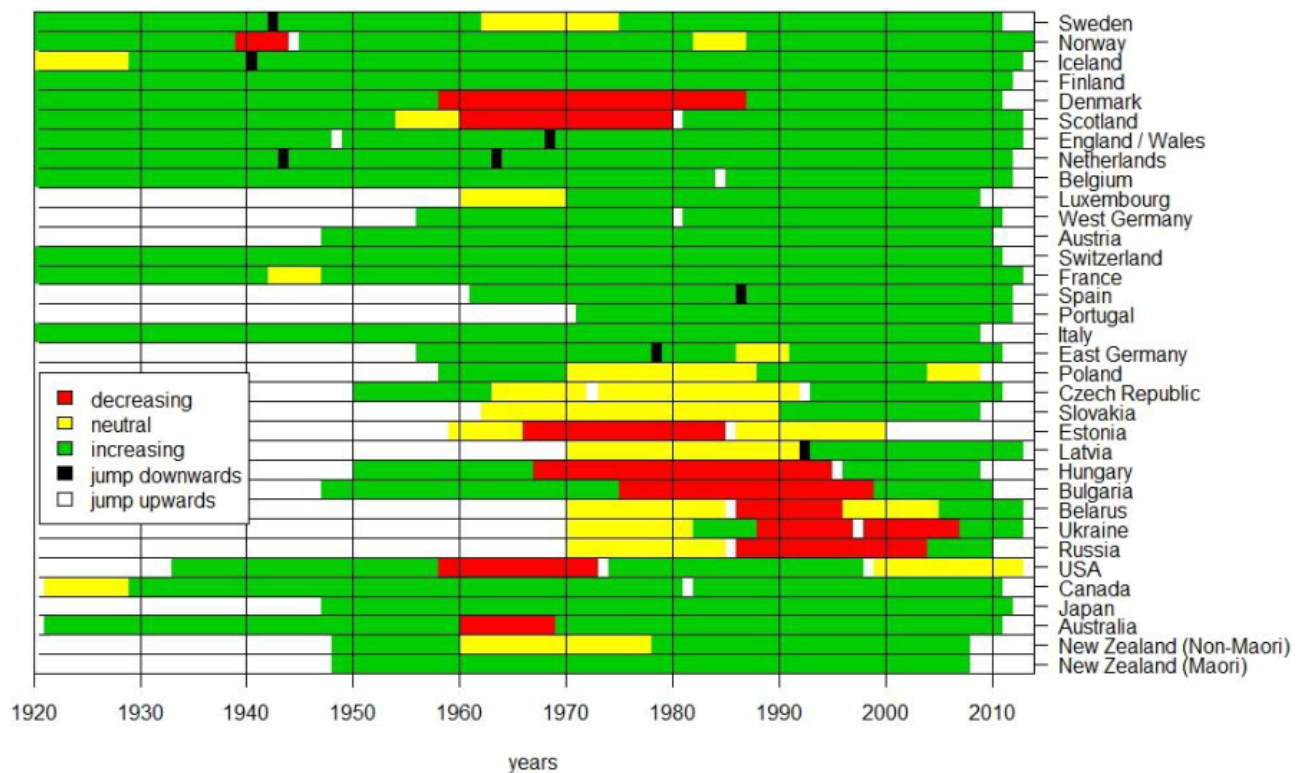


Figure 9: Trends in the Evolution of Degrees of Inequality (DoI_0): Females, Starting Age 0

eastern European populations, these differences between sexes become apparent. For the eastern European cluster in general, we observe less diffusion for females than for males. In contrast, for few northern European populations (e.g., Denmark), we observe diffusion between about 1960 and 1990, which we cannot find for males.

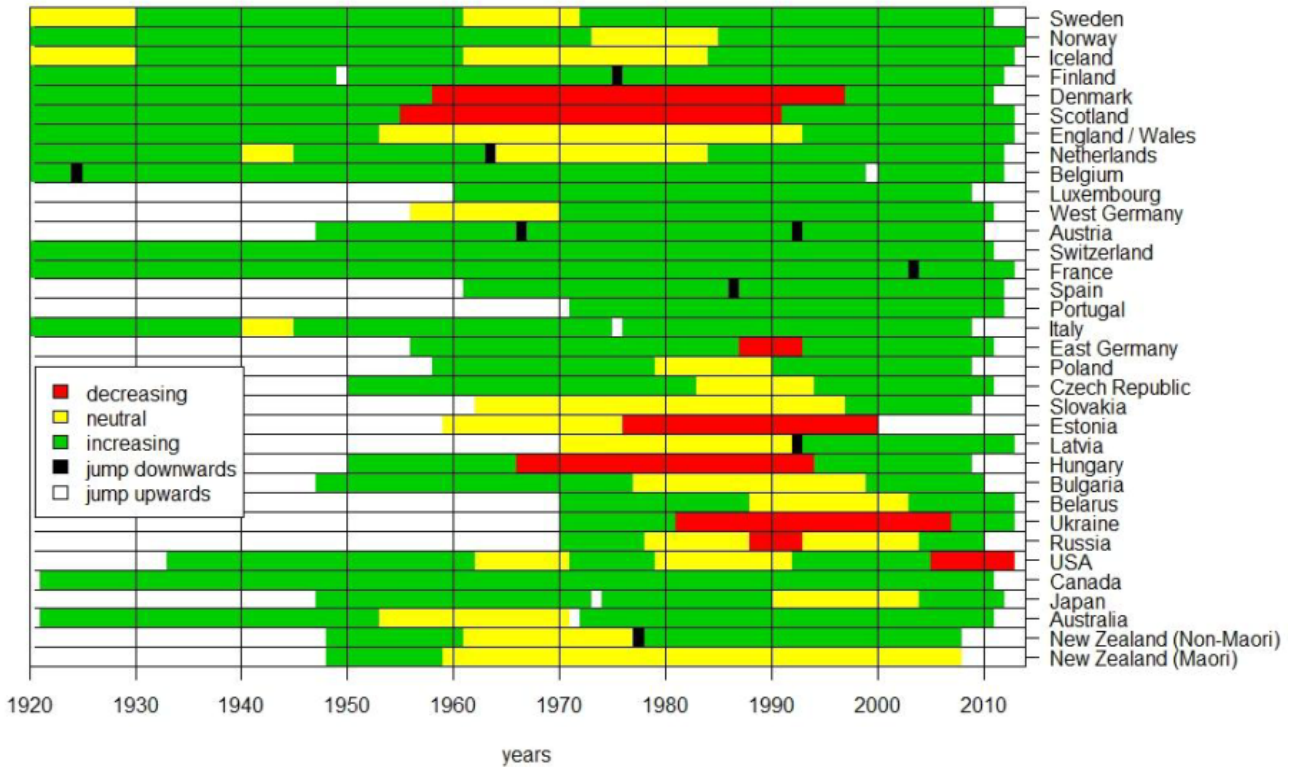


Figure 10: Trends in the Evolution of the Numbers of Deaths at the Modal Age at Death ($d(M)_0$): Females, Starting Age 0

Across statistics, we find increasing trends for females during the most recent years. Thus, then the trends of the deaths curve's evolution coincide for both sexes for most populations. However, there are apparent sex-specific differences in the trends of the deaths curve's evolution over the entire observation period. This is supported by Figure 11, which shows the relative similarity (RS) between sexes for the starting age 0.⁴ For the majority of the populations (27 out of 34), the relative similarity is smaller than 70 percent, and there are only four populations (Luxembourg, Switzerland, Italy and Japan) where the RS exceeds 80 percent. Not least, this finding illustrates the importance of a sex-specific consideration of mortality evolutions, as females and males have experienced different trends in the mortality evolution during the last 100 years. Thus, for example, for the usage of unisex models, it must be carefully checked whether the model is applicable and reasonable, depending on the question at hand.

Moreover, we can see from Figure 11 that the sex-related differences of the trends in the deaths curve's evolution are relatively high (i.e., low RS) for most of the eastern European populations.

⁴We first calculate the relative similarity between sexes (instead of populations) per statistic and population. After that, we determine the arithmetical mean between relative similarities for the four statistics per population. The formulae we use here are analogous to those described in Section 4.

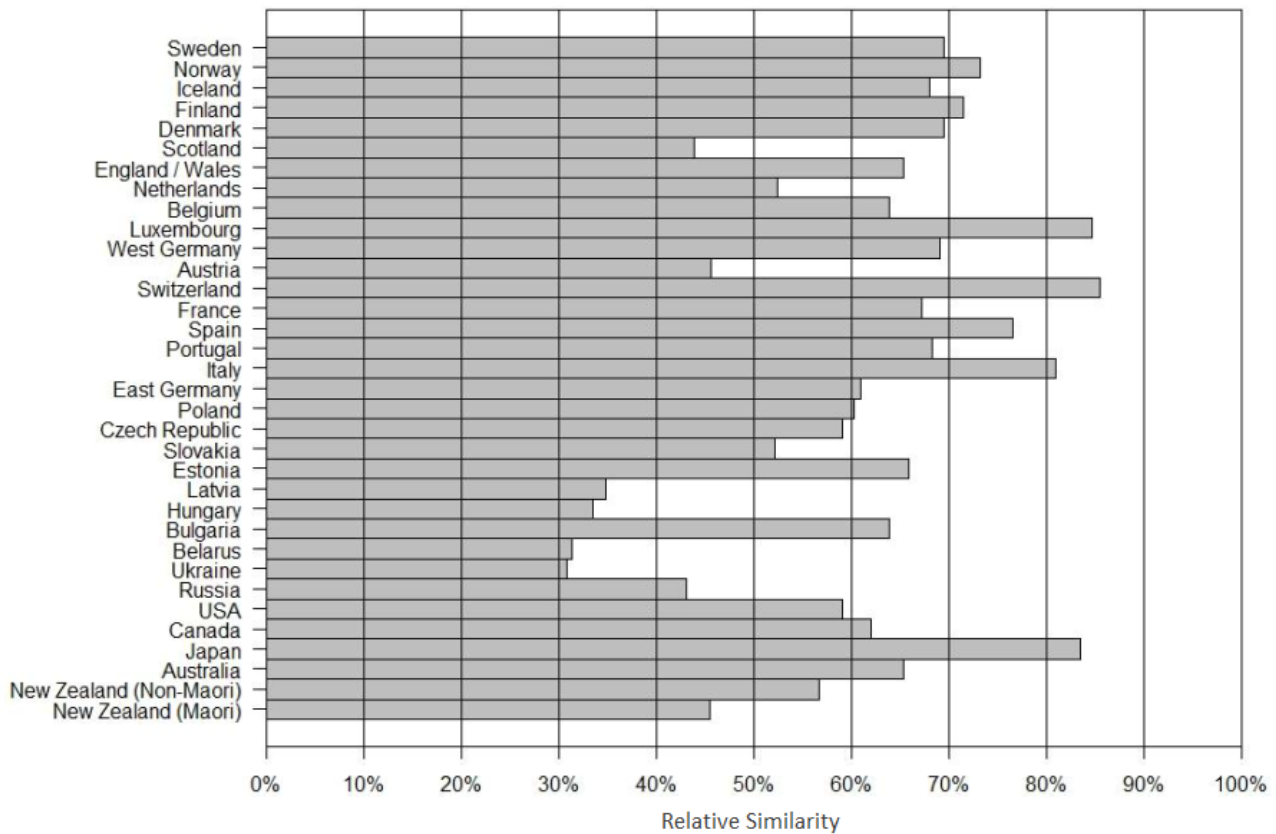


Figure 11: Relative Similarity (RS) Between Females and Males: Average of All Statistics, Starting Age 0

For this cluster, we can find only three populations whose RS exceeds 60 percent: East Germany, Estonia and Bulgaria. In comparison, the RS between sexes for central and southwestern European populations, for example, is rather high. Thus, we can identify regions where the sex-related differences in the deaths curve's evolution are usually larger than in other regions. Moreover, also in neighboring countries (e.g. Austria and Switzerland), the RS between sexes can differ significantly. Note, however, that a low RS for sexes does not necessarily imply a great difference in the level of mortality between both sexes.

To conclude the analysis of the sex-related similarities and differences of the trends in the deaths curve's evolutions, we can state three major findings:

1. Regarding left- or right-shifting mortality, the differences between females and males appear to be significant. However, during the most recent years, the direction of the trends for both females and males tends to converge toward right-shifting mortality.
2. Regarding compression or decompression, we cannot find any clear difference between females and males. Despite some sex-specific differences for some populations, the generally observed trends seem to be unisex phenomena.
3. Regarding extension or contraction and concentration or diffusion, respectively, there are

few sex-related differences. However, these differences cannot be regarded as immaterial.

Starting Age 0 vs. 60

Before we compare results for different age ranges, we recall that, as mentioned in Section 3, the modal age at death and the upper bound of the deaths curve's support are independent of the choice of the starting age. Thus, periods of left- or right-shifting mortality as well as extension or contraction coincide for both age ranges. Therefore, in this subsection, we only consider the trends in DoI_{60} and $d(M)_{60}$ and compare them with DoI_0 and $d(M)_0$, respectively.

Figure 12 displays the time bars for the degrees of inequality DoI_{60} for males. At the first glance, we can see that there are many more and much longer periods of neutral trends or even decreases in DoI_{60} than in DoI_0 . Indeed, for DoI_0 , we observed an increasing trend for many European populations after World War II, which seemed to be interrupted by relatively short periods of decompression or jumps. Though we observe a long-term compression for the retirees in a few populations (e.g., Sweden, the Netherlands, Belgium, Japan and Australia), the majority of the European populations do not experience an increase in DoI_{60} before the 1980s or even later.

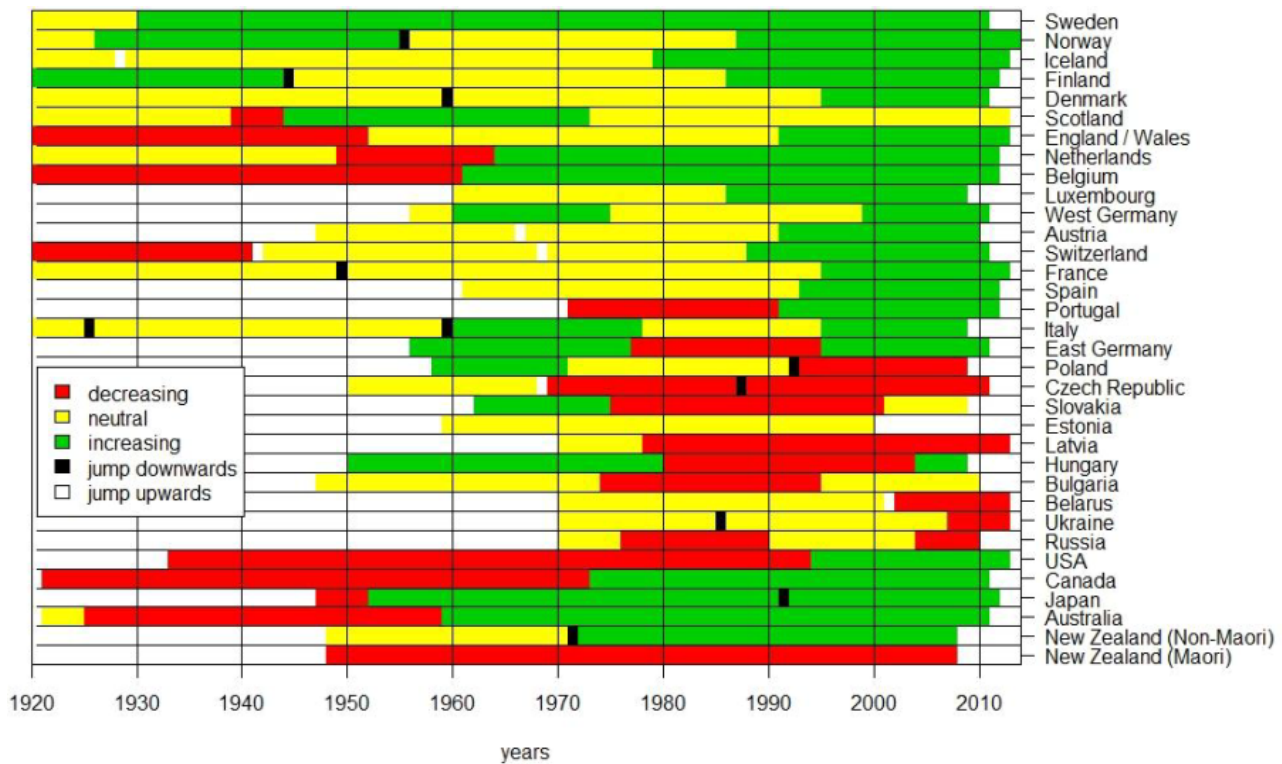


Figure 12: Trends in the Evolution of the Degree of Inequality (DoI_{60}): Males, Starting Age 60

For most of the eastern European populations, we do not even observe any period of compression, but rather long-term decompression for the starting age 60. Thus, we have no apparent trend change there during the 1990s. For the other European populations (except Italy), however, we do not observe any decompression after the early 1960s, but rather neutral trends. This

observation again underlines the differences in the trends of the mortality evolution between the eastern European populations and the other (European) populations even for the starting age 60.

For some non-European populations, we found some neutral periods in the most recent decades for the starting age 0. This also is not true for the starting age 60. Moreover, in the early decades of the observation period, we find compression for the starting age 0 and decompression for the starting age 60 for these populations. Thus, we can even observe opposing trends at that time for different starting ages.

Figure 13 shows the trends in the numbers of deaths in the modal age at death $d(M)_{60}$ for males. For $d(M)$, the differences between the starting ages are even bigger than for DoI . Before the late 1960s, concentration occurred only for single populations and for rather short time periods. Instead, we globally observe neutral trends or even diffusion for most populations. Starting in the late 1960s, we find a short trend of concentration for single populations across the world (e.g., England and Wales, Italy, East Germany, Bulgaria and the United States). However, during the 1980s, we have a rather comprehensive trend of diffusion. Only in the last few decades we observe a comprehensive trend of concentration for almost all non-eastern European populations. In contrast, for most eastern European populations, we observe a continued diffusion until the end of the observation period.

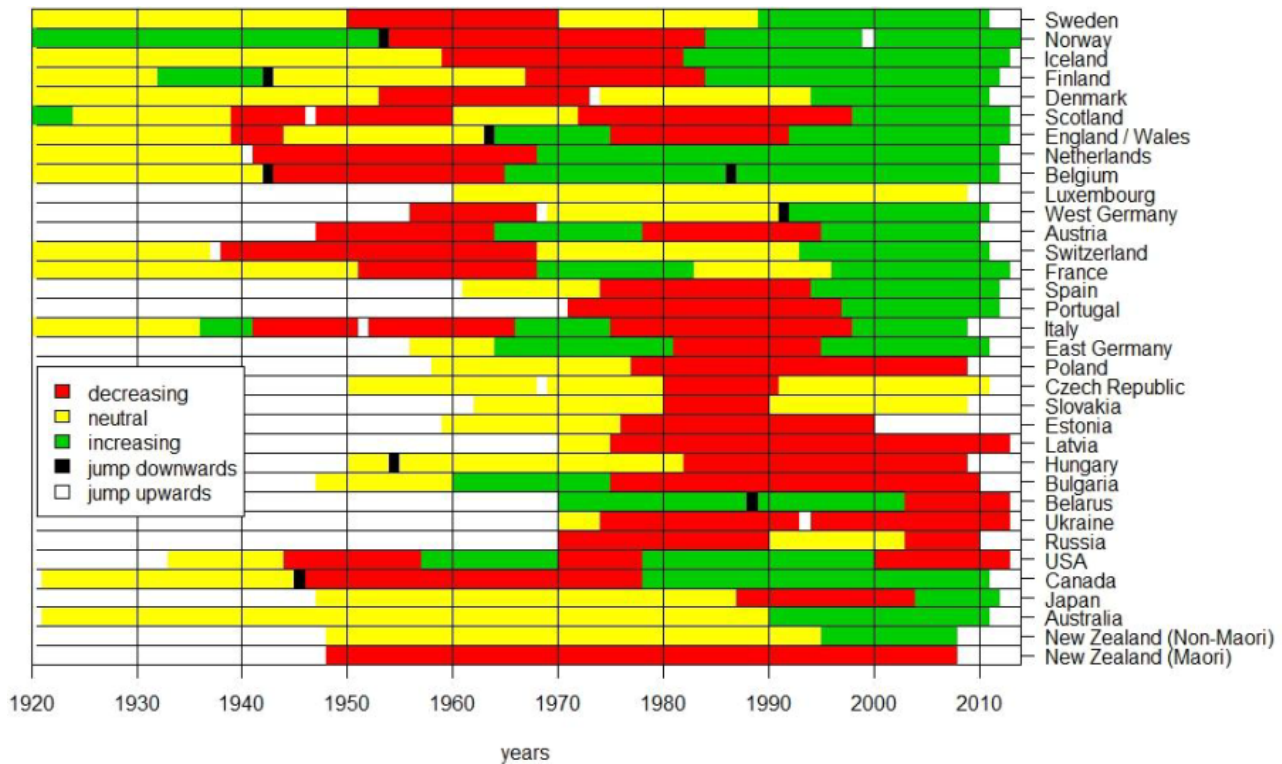


Figure 13: Trends in the Evolution of the Numbers of Deaths at the Modal Age at Death $d(M)_{60}$: Males, Starting Age 60

For females, the differences between DoI_0 and DoI_{60} and between $d(M)_0$ and $d(M)_{60}$, respec-

tively, are much smaller than for males (see Figures 16 and 17 in Appendix A for starting age 60). Since we cannot obtain any new insights here, we do not carry out the comparisons in detail. However, in what follows, we briefly discuss the similarity (or dissimilarity) of the trends in the deaths curve's evolution between the starting ages for both sexes.

Figures 14 and 15 show the relative similarities between the starting ages 0 and 60 for males and females, respectively, as an average over all four statistics over the entire observation period. We observe that the relative similarities between the different starting ages are much smaller for males than for females in general. For females, the trends in the deaths curve's evolution appear to be dominated by changes in the mortality at older ages, thus leading to similar trends for both starting ages. For males, in contrast, changes in the deaths curve below age 60 seem to have a more relevant impact, as they significantly change the trends observed only for ages above 60. This once again illustrates the differences in the trends of the mortality evolution between females and males.

In conclusion, we can state that there are considerable differences in the trends of DoI and $d(M)$ between the starting ages 0 and 60. These differences are more significant for males than for females.

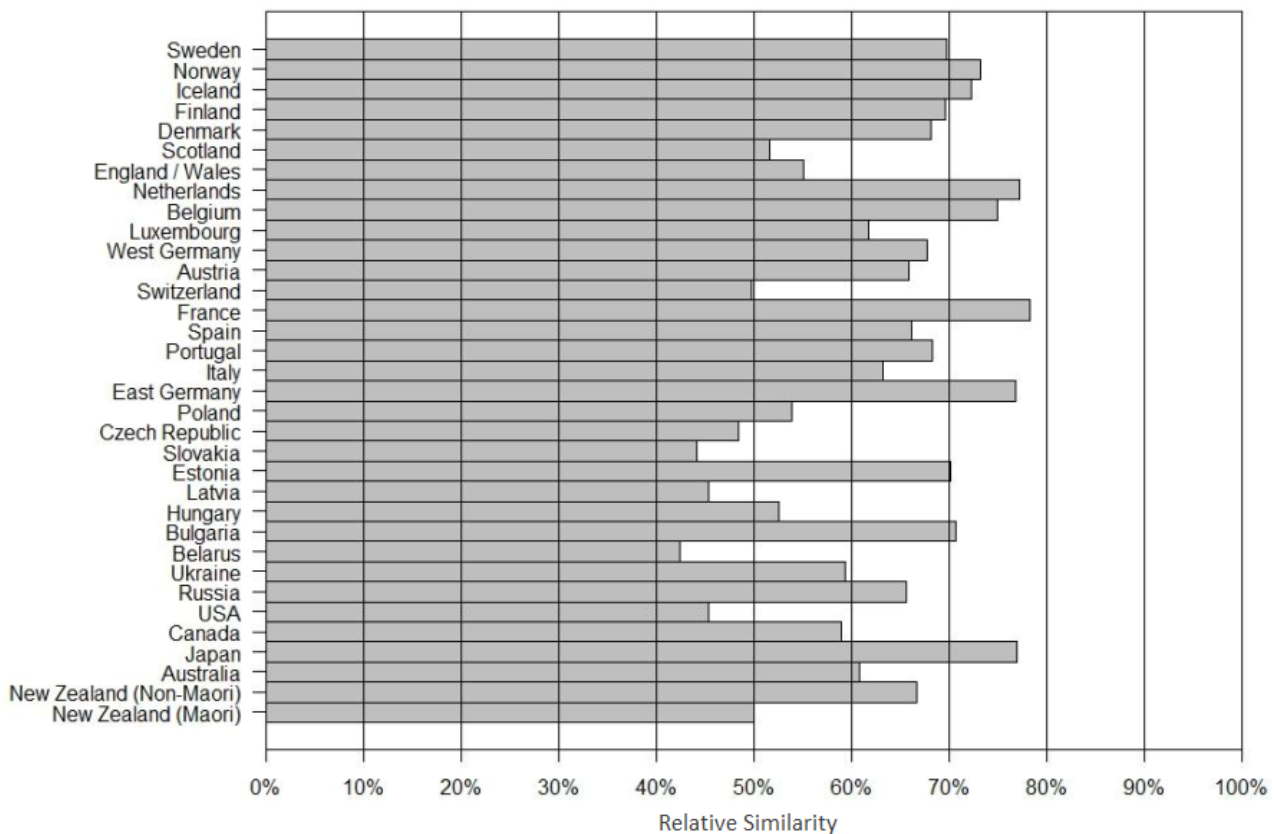


Figure 14: Relative Similarities Between Starting Ages 0 and 60: Average of All Statistics, Males

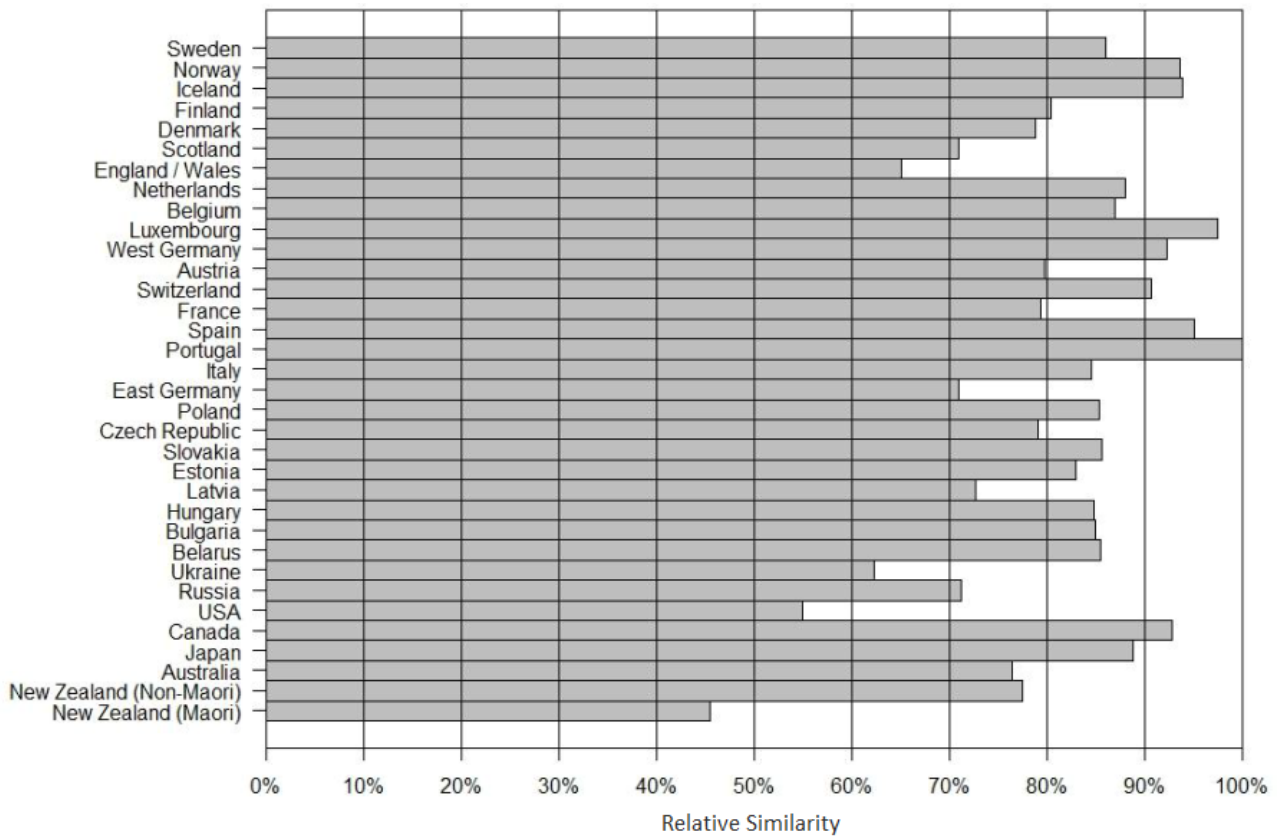


Figure 15: Relative Similarities Between Starting Ages 0 and 60: Average of All Statistics, Females

6 Conclusion

In this article, we discuss and apply a classification framework for mortality evolution patterns that was recently introduced by Börger et al. (2016). This framework consists of four components, which together uniquely define a scenario of mortality evolution. The modal age at death M and the upper bound of the deaths curve's support UB indicate changes in the position of the deaths curve, where M measures right- or left-shifting mortality and UB measures extension or contraction, respectively. The degree of inequality (DoI) and the number of deaths in the modal age at death $d(M)$ indicate changes in the shape of the deaths curve, of which DoI measures compression or decompression and $d(M)$ measures concentration or diffusion, respectively.

We calculate these statistics for 34 populations separately for males and females and for the starting ages 0 and 60. Thus, we obtain 544 time series, which we have to rework. For this purpose, we discuss several methods. In addition, we introduce the relative similarity (RS), which enables us to efficiently compare each pair of trend processes with three attainable states. In the discussion of the results, we first focus on the trends in the deaths curve's evolution of males with starting age 0. Here we obtained three major findings. First, during the most recent years, all four statistics increase for almost all populations, which generally is a long-term effect for all populations outside eastern Europe. Second, at least until the 1990s, the eastern

European populations experience different trends than the other populations. And third, during the 1960s, we observe a comprehensive decreasing trend, especially for M , and an inversion of this trend around 1970 for many populations at more or less the same time. This allows for two conclusions: there are supra-regional patterns in the trends of the change of the deaths curve, and there is no single global pattern for these trends. Both findings must be taken into account in future research on this topic.

By comparison of these findings to the trends in the deaths curve's evolution of females, we obtain different results. Regarding left- or right-shifting mortality (i.e., the trends in M), the differences between males and females are considerable. In contrast, the differences between females and males regarding compression or decompression (i.e., the trends in DoI) appear to be immaterial. Finally, for extension or contraction (i.e., the trends in UB) and concentration or diffusion (i.e., the trends in $d(M)$), respectively, we can find certain differences between sexes but also some similarities. However, the increase in all four statistics during the most recent decades seems to be an almost-global unisex phenomenon. Analyzing the RS between females and males per population, we indeed find large differences in the trends of the deaths curve's evolution between sexes. Moreover, we find regions where these sex-related differences are larger or smaller than in other regions. For instance, in eastern Europe, the RS between sexes in general is smaller (meaning larger differences) than for other populations.

The trends in DoI and in $d(M)$, respectively, for the age range beyond age 60 in general show fewer long-term increasing trends than for the complete age range. Consequently, we observe multiple periods where these statistics show opposing trends for the two starting ages for males. In contrast, for females, the differences of the trends in these two statistics between the starting ages are rather small. We therefore analyze the RS between the starting ages for both sexes and find that these similarities indeed are smaller for males than for females. This repeatedly illustrates the need for sex-specific analyses of mortality structures and sex-specific mortality modeling.

The insights of the present analysis especially can be helpful for the application of mortality models. There, the first step, of course, is to find a suitable model. The second step, however, is the calibration of the model to historical data. Conventional mortality models do not incorporate trend changes. Therefore, such models should be calibrated to periods of time series where the change of the deaths curve follows a constant trend. With this analysis, we can exactly refer to such periods. Moreover, for the application of more sophisticated models (e.g. multi-population mortality models), this analysis provides findings about coherent populations. For example, is it reasonable to apply a multi-population mortality model to data from Belgium, the Netherlands and Luxembourg in order to obtain more reliable mortality rates for the comparatively small population of Luxembourg? And if that appears to be reasonable, is this still true if we considered sex-related mortality and/or old-age mortality?

Moreover, the methods introduced here can be very useful for governments, life insurers and pension funds whenever trends in the mortality evolution need to be analyzed in the context of

surrounding, dependent or superior populations. For instance, these methods enable life insurers to test the trends in the mortality structure of their portfolios for consistency with the country's general public.

Not least, these findings hopefully will initiate further research in more explanatory disciplines, such as epidemiology, medicine, biology, sociology and demography. To our knowledge, for example, the sex-related differences in the mortality structure (and consequently also in the trends of the mortality evolution) are still an open field of research. Moreover, from our point of view, the differences in the trends of the deaths curve's evolution between the starting ages, especially for males, are worth further research.

Acknowledgements

The author wants to thank Matthias Börger and Jochen Ruß for their support and their guidance and also Johannes Schupp for helpful discussions regarding the methods used for trend detection.

References

- Börger, M., Genz, M., and Ruß, J. (2016). *Extension, Compression, and Beyond: A Unique Classification System for Mortality Evolution Patterns*. Preprint, Ulm University and Institute for Finance- and Actuarial Science, Ulm. URL: http://www.ifa-ulm.de/fileadmin/user_upload/download/forschung/2016_ifa_Boerger-Genz-Russ_Extension-Compression-and-Beyond.pdf.
- Canudas-Romo, V. (2008). The modal age at death and the shifting mortality hypothesis. *Demographic Research* 19(30): 1179–1204. DOI: 10.4054/DemRes.2008.19.30.
- Cheung, S.L.K., Robine, J.-M., Paccaud, F., and Marazzi, A. (2009). Dissecting the Compression of Mortality in Switzerland, 1876–2005. *Demographic Research* 21(19): 569–598.
- Debón, A., Martínez-Ruiz, F., and Montes, F. (2011). *Temporal evolution of some mortality indicators: Application to Spanish data*. Paper presented at the 2011 Living to 100 Society of Actuaries International Symposium, Orlando, FL. URL: <https://www.soa.org/essays-monographs/2011-living-to-100/mono-li11-1b-debon.pdf>.
- Edwards, R.D. (2011). Changes in World Inequality in Length of Life: 1970–2000. *Population and Development Review* 37(3): 499–528. DOI: 10.1111/j.1728-4457.2011.00432.x.
- 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. DOI: 10.1111/j.1728-4457.2005.00092.x.
- HMD (2015). *Human Mortality Database*. University of California, Berkeley, and Max Planck Institute for Demographic Research. URL: www.mortality.org.

- Kannisto, V. (2000). Measuring the compression of mortality. *Demographic Research* 3(6). DOI: 10.4054/DemRes.2000.3.6.
- Kannisto, V. (2001). Mode and Dispersion of the Length of Life. *Population: An English Selection* 13(1): 159–171. URL: <http://www.jstor.org/stable/3030264>.
- Nusselder, W.J. and Mackenbach, J.P. (1996). Rectangularization of the Survival Curve in the Netherlands, 1950-1992. *The Gerontologist* 36(6): 773–782. DOI: 10.1093/geront/36.6.773.
- Ouellette, N. and Bourbeau, R. (2011). *Recent adult mortality trends in Canada, the United States and other low mortality countries*. Paper presented at the Living to 100 Society of Actuaries International Symposium, Orlando, FL. URL: <https://www.soa.org/essays-monographs/2011-living-to-100/mono-li11-4b-ouellette.pdf>.
- Robine, J.-M., Cheung, S.L.K., Horiuchi, S., and Thatcher, A. R. (2008). *Is there a limit to the compression of mortality?* Paper presented at the Living to 100 and Beyond Society of Actuaries Symposium Orlando, FL. URL: <https://www.soa.org/essays-monographs/2008-living-to-100/mono-li08-03-cheung.pdf>.
- Thatcher, A.R., Cheung, S.L.K., Horiuchi, S., and Robine, J.-M. (2010). The compression of deaths above the mode. *Demographic Research* 22(17): 505–538. DOI: 10.4054/DemRes.2010.22.17.
- Viner, R.M., Coffey, C., Mathers, C., Bloem, P., Costello, A., Santelli, J., and Patton, G.C. (2011). 50-Year Mortality Trends in Children and Young People: A Study of 50 Low-Income, Middle-Income, and High-Income Countries. *Lancet* 377(9772): 1162–1174. DOI: 10.1016/S0140-6736(11)60106-2.

Appendices

A Trends in the Mortality Evolution of Females With Starting Age 60

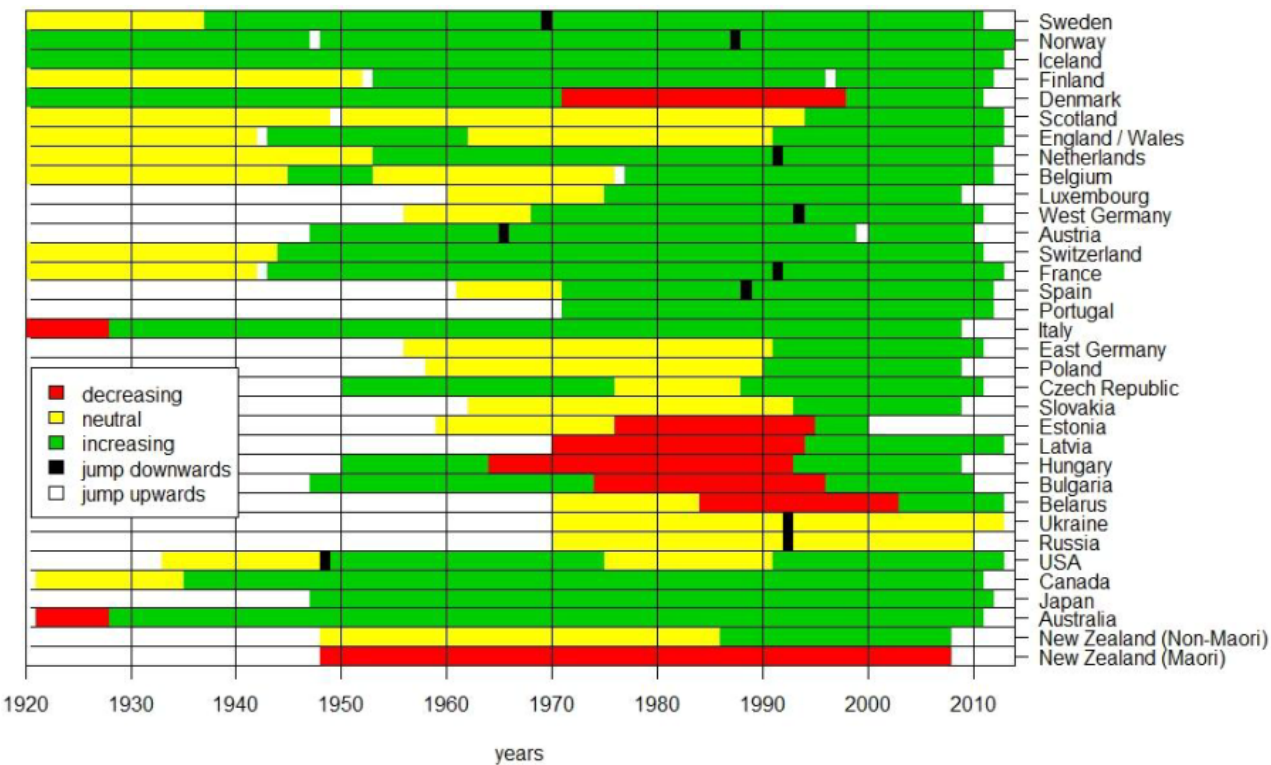


Figure 16: Trends in the Evolution of Degrees of Inequality (DoI_{60}): Females, Starting Age 60

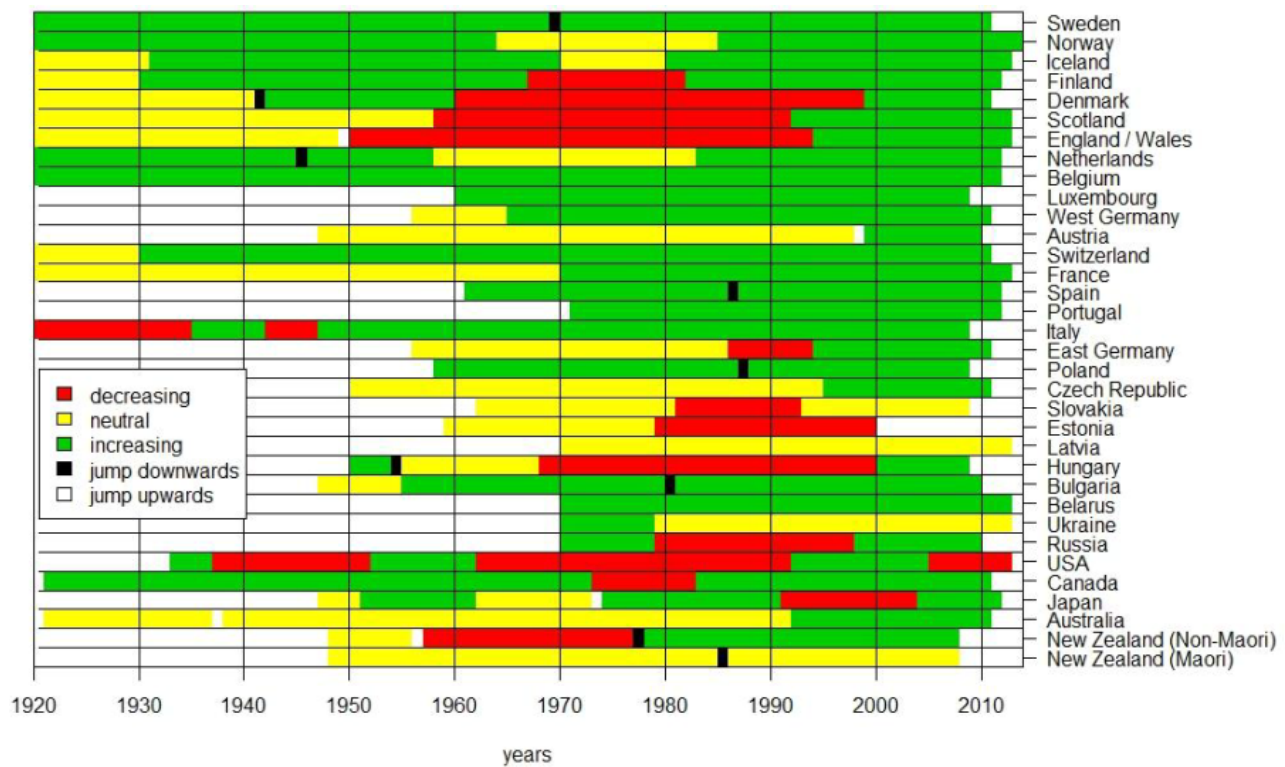


Figure 17: Trends in the Evolution of the Numbers of Deaths at the Modal Age at Death $d(M)_{60}$: Females, Starting Age 60

3 The Myth of Immortality: An Analysis of the Maximum Lifespan of US Females

Source:

Feifel, J., Genz, M., and Pauly, M. (2018). The Myth of Immortality: An Analysis of the Maximum Lifespan of US Females. Working Paper.

The Myth of Immortality: An Analysis of the Maximum Lifespan of US Females

Jan Feifel*

Martin Genz[†]Markus Pauly[‡]

Abstract

Questions related to the existence and specification of a limit to human lifespan lead to heated discussions in several scientific fields such as biology, demography, medicine, or actuarial sciences. In the present paper, we contribute to this discussion from a statistical point of view. To this end, we use combined mortality data of US females obtained from the International Database on Longevity as well as the Human Mortality Database. The use of old-age mortality data typically raises two issues: sparse information on the old ages and censored observations. Up to the present paper, this censoring issue mostly has been ignored in previous investigations on the maximum human lifespan. We address this accordingly by combining sub-sampling and cross-validation techniques with recent results on censored extreme value theory. As the main result, we estimate the maximum lifespan of US females for moving intervals of nine calendar years between 1980 and 2003.

Keywords: Censoring, Extreme value theory, Subsampling, Human Mortality Database, Supercentenarians, Old-age mortality

*Institut für Statistik, Universität Ulm, Helmholtzstraße 20, 89081 Ulm, Germany, email: jan.feifel@uni-ulm.de (corresponding author)

[†]Institut für Finanz- und Aktuarwissenschaften (ifa) and Institut für Versicherungswissenschaften, Universität Ulm, Lise-Meitner-Straße 14, 89081 Ulm, Germany, email: m.genz@ifa-ulm.de

[‡]Institut für Statistik, Universität Ulm, Helmholtzstraße 20, 89081 Ulm, Germany, email: markus.pauly@uni-ulm.de

1 Introduction

The dream of immortality is not only subject to legends, science fiction, and fairy tales but also has attracted the attention of researchers from many different fields of science. Who has not yet dreamed about the possibility to extend the individual lifespan to ages that have never been reached before? Following, for example, de Grey (2003) immortality may become reality in near future. On the other hand, there may be a natural limit to human life which probably is just not yet in sight. Or maybe we have already reached this limit. For example Dong et al. (2016) found "evidence for a limit to human lifespan" which is supported by a comprehensive data analysis. However, they base their claims on statistics like the maximum reported age at death which completely ignores people who are still alive at highest ages. Others like, e.g., Fries (1980) found natural constraints to longevity while a few decades later reality exceeded their forecasts. Recently, Rootzén and Zholud (2017) concluded that "human life is unbounded but short". However, their statistical analysis is based on small sample sizes and thus may lack power, see also Section 3 below. In any case, as pointed out by Weon and Je (2009), "the existence of maximum human lifespan remains a puzzle in aging research".

From a statistical point of view, all of these questions are related to the support of the density function of the underlying age distribution at death (the so-called deaths curve) and lead to investigations about its right endpoint. The current paper studies them for US females by means of modern statistical methods from extreme value theory (EVT). Since the deaths curve typically changes over time (see e.g. Börger et al., 2018) its right-endpoint also changes over time (see e.g. Cohen and Oppenheim, 2012; Wilmoth et al., 2000). We, therefore, study the evolution of the deaths curve's right tail over moving time periods.

For this analysis, we use data from the International Database on Longevity (IDL, 2017) and the Human Mortality Database (HMD, 2015). To avoid gender effects and to ensure an acceptable data quality and quantity we focus on data of US females. The HMD is one of the largest publicly available databases on mortality in the world. However, many observations in the HMD are right-censored, particularly at the age of 110. This prohibits a gain of information about the right tail behavior of the deaths curve. However, it is well-known that there have been humans which survived the age of 110 (see e.g. Robine and Allard (1998) for the famous case of Jeanne Calment) and thus the HMD data alone is not suitable for any EVT analysis. In contrast, the IDL is more informative regarding the death counts of the highest ages: It provides the exact age and time of death of so-called supercentenarians, i.e. individuals who at least reached their 110th birthday. The IDL covers death counts of 309 US female supercentenarians in the time span 1980–2003. This allows for a statistical analysis in the classical EVT framework (see de Haan and Ferreira, 2006; Falk et al., 2010; Reiss and Thomas, 2007; Resnick, 2013) which for example has been applied to estimate the maximum attainable age for the Netherlands and Belgium by Aarsen and de Haan (1994) and Gbari et al. (2017) and Einmahl et al. (2019), as well as for Canada and Japan by Watts et al. (2006). Since such an analysis has not been

carried out for the US alone, we start our investigation with the IDL. Since 309 is a rather small number for quite a large time horizon we increase the inferential accuracy by subsequently combining the datasets of the IDL and the HMD. We then study this combined data set (CDS) over moving time intervals, each of 9 years length. In this way, we evaluate the progression of the lifespan distribution from 1980–2003. As a by-product, working at smaller time intervals is more convincing with respect to the usual iid assumption underlying EVT methods. Moreover, it allows for an adequate treatment of the present truncation by incorporating survivors of each time interval as right-censored observations. To take care of these issues in an adequate way we employ methods from *censored* EVT following Einmahl et al. (2008), see also Gomes and Neves (2011) and Worms and Worms (2014) or Gomes and Guillou (2015) for other censored EVT approaches.

To our knowledge, this is the first paper on estimating the maximum lifespan in a population that

- employs methods from censored EVT and
- includes old-aged survivors in the analysis
- while adequately dealing with the underlying censoring issue.

The remainder of this paper is organized as follows: Section 2 gives a detailed description of two employed databases and discuss their structure. Moreover, in this section, we describe how we construct a combined data set which exploits the advantages of both databases and we make a quantitative analysis of this data set. Thereafter, in Section 3 we introduce and apply classical EVT methods to analyze the IDL data alone and illustrate our findings. As this leads to results which are neither statistically significant nor rigorous, we analyze the combined data set by means of censored EVT in Section 4. The results are illustrated and discussed in Section 5 and the paper closes with some conclusions and an outlook in Section 6.

2 Data Description

In this section we cover different aspects of the data used for our research. Unfortunately, there is no publicly available database that includes complete mortality and survival data for a sufficiently long history while providing the complete age range for all covered populations. Nevertheless, we admit that there are databases giving detailed information on survival as well as the time and age at death. But often these databases are not publicly available and only cover single populations. In particular, this complicates the comparison of results. We therefore decided to analyze publicly available data. In what follows, we first describe some characteristics of the databases underlying our investigations. We then describe how to exploit their advantages, and finally make a quantitative analysis for the estimation of the deaths curve’s right endpoint.

2.1 Characteristics of the HMD and IDL Databases

Our research is based on data from two different databases: the International Database on Longevity (IDL, 2017) and the Human Mortality Database (HMD, 2015). As outlined in the introduction we focus on US females to have a rather homogeneous population and a sufficiently large sample size.

The **IDL** contains data for 15 industrialized countries. Each of its records relates to one individual dying after its 110th birthday, where the individual's age at death is given exactly in days from their birthday and it is supposed to be free of age ascertainment bias (see Maier et al., 2010). For the US the IDL provides the age and time of death of 309 females dying at the age of 110 or beyond between 1980 and 2003. The oldest US female recorded in the IDL reached an age of 43560 days (i.e. 119 years and 97 days). These figures illustrate a major issue of the IDL: Its exactness and lack of ascertainment bias inevitably lead to a reduction in the number of observations. Thus, the data may underestimate the count of supercentenarians dying in the years between 1980 and 2003. Maier et al. (2010) explain the process of validation of the IDL data and conclude that the IDL fits the profile of the US female population well in general though it probably underestimates the absolute number of people dying in this age range. A further disadvantage of the IDL is the lack of survival data: We only know the time and age of death of some very old US females but there might be some individuals still alive in similar or even higher ages. Moreover, the IDL stopped collecting data after 2003.

In contrast to the IDL the **HMD** covers mortality and population size (i.e. survival) data of 38 different countries, where death counts are obtained in up to three dimensions: age at death, calendar year of death, and calendar year of birth. The HMD edits the input data to obtain a homogeneous structure over calendar years per population but also across populations. To reduce complexity we pass on the information of the year of birth and only employ the single-year and single-age death counts and population size (survivor) data. Unfortunately the HMD only supplies age and time discrete data, i.e. we only know how many members of a population die during the (integer) year t and in the (integer) age x , but we neither know the exact date of death, nor their exact age at death. We therefore follow Wilmoth et al. (2017) and assume the date of death as well as the date of birth during a calendar year to be uniformly distributed. In this way we can redistribute the death counts per calendar year and thus work with a non-integer time and age at death. Similarly, we can derive an age structure of the population at the end of a calendar year.

For our analysis, we extracted HMD population size data for US females for the time horizon from 1980 – 2003. In sum it includes almost 26 million deaths over all ages whereof 1853 females are deterministically right-censored at 110, i.e. died at the age of 110 or beyond. This means that the HMD only contains single age mortality data up to the age of 109 and a cumulative entry $110+$ thereafter. This constricts the estimation of the maximum attainable age. For this reason, we additionally merge the IDL into the HMD for the analysis. This combination exploits

the quantity of the HMD and the quality of the IDL data.

2.2 Combining the Data – the CDS

The input data of the HMD for US females builds on several official population and health statistics of the United States. In this sense, the data is likely to be complete with regard to the US population in general. Thus, it is reasonable to assume the IDL population being a sub-population of the HMD population. With this assumption we combine them as follows: First, as we are interested in old-age mortality, we only include the HMD data (death counts and survivors) beyond the age of 97. The inclusion of survivors thereby adds right-censored observations to our time interval (1980–2003) of interest. Second, for each calendar year, we randomly delete as many deaths counts from the HMD deaths counts' entry 110+ as we observe deaths in this age range in the IDL. Finally, we derive the number of survivors of each calendar year from the data of the IDL. This is possible as we know the exact age and time of death of each individual recorded in the IDL. We randomly delete as many survivors of the HMD as we have survivors implicitly given by the IDL for each calendar year and each age in the age span beyond 97. This merges the IDL into the HMD and avoids double counts in the resulting data set which we denote as **CDS** (combined data set). Later on, we analyze the CDS with censored EVT techniques introduced in Einmahl et al. (2008) which require the assumption of iid observations. However, since the structure of mortality typically changes over time, this presumption may not be reasonable for the complete time horizon of the CDS of 24 calendar years. Since evaluating on an annual basis would lead to rather small sample sizes, particularly with respect to the proportion of the IDL data in the CDS per period, we found a compromise: We run our analysis on time windows of nine adjacent calendar years. In each step, these windows are shifted by one-year, i.e. we start with the window between 1980 and 1988, proceed with the window between 1981 and 1989, and so on, ending with the window between 1995 and 2003. We treat the records for fixed windows as realizations from independent and identically distributed random variables that may be subject to independent right-censoring. The assumption on identically distributed random variables on each window appears to be more convincing compared to the entire period 1980–2003. Moreover, analyzing moving windows enables us to evaluate the evolution of the deaths curve's right endpoint over time.

2.3 Quantitative Analysis of the Databases

To get a picture of the data bases we compare the values of the IDL and the HMD (for the records beyond 97 years; without double counts) in Table 1. For each window the absolute numbers of death counts and censored observations are shown. For the HMD the latter is split into censored individual surviving the time window (i.e. the subject died at a later time point) and censoring due to HMD-specific missing information for ages beyond 110 (110+ entries)

Window	HMD			IDL		
	Observed death counts	Censored observations due to 110+ entries survivors		Observed death counts	% of all CDS deaths in the range 110+	Censored due to survival
1980-88	184752	624	85577	72	10.34	14
1981-89	198520	616	90676	79	11.37	14
1982-90	212466	624	95641	87	12.24	18
1983-91	227203	614	101020	84	12.03	32
1984-92	240896	617	105983	102	14.19	24
1985-93	257341	656	108517	118	15.25	23
1986-94	272599	657	110999	129	16.41	22
1987-95	287816	666	112084	135	16.85	28
1988-96	301362	675	115006	145	17.68	24
1989-97	314376	706	118120	152	17.72	30
1990-98	327949	732	119849	157	17.66	36
1991-99	342915	729	121968	168	18.73	45
1992-00	356859	752	123306	181	19.4	27
1993-01	370065	760	124548	173	18.54	11
1994-02	380731	751	125481	159	17.47	3
1995-03	391504	746	127637	142	15.99	0

Table 1: Shown are the window-specific sample sizes of the censored and uncensored observations in HMD and IDL together with the portion of the IDL in the higher age entries (older than 110 years) of the CDS.

which is responsible for less than 1% of the censored observations. As the IDL collects the exact age at death, observations from the IDL are only censored due to survival of the specific time window. This quantity decreases in the last four windows since the IDL only covers data up to the year 2003. As a small caveat, we note that this issue may lead to a slight under-estimation of the right endpoint in the last four windows. In addition to the absolute death counts in the IDL also its portion in the number of death counts in the age range beyond 110 from the CDS is shown. This portion ranges between 10% and 20%, which means that for each window at least one of ten US female supercentenarians dying is an uncensored observation in the CDS. The portion of the IDL data in all deaths of the CDS per window is much smaller (between about 0.03% and 0.05%) but this is mainly due to the high number of observations from the HMD in the age range between 97 and 109. Finally, the corresponding CDS death counts are obtained by summing up column two and five, whereas for the censoring counts columns three, four and seven are required. It is apparent that these values reveal an increasing trend over the windows. This is in line with the findings of, e.g. Robine and Paccaud (2005), who found that the number of old Swiss people increases over time.

In what follows we first study the IDL data with classical EVT techniques for the whole time interval 1980–2003 as the IDL counts within each window are rather small (Table 1). We then

use the results as motivation for a more sophisticated analysis of the CDS with censored EVT methods for the moving windows.

3 Classical EVT Analysis of the IDL

Consider the classical EVT framework given by iid random variables $X_i \sim F$ with unknown cumulative distribution function (cdf) F . Here X_i indicates the age at death of subject i and we are particularly interested in the *right endpoint* of the cdf F , i.e. $x_F := \sup\{x : F(x) < 1\} \leq \infty$. Since we do not want to presume the existence of a limit to human lifespan, we should also allow for an infinite lifetime, i.e. for $x_F = \infty$. For estimating x_F there exist two well-established EVT approaches, each employing a different class of distributions: the *generalized extreme value distributions* (GEV) and the *generalized Pareto distributions* (GPD). In what follows, we briefly describe these two approaches and refer to de Haan and Ferreira (2006) for more details and proofs.

The GEV is based on the Fisher-Tippett Theorem (see Fisher and Tippett, 1928). It describes the class of potential limit distributions G_γ of the standardized maximum $(\max_{1 \leq i \leq n} \{X_i\} - b_n)/a_n$ for suitable sequences $a_n > 0$ and b_n . The class of GEV $G_\gamma(ax + b)$, is given by

$$G_\gamma(x) = \begin{cases} \exp\left(-(1 + \gamma x)^{-\frac{1}{\gamma}}\right), & 1 + \gamma x > 0 \quad \gamma \neq 0 \\ \exp(-\exp(-x)), & x \in \mathbb{R} \quad \gamma = 0, \end{cases} \quad (1)$$

where a and b are scaling and location parameters which can be estimated from the current dataset.

While this concept is rather general, a GPD approach seems to be more suitable with respect to the left truncated IDL data since it considers values above a given threshold t (see Balkema and de Haan, 1974), say $t = 110$ in case of the IDL. To this end, we are interested in the distribution $F_t(x) = \mathbb{P}(X_1 - t \leq x | X_1 > t)$. For example de Haan and Ferreira (2006) show that the limit equation $\lim_{t \rightarrow x_F} \sup_{x_F - t < x < 0} |F_t(x) - H_{\gamma, \sigma}(x)| = 0$ can only be fulfilled for distributions from the GPD given by

$$H_{\gamma, \sigma}(x) = \begin{cases} 1 - \left(1 + \frac{\gamma x}{\sigma}\right)^{-\frac{1}{\gamma}} & \text{for } \gamma \neq 0 \\ 1 - \exp\left(-\frac{x}{\sigma}\right) & \text{for } \gamma = 0, \end{cases} \quad (2)$$

where σ is again a scaling parameter.

In both cases, the *extreme value index* (EVI) γ is the key parameter that characterizes the tail of the limit distribution. In particular, the value of γ separates two different cases: If $\gamma < 0$ holds (case (i)), the considered distribution has a finite right endpoint which corresponds to a finite maximal age. Contrariwise, if $\gamma > 0$ holds (case (ii)), the right endpoint is infinite and US females may have the potential to live forever. For the decision between the cases (i) and (ii)

and for the estimation of x_F we need adequate estimators for γ . The recent literature discusses several different estimators, some of which are valid only for EVIs in certain intervals (see e.g. Pickands (1975) or Dekkers et al. (1989) for early work on this topic and de Haan and Ferreira (2006) and Resnick (2013) for reviews). However, we particularly allow for positive values of the EVI. This basically leaves us with two choices of estimators for γ : the *moment estimator (Mom)* for the GEV approach and the *maximum likelihood estimator (MLE)* for the GPD approach. The latter is defined as that value of γ which maximizes the likelihood during a fit of the GPD model to the data while the moment estimator is defined as

$$\hat{\gamma}_n^{Mom}(k) := M_n^{(1)}(k) + 1 - \frac{1}{2} \left(1 - (M_n^{(1)}(k))^2 / M_n^{(2)}(k) \right)^{-1}, \quad (3)$$

where $M_n^{(j)}(k) := \frac{1}{k} \sum_{i=1}^k (\log(X_{n-i+1,n}) - \log(X_{n-k,n}))^j$ and $X_{i,n}$ denotes the i -th order statistic. Other estimators such as Pickands or Hill estimators have not been considered due to restricted EVI values (Hill) or less efficiency (Pickands), see e.g. the discussion in Section 3 of de Haan and Ferreira (2006). In what follows, $\hat{\gamma}_n^{(\cdot)}(k)$ denotes the EVI estimator, where k is the number of upper order statistics used and (\cdot) indicates the short-form of the estimator (i.e. MLE or Mom, respectively). The choice of k as a function of the number of observations n is crucial for a good estimate. Since both estimators are consistent for γ and even asymptotically normal if $k \rightarrow \infty$ while $\frac{k}{n} \rightarrow 0$, we should not choose k too small or too large. To this end, e.g. de Haan et al. (2016) describe the following observation: the smaller the values of k , the higher the variance of the EVI estimator while for increasing k the bias of the EVI estimator gets larger. Finding a suitable k between these extremes is called the *bias-variance trade-off*. To get a first idea of the choice of k , we perform a graphical inspection: Figure 1 shows both the moment and maximum likelihood estimates for the IDL depending on k and the point-wise 95% confidence intervals (CI) for both estimates (where we employed the methods of Aarsen and de Haan (1994) for the construction of the Mom-based CI and carried them over to the MLE-based CI for a better comparison).

From Figure 1 we can see that both EVI estimates are positive for $k \leq 46$ and negative for $k > 46$. Interpreting values of $k \leq 46$ as too small this suggests a negative EVI-estimate corresponding to a finite maximal attainable age. However, the CI or more decisively the one-sided CI (not shown here) portend that the null hypothesis $H_0 : \gamma > 0$ of a positive EVI cannot be rejected at the significance level of 5%. In particular, the asymptotic one-sided z -test for H_0 based on the MLE (Mom) gives an approximate p -value of 0.2349 (0.2183). Thus, the hypothesis of a potentially infinite lifetime cannot be rejected based on data from the IDL and the methods described above. If we ignore this statistical insignificance for now and only take into account that the EVI-estimates are negative for $k > 46$, we can nevertheless estimate a maximum attainable age. To this end, we have to find a suitable k . A general recommendation, or rule-of-thumb, is to choose k in a region where these estimates stabilize. Incorporating the bias-variance trade-off as well as the asymptotic framework, this points us to a choice of k

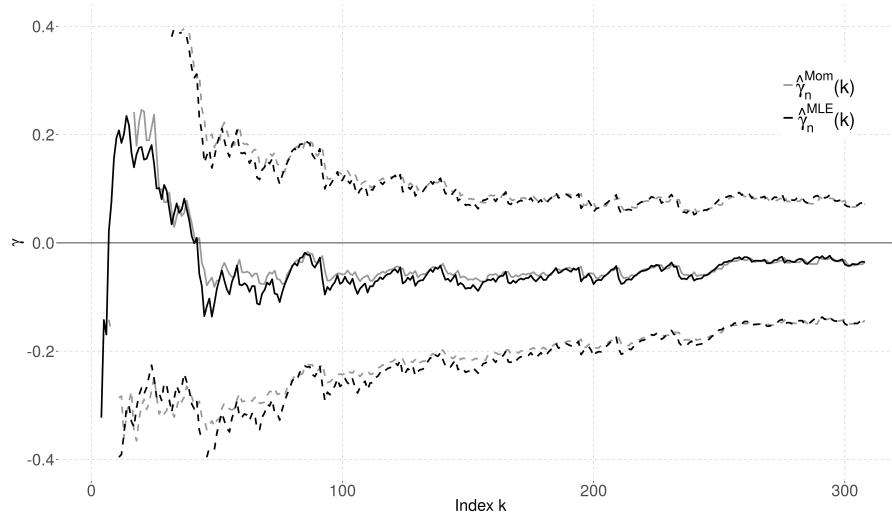


Figure 1: Point estimates $k \mapsto \hat{\gamma}_n^{(\cdot)}(k)$ of the EVI based on the MLE and Mom approach as a function of k together with point-wise 95% CI for the IDL.

between 100 and 200. The minimum, maximum and median of the MLE and Mom estimates for γ with k between 100 and 200, the corresponding value of k , and the 95% confidence intervals are given in Table 2.

		$\hat{\gamma}_n^{(\cdot)}(k)$	(95% CI for γ)	k	$\hat{x}_F^{(\cdot)}(k)$	(95% CI for x_F)
min	MLE	-0.0711	(-0.2173, 0.0751)	155	132.30	(91.62, 172.98)
	Mom	-0.0942	(-0.2773, 0.0888)	106	127.49	(99.53, 155.45)
median	MLE	-0.0572	(-0.2205, 0.1061)	128	136.71	(68.37, 205.05)
	Mom	-0.0663	(-0.2247, 0.0921)	142	133.11	(84.14, 182.08)
max	MLE	-0.0385	(-0.1978, 0.1207)	140	147.60	(02.92, 292.28)
	Mom	-0.0418	(-0.2032, 0.1196)	139	144.49	(21.32, 267.66)

Table 2: MLE- and Mom-estimates of the EVI (γ) and the right-endpoint (x_F) for the IDL data together with 95% confidence intervals for different choices of k .

As already seen in Figure 1 the CIs are rather wide and each contains the value of 0, i.e. a potential infinite lifetime cannot be rejected at level 5%. This may be explained by the small sample size and a potentially negative γ that is quite close to zero (i.e. does not lie far in the alternative). However, each of the estimated EVI values corresponds to a different GPD (MLE case) or GEV (Mom case) and can be used to estimate the right-endpoint of the underlying distribution.

For illustration, exemplary QQ-plots for both median values are given in Figure 2. There the empirical quantiles of the IDL data are plotted against simulated quantiles of the respective distribution functions and the solid lines denote the corresponding regression line, respectively. Both plots show an acceptable fit with a considerable advantage for the GPD model (left panel). This additionally reassures the potential applicability of both EVI approaches. We also investigated whether the underlying distribution function lies in the maximum domain of attraction

of G_γ for the above choices of $\gamma < 0$. By means of the test of Dietrich et al. (2002) with the critical value proposed in Huesler and Li (2006) we could not reject this.

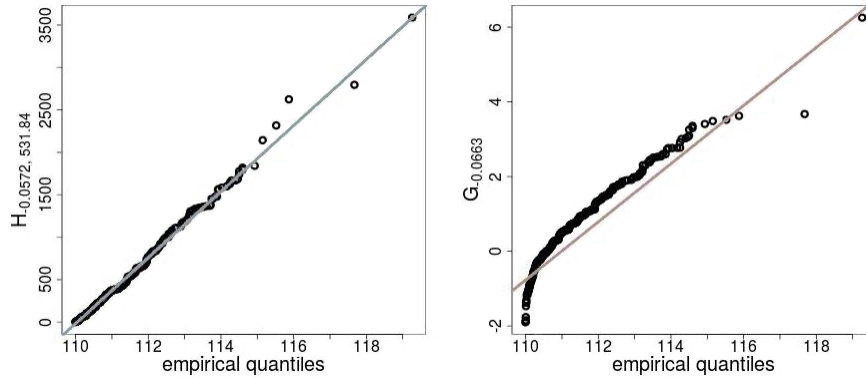


Figure 2: QQ-plots of empirical quantiles over simulated quantiles; left panel: pareto distribution with $\gamma = -0.0572$; right panel: generalized extreme value distribution with $\gamma = -0.0663$. The solid lines visualize the regression line for the plots.

In the next step, we determine the right endpoint of the corresponding models. As explained around Equation (1) above, we need to estimate the scaling and location parameters a and b , in order to find a GEV. To this end we apply the method described in Section 3 of de Haan and Ferreira (2006) and finally obtain estimates for the maximum lifespan $\hat{x}_F^{Mom}(k)$ (GEV) and $\hat{x}_F^{MLE}(k)$ (GPD), respectively. The estimators again depend on the tuning parameter k and are consistent and asymptotically normal under similar regularity assumptions. With this, we can construct 95% CIs. The last two columns of Table 2 show the minimum, median, and maximum of the MLE and Mom endpoint estimates with the corresponding asymptotic 95% CIs and the value of k referring to the endpoint estimates $\hat{x}_F^{Mom}(k)$ and $\hat{x}_F^{MLE}(k)$, respectively. The obtained estimates for the maximum lifespan fluctuate between 127.49 (GEV with minimum EVI Mom-estimate) and 147.6 years (GPD with maximum EVI MLE-estimate). This suggests that the sample size is probably not sufficient for more precise results. Moreover, the confidence intervals are again (too) wide. In particular, for the larger EVI estimates the asymptotic behavior of the underlying variance estimator (which is of order $O(1/\gamma^4)$ for $\gamma \rightarrow 0$) cannot be countervailed by the sample size which leads to the given CIs with poor relevance.

4 A Censored EVT Analysis of the CDS

The results of the previous section are neither statistically significant nor rigorous. As a consequence, we need to significantly increase the sample size of the dataset. Moreover, the neglect of present censoring as well as the underlying iid assumption for the whole time span may lead to implausible estimates. To take care of these issues we analyze the CDS over 16 moving time windows as described in Section 2 by means of censored EVT. To this end, Einmahl et al. (2008) provide several theoretical results with regard to the statistical analysis

of randomly right-censored extreme value data. In their framework, the age at death of the i -th individual is modeled via iid random variables X_i with distribution function F_X . These random variables may be independently right censored by censoring random variables $C_i \stackrel{iid}{\sim} F_C$, where F_C denotes their distribution function. Consequently, the observations in the dataset are given by $Z_i = \min(X_i, C_i)$ and a censoring indicator is given by $\delta_i = \mathbf{1}\{Z_i = X_i\}$. Let $\gamma(\cdot)$ and $x_{F(\cdot)}$ denote the EVI and the endpoint of the distribution function of X, C , and Z , respectively. To answer the target questions on the maximum lifespan we need to determine γ_X and x_{F_X} based on the available data. The classical EVT approaches from the previous subsection, however, would only lead to reasonable estimates for γ_Z and x_{F_Z} (when only using the observations Z_i). On the other hand, a complete case analysis (when deleting all censored observations) would generally result in estimates that are strongly biased. To address censoring adequately, Einmahl et al. (2008) introduce estimators for γ_X and x_{F_X} which are valid in the following cases:

$$\begin{cases} \text{case 1: } \gamma_X > 0, \gamma_C > 0 \\ \text{case 2: } \gamma_X < 0, \gamma_C < 0, x_{F_X} = x_{F_C} \\ \text{case 3: } \gamma_X = \gamma_C = 0, x_{F_X} = x_{F_C} = \infty. \end{cases} \quad (4)$$

With regard to the estimation of the maximal attainable age at death only the second case corresponds to a finite right-endpoint. For the sake of generality we nevertheless also allow for non-negative EVIs γ_X and γ_C .

In what follows, we consider intervals of nine adjacent calendar years for the time span 1980–2003. Starting with 1980–1988 these windows are shifted by steps of one calendar year, see Table 1. As the CDS records of death and censoring counts for the 16 time windows (see Table 1) do not refuse the first two cases of Assumption (4) at first sight we presume that the methods of Einmahl et al. (2008) are applicable to estimate the right endpoint of the deaths curve for each window. The windows contain between about 300,000 and 500,000 (partially censored) observations. Since this figure is still too large for most algorithms with respect to the computing time (even when using modern computing clusters), we follow a sub-sampling approach: Let $\mathcal{Z}_n := \{Z_1, \dots, Z_n\}$ denote the CDS sample for one representative window. We then randomly draw m times from \mathcal{Z}_n without replacement to construct a sub-sample of size m , where $m < n$. Independently repeating this N times we obtain sub-samples $\mathcal{Z}_m^i := \{Z_1^i, \dots, Z_m^i\}$, $1 \leq i \leq N$. We then estimate the EVIs (of the given window) separately for each sub-sample. Following Politis et al. (1999) we should choose m from a theoretical perspective such that $\frac{m}{n} \rightarrow 0$ as $m \rightarrow \infty$. Moreover, the size of the sub-samples m should be sufficiently large, such that each sub-sample includes IDL data with a sufficiently high probability. Taking these issues as well as the computation time into account we calculated estimates for $m = 7500$ and $N \in \{1000, 5000\}$. For the choice of k we employ a cross-validation type procedure that is related to the Leave-One-Out method. This allows for a more objective choice of the tuning parameter k , at least compared to the rule of thumb approach from Section 3. To this end, we draw one more sub-sample \mathcal{Z}_m^0 from

\mathcal{Z}_n of length m that serves as a test sample. All other sub-samples $\mathcal{Z}_m^i, 1 \leq i \leq N$, serve as training samples. For each of these sub-samples $\mathcal{Z}_m^i, 0 \leq i \leq N$, we estimate the EVI $\hat{\gamma}_{X,i}^{(\cdot)}(k_i)$ and thereby vary the number of upper order statistics $1 \leq k_i \leq m$. To this end, we use the methods described by Einmahl et al. (2008).

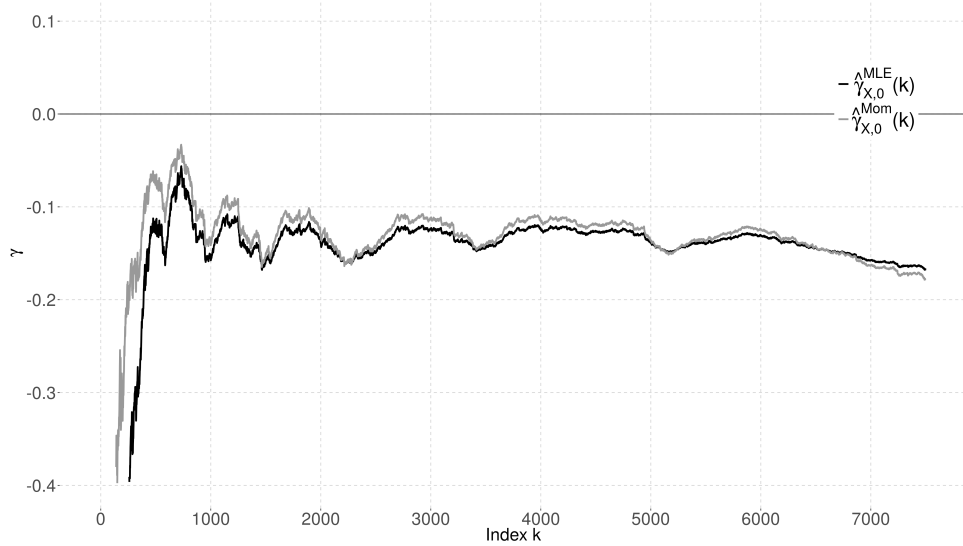


Figure 3: Different EVI estimates as a function of the number of upper order statistics k_0 in case of the trainings dataset of the first window between 1980 and 1988 from the CDS.

Figure 3 shows different EVI estimates $k_0 \mapsto \hat{\gamma}_{X,0}^{(\cdot)}(k_0)$ for the test sample of the first interval between 1980 and 1988 for $N = 5000$. Since there are several stable plateaus visible, it is not clear where the upper order statistic should be chosen to get a good bias-variance tradeoff. We therefore compare the EVI estimates for the test set $k_0 \mapsto \hat{\gamma}_{X,0}^{(\cdot)}(k_0)$ with the corresponding EVI estimates from each training set $k_i \mapsto \hat{\gamma}_{X,i}^{(\cdot)}(k_i)$. For each of these comparisons we obtain the estimator

$$\hat{k}_{0,i} := \operatorname{argmin}_{1 \leq k_0 \leq m} \sum_{k_i=1}^m k_i^{5/4} \left(\hat{\gamma}_{X,0}^{(\cdot)}(k_0) - \hat{\gamma}_{X,i}^{(\cdot)}(k_i) \right)^2, \quad 1 \leq i \leq N. \quad (5)$$

This corresponds to the EVI estimate from the test set \mathcal{Z}_m^0 that has the smallest penalized quadratic distance from all EVI estimates of the i^{th} trainings set \mathcal{Z}_m^i . The choice of the penalty $k_i^{5/4}$ reduces the influence of the bias since it supports intermediate values and penalizes larger values of k (for which the bias is supposed to become too large). Now we can derive an estimator for the EVI by taking the mean of all $\hat{\gamma}_{X,0}^{(\cdot)}(\hat{k}_{0,i}), 1 \leq i \leq N$, obtained from Equation (5):

$$\hat{\gamma}_{X,0}^{(\cdot)} = N^{-1} \sum_{i=1}^N \hat{\gamma}_{X,0}^{(\cdot)}(\hat{k}_{0,i}).$$

To estimate corresponding 95% confidence intervals of the form $(\hat{\gamma}_L, \hat{\gamma}_U)$ we use the asymptotic normality of the EVI estimators given in Einmahl et al. (2008). Proceeding in this way for

all time windows we obtain the MLE- and Mom-based EVI estimates with corresponding 95%-confidence intervals stated in Table 3 (for $N = 5000$). Moreover, the limiting distribution

window	$\hat{\gamma}_n^{(\cdot)}(k)$	Mom (95% CI for γ)	$\hat{\gamma}_n^{(\cdot)}(k)$	MLE (95% CI for γ)
1980-88	-0.14266	(-0.14307,-0.14226)	-0.13500	(-0.13533,-0.13467)
1981-89	-0.14225	(-0.14266,-0.14183)	-0.13827	(-0.14075,-0.13579)
1982-90	-0.15943	(-0.15973,-0.15913)	-0.14813	(-0.14966,-0.14660)
1983-91	-0.15920	(-0.15954,-0.15886)	-0.13979	(-0.14017,-0.13941)
1984-92	-0.13900	(-0.14761,-0.13038)	-0.12407	(-0.12458,-0.12356)
1985-93	-0.13899	(-0.13956,-0.13843)	-0.12448	(-0.12513,-0.12382)
1986-94	-0.14553	(-0.14596,-0.14511)	-0.12650	(-0.12707,-0.12594)
1987-95	-0.14616	(-0.14663,-0.14569)	-0.12669	(-0.12791,-0.12547)
1988-96	-0.15376	(-0.15430,-0.15322)	-0.15025	(-0.15080,-0.14970)
1989-97	-0.15359	(-0.15585,-0.15134)	-0.15062	(-0.15102,-0.15022)
1990-98	-0.14601	(-0.14819,-0.14383)	-0.14717	(-0.14757,-0.14677)
1991-99	-0.14524	(-0.14564,-0.14484)	-0.14676	(-0.14717,-0.14635)
1992-00	-0.15112	(-0.15162,-0.15061)	-0.13735	(-0.13789,-0.13681)
1993-01	-0.13187	(-0.13607,-0.12768)	-0.13764	(-0.14035,-0.13492)
1994-02	-0.15253	(-0.15318,-0.15188)	-0.15808	(-0.15859,-0.15757)
1995-03	-0.14023	(-0.14231,-0.13815)	-0.16646	(-0.16818,-0.16475)

Table 3: MLE- and Mom-estimates of the EVI (γ), upper order statistics k , and 95% confidence intervals for the CDS for $N = 5000$.

of the EVI-estimators directly yields estimates for the right endpoint given any estimate for the EVI. However, no mathematically sound confidence intervals for x_F have been developed so far. We therefore seize on the former suggestion that every EVI-estimate yields a right endpoint estimate and proceed as follows: For the upper and lower limits $\hat{\gamma}_L^{(\cdot)}$ and $\hat{\gamma}_U^{(\cdot)}$ of the 95 % empirical confidence interval of $\gamma_{X,0}$ we calculate corresponding endpoint estimates, say $\hat{x}_{F_X,L}^{(\cdot)}$ and $\hat{x}_{F_X,U}^{(\cdot)}$, to obtain an (at least descriptive) confidence interval $(\hat{x}_{F_X,L}^{(\cdot)}, \hat{x}_{F_X,U}^{(\cdot)})$ for x_F . The results for $N = 5000$ are shown in Table 4 and discussed below.

5 Discussion of the CDS Results

In this section, we discuss our results and their plausibility. We start by describing the results on the EVI given in Table 3 (the results for a subsample size of $N = 1000$ are in the same range and thus not shown here). First, it is apparent that all EVI estimates are below zero for every window. Moreover, even the upper bounds of their 95% confidence intervals are below zero. Since this even holds after multiplicity adjustment for the type I error, e.g. using the Bonferroni correction method (results not shown) (Bretz et al., 2016, Sec. 2), we can argue for a significantly negative EVI ($\gamma_{F_X} < 0$). This indicates the existence of a limit to the lifespan of US females in every window; at least in this retrospective view. A similar investigation for

the censored observations alone also indicated a significant negative $\gamma_{F_C} < 0$ so that Case 2 of Assumption (4) appears to be plausible. Accordingly, we are able to estimate the right endpoint of the deaths curve as described in the former section.

window	Mom		MLE	
	\hat{x}_{F_X}	(95% approx. CI for \hat{x}_{F_X})	\hat{x}_{F_X}	(95% approx. CI for \hat{x}_{F_X})
1980-88	122.802	(120.240, 127.740)	125.409	(122.072, 129.602)
1981-89	122.898	(120.304, 127.801)	124.839	(122.642, 129.510)
1982-90	120.781	(117.280, 123.487)	122.880	(121.530, 124.833)
1983-91	120.732	(117.432, 123.566)	124.099	(123.995, 129.278)
1984-92	123.001	(120.793, 124.673)	126.510	(123.700, 130.830)
1985-93	122.871	(120.881, 125.343)	126.610	(123.661, 130.671)
1986-94	121.832	(118.322, 126.409)	126.310	(123.391, 127.733)
1987-95	121.661	(118.465, 126.420)	126.278	(123.533, 127.622)
1988-96	120.562	(119.143, 122.889)	121.897	(119.420, 125.152)
1989-97	120.797	(119.508, 122.043)	122.206	(119.408, 125.067)
1990-98	122.107	(119.987, 123.567)	122.380	(119.214, 124.939)
1991-99	122.186	(118.805, 126.413)	122.354	(119.232, 124.976)
1992-00	121.107	(118.078, 124.314)	123.482	(119.541, 128.454)
1993-01	123.489	(120.877, 125.906)	124.340	(119.267, 128.550)
1994-02	120.278	(118.502, 122.782)	120.704	(117.576, 124.642)
1995-03	122.478	(120.417, 124.407)	120.946	(119.382, 122.930)

Table 4: MLE- and Mom-estimates of the right endpoint (x_{F_X}) and 95% approximate confidence intervals for the CDS for $N = 5000$.

Table 4 shows these estimates together with their confidence intervals for both the maximum likelihood and the moment estimator in case of $N = 5000$ subsamples (the results for $N = 1000$ are again almost equal and thus not shown). As we can see from this table, both right endpoint estimates range between 120.278 and 126.610 over all time windows. Further, we cannot detect any major trends in the results for the maximum likelihood estimator nor in those of the moment estimator.

Moreover, the confidence intervals range between 117.280 and 130.830 with a maximum interval length of 9.283 years (MLE estimation in the window between 1993 and 2001). For the interpretation of these results, we recall from Section 2 that the IDL stopped collecting data after 2003. Thus, the later time spans contain fewer IDL survival data which may lead to a slight underestimation. Anyhow, the observed rather constant behavior of the deaths curve's right endpoint is in line with recent findings of Einmahl et al. (2019) who also did not find '*indications of trends in these upper limits over the last 30 years*' for Dutch residents based on classical (non-censored) EVT.

As explained in Section 2.1, the oldest US female recorded in the IDL reached an age of 119 years and 97 days and she died in the year 1999. Though the lower bound of the confidence intervals for some windows is lower than 119 years, the estimates for the right endpoint are always larger. Since the HMD is regarded to be complete with respect the US population our

results seem to be plausible.

Advantages of the CDS approach. We finally indicate the advantages of working with the CDS. First, utilizing the HMD alone did not lead to reliable estimates (results shown in Appendix A). In particular, in the analysis of the HMD alone we obtained EVT estimates of the right endpoint that are below the age of 119 years. However, from the IDL we know at least one proven case of a US female surviving the age of 119 (see Section 2.1). On the other hand, we could infer more information from investigating the CDS than building on the IDL alone: The evaluation of the deaths curve's right endpoints in moving windows was not possible on the IDL data alone due to too small sample sizes per window (see Table 1). We, therefore, performed its EVT analysis for the whole time span 1980–2003 in Section 3 for which the stated EVI confidence intervals even include the value of zero (see Table 2). However, the existence of a finite lifespan could be accepted by means of the significantly larger CDS. Since the latter database is larger (in particular it contains the IDL) and incorporates survival data, the corresponding results from Table 3 are more reliable than the results for the IDL alone.

Setting the numbers into context. Comparing our results to previous findings in recent literature is difficult because the basics (i.e. data and methods) differ considerably. In particular, the present statistical analysis is the first that accounts for censoring of the underlying data accordingly. Nevertheless, to at least associate the numbers with the context of a maximum human lifespan we summarize some of the most recent results: Utilizing HMD data alone, de Beer et al. (2017) find that the maximum human lifespan may increase to 125 years. Li et al. (2011) estimate the maximum lifespan for Australians at 112.2 and at 109.43 for New Zealand females and males, respectively. To this end, they use HMD data. Bravo and Corte-Real (2012) also use HMD data and obtain similarly low estimates at 112.77 and 111.78 for Spanish and Portuguese females, respectively. However, at least for Spain and Australia, we can find IDL records for females dying at higher ages (both after their 114th birthday). Einmahl et al. (2019) analyzed about 285,000 death counts of male and female Dutch residents during 1986–2015 (who at least reached an age of 92) for single years (which makes the sample size for each estimation even smaller) and found an average estimated upper endpoint at 115.7 years. Not only in relation to our results, these figures seem to be exceptionally low. For instance, even more than two decades ago, Aarsen and de Haan (1994) infer a 95% confidence interval for the Netherlands between 113 and 124 years (based on about 20,000 combined observations from Dutch men and women). More recently, Weon and Je (2009) obtain maximum age estimates around 126 years for Swedish females and Hanayama and Sibuya (2015) find estimates for the maximum lifespan of Japanese at 123 years. For both analyses, HMD data is used. In contrast, Gbari et al. (2017) use a country-specific database for Belgian citizens which includes 36,616 death counts for females dying in the age range 95+ after 1981. Based on this data they quantify the maximum attainable age of Belgian females between 120.3 and 122.73 years.

Barbi et al. (2018) use data for Italian females to find decelerating death hazards at high ages and conclude that there might be no limit to human longevity. In contrast to that, but in consensus with the aforementioned researches, our estimates show that there is a limit to longevity (in our case) for US females. However, compared to the latter, our results are in the upper range of the results. These disparities in the results may also be caused by different sources of data, differences between the populations under study, or the diversity of the employed methods. In particular, the sample size of the data plays an important role in the quality of the results. In this regard, the data we use outperforms the data from many previous researches (see Table 1). Moreover, for example, the case of Jeanne Calment (see Robine and Allard, 1998) indicates that it is possible to survive to ages beyond 120 years. In this spirit we consider our results to be reasonable.

6 Conclusion and Outlook

The present paper investigates the question of quantifying the maximum human lifespan for US females. To this end, we used two different sources of data: the International Database on Longevity (IDL) and the Human Mortality Database (HMD). We discussed the specific characteristics of each database and additionally exploited their advantages by constructing a comprehensive combined data set, the CDS. These data include both death counts and censored survival data. We then started our investigation of inferring the existence of a possibly finite lifespan by means of extreme value theory (EVT). Starting with a classical EVT analysis of the IDL alone, the hypothesis of an infinite lifespan could not be rejected. This is presumably due to its rather small sample size of 309 for the analyzed time span 1980–2003. Moreover, the obtained point estimates for the maximum attainable age fluctuated considerably and are thus not very meaningful. To obtain more reliable results we subsequently analyzed the CDS which we arranged into intersecting windows to study the evaluation of the lifespan over time. Due to the involved censoring, we employed more convenient censored EVT methods from Einmahl et al. (2008) which we equipped with a computationally efficient algorithm based on sub-sampling and cross-validation.

Finally, we illustrated and discussed our results. For US females, we found significant evidence for a finite lifespan in the combined data and obtained reliable point estimates for the maximum attainable age for each window. These estimates vary between 120 and 127 years which appears to be reasonable compared to existing results in the literature (particularly, the oldest US female recorded in the IDL reached an age of more than 119 years).

Independently of our analysis, other authors like Einmahl et al. (2019) and Rootzén and Zholud (2017) have been motivated to apply (uncensored) EVT methods on different datasets after reading the nature paper by Dong et al. (2016). However, judging from our findings, we recommend the application of censored EVT methods to other populations for future research;

employing the idea of the current CDS: Whenever there is a comprehensive data set which is right censored (as e.g. in the HMD) and a comparatively small but exact dataset (as e.g. given by the IDL) on the same population this method can be used and benefits from much larger and reliable data. In addition, for enhanced analysis additional research is needed from a methodological point of view to construct asymptotically correct confidence limits for the right endpoint, to develop Bayesian EVI estimate versions under censoring and to treat more complicated censoring issues such as, e.g., interval-censored data for the Weibull model.

Acknowledgements

We are grateful for the helpful comments from Joel E. Cohen, who motivated us to introduce moving time windows.

References

- Aarsen, K. and de Haan, L. (1994). On the maximal life span of humans. *Mathematical Population Studies* 4(4): 259–281. DOI: 10.1080/08898489409525379.
- Balkema, A.A. and de Haan, L. (1974). Residual Life Time at Great Age. *The Annals of Probability* 2(5): 792–804. URL: <http://www.jstor.org/stable/2959306>.
- Barbi, E., Lagona, F., Marsili, M., Vaupel, J.W., and Wachter, K.W. (2018). The plateau of human mortality: Demography of longevity pioneers. *Science* 360: 1459–1461. DOI: 10.1126/science.aat3119.
- Börger, M., Genz, M., and Ruß, J. (2018). Extension, Compression, and Beyond - A Unique Classification System for Mortality Evolution Patterns. *Demography* 55(4): 1343–1361. DOI: 10.1007/s13524-018-0694-3.
- Bravo, J.M. and Corte-Real, P. (2012). *Modeling Longevity Risk using Extreme Value Theory: An Empirical Investigation using Portuguese and Spanish Population Data*. Proceedings of the 7th Finance Conference of the Portuguese Finance Network, July 5-7, 2012, Aveiro, Portugal. URL: <http://hdl.handle.net/10174/7344>.
- Bretz, F., Westfall, P., and Hothorn, T. (2016). *Multiple comparisons using R*. New York: Chapman and Hall/CRC.
- Cohen, J.E. and Oppenheim, J. (2012). Is a limit to the median length of human life imminent? *Genus* 68(1): 11–40. URL: <http://www.jstor.org/stable/genus.68.1.11>.
- de Beer, J., Bardoutsos, A., and Janssen, F. (2017). Maximum human lifespan may increase to 125 years. *Nature* 546(7660): E16–E17.
- de Grey, A.D.N.J. (2003). The foreseeability of real anti-aging medicine: focusing the debate. *Experimental Gerontology* 38(9): 927–934. DOI: 10.1016/S0531-5565(03)00155-4.

- de Haan, L. and Ferreira, A. (2006). *Extreme value theory: an introduction*. New York; [Heidelberg]: Springer.
- de Haan, L., Mercadier, C., and Zhou, C. (2016). Adapting extreme value statistics to financial time series: dealing with bias and serial dependence. *Finance and Stochastics* 20(2): 321–354. DOI: 10.1007/s00780-015-0287-6.
- Dekkers, A. L. M., Einmahl, J.H.J., and de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution. *The Annals of Statistics* 17(4): 1833–1855.
- Dietrich, D., de Haan, L., and Huesler, J. (2002). Testing extreme value conditions. *Extremes* 5(1): 71–85. DOI: 10.1023/A:1020934126695.
- Dong, X., Milholland, B., and Vijg, J. (2016). Evidence for a limit to human lifespan. *Nature* 538(7624): 257–259.
- Einmahl, J., Einmahl, J.H.J., and de Haan, L. (2019). Limits to human life span through extreme value theory. *Journal of the American Statistical Association* 0(ja): 1–15. DOI: 10.1080/01621459.2018.1537912.
- Einmahl, J.H.J., Fils-Villetard, A., and Guillou, A. (2008). Statistics of extremes under random censoring. *Bernoulli* 14(1): 207–227.
- Falk, M., Hüsler, J., and Reiss, R.-D. (2010). *Laws of small numbers: extremes and rare events*. New York; [Heidelberg]: Springer.
- Fisher, R.A. and Tippett, L.H.C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society* 24 (2): 180–190. DOI: 10.1017/S0305004100015681.
- Fries, J.F. (1980). Aging, natural death, and the compression of morbidity. *New England Journal of Medicine* 303(3): 130–135.
- Gbari, S., Poulain, M., Dal, L., and Denuit, M. (2017). Extreme Value Analysis of Mortality at the Oldest Ages: A Case Study Based on Individual Ages at Death. *North American Actuarial Journal* 21(3): 397–416.
- Gomes, M.I. and Guillou, A. (2015). Extreme value theory and statistics of univariate extremes: A review. *International Statistical Review* 83(2): 263–292. DOI: 10.1111/insr.12058.
- Gomes, M.I. and Neves, M.M. (2011). Estimation of the extreme value index for randomly censored data. *Biometrical Letters* 48(1): 1–22.
- Hanayama, N. and Sibuya, M. (2015). Estimating the Upper Limit of Lifetime Probability Distribution, Based on Data of Japanese Centenarians. *The Journals of Gerontology: Series A* 71(8): 1014–1021. DOI: 10.1093/gerona/glv113.
- HMD (2015). *Human Mortality Database*. University of California, Berkeley, and Max Planck Institute for Demographic Research. URL: www.mortality.org.
- Huesler, J. and Li, D. (2006). On testing extreme value conditions. *Extremes* 9(1): 69–86. DOI: 10.1007/s10687-006-0025-8.
- IDL (2017). *International Database on Longevity*. Max Planck Institute for Demographic Research. URL: <http://www.supercentenarians.org/>.

- Li, J.S.H., Ng, A.C.Y., and Chan, W.S. (2011). Modeling old-age mortality risk for the populations of Australia and New Zealand: An extreme value approach. *Mathematics and Computers in Simulation* 81(7): 1325–1333. DOI: 10.1016/j.matcom.2010.04.025.
- Maier, H., Gampe, J., Jeune, B., Robine, J.-M., and Vaupel, J.W., eds. (2010). *Supercentenarians. Demographic Research Monographs (A series of the Max Planck Institute for Demographic Research)*. Berlin, Heidelberg: Springer.
- Pickands, J. (1975). Statistical Inference Using Extreme Order Statistics. *The Annals of Statistics* 3(1): 119–131.
- Politis, D.N., Romano, J.P., and Wolf, M. (1999). *Subsampling*. New York; [Heidelberg]: Springer.
- Reiss, R.-D. and Thomas, M. (2007). *Statistical analysis of extreme values*. New York; [Heidelberg]: Springer.
- Resnick, S.I. (2013). *Extreme values, regular variation and point processes*. New York; [Heidelberg]: Springer.
- Robine, J.-M. and Allard, M. (1998). The oldest human. *Science* 279(5358): 1831. DOI: 10.1126/science.279.5358.1831h.
- Robine, J.-M. and Paccaud, F. (2005). Nonagenarians and centenarians in Switzerland, 1860–2001: a demographic analysis. *Journal of Epidemiology & Community Health* 59(1): 31–37. DOI: 10.1136/jech.2003.018663.
- Rootzén, H. and Zholud, D. (2017). Human life is unlimited – but short. *Extremes* 20(4): 713–728. DOI: 10.1007/s10687-017-0305-5.
- Watts, K.A., Dupuis, D.J., and Jones, B.L. (2006). An Extreme Value Analysis Of Advanced Age Mortality Data. *North American Actuarial Journal* 10(4): 162–178. DOI: 10.1080/10920277.2006.10597419.
- Weon, B.M. and Je, J.H. (2009). Theoretical estimation of maximum human lifespan. *Biogerontology* 10(1): 65–71.
- Wilmoth, J.R., Andreev, K., Jdanov, D., and Gleijeses, D.A. (2017). *Methods protocol for Human Mortality Database (Version 6)*. URL: <http://www.mortality.org/Public/Docs/MethodsProtocol.pdf>.
- Wilmoth, J.R., Deegan, L. J., Lundström, H., and Horiuchi, S. (2000). Increase of Maximum Life-Span in Sweden, 1861–1999. *Science* 289(5488): 2366–2368. DOI: 10.1126/science.289.5488.2366.
- Worms, J. and Worms, R. (2014). New estimators of the extreme value index under random right censoring, for heavy-tailed distributions. *Extremes* 17(2): 337–358. DOI: 10.1007/s10687-014-0189-6.

Appendices

A Analysis of the HMD Data

The valuable information added to analysis by including the supercentenarians of the IDL becomes apparent when considering the HMD in a stand-alone analysis. The HMD contains single year and single age death counts and survivors up to the age of 109. However, the age range 110+ only gives cumulative numbers, i.e. the total number of people alive / dying in some age beyond 110. Thus, the HMD data is right censored.

The use of survival data of the HMD besides the death counts per calendar year and age, results in a *randomly* right censored data set (as the age at death of the US females surviving a time interval is unknown). Thus, in estimating the right endpoint of the deaths curve on only the HMD data, we can proceed analogously to the methods of the main part of the article. This results in the estimates of Table 5. For the sake of brevity and conciseness here we only illustrate the results for the Moment estimator and for $N = 5000$. This table corresponds to the tables in the main part of the article, where we illustrate our results for the CDS analysis. The estimates

window	$\hat{\gamma}_n^{(\cdot)}(k)$	(95% CI for γ)	\hat{x}_{F_X}	(95% CI for \hat{x}_{F_X})
1980-1988	-0.182	(-0.18271,-0.18219)	118.501	(116.602,121.138)
1981-1989	-0.191	(-0.19116,-0.19006)	117.333	(116.177,119.231)
1982-1990	-0.188	(-0.18862,-0.18800)	117.876	(116.555,119.587)
1983-1991	-0.206	(-0.20624,-0.20548)	116.361	(115.567,117.716)
1984-1992	-0.215	(-0.21503,-0.21434)	115.672	(114.757,117.909)
1985-1993	-0.204	(-0.20478,-0.20323)	116.678	(115.112,119.085)
1986-1994	-0.210	(-0.21067,-0.20993)	115.748	(114.153,117.503)
1987-1995	-0.210	(-0.21040,-0.20962)	115.704	(115.011,118.413)
1988-1996	-0.205	(-0.20515,-0.20419)	115.963	(115.090,119.293)
1989-1997	-0.209	(-0.20979,-0.20876)	115.637	(114.947,117.192)
1990-1998	-0.216	(-0.21607,-0.21550)	115.264	(113.997,116.611)
1991-1999	-0.246	(-0.24668,-0.24594)	113.702	(112.969,114.008)
1992-2000	-0.230	(-0.23094,-0.22986)	114.125	(113.045,115.886)
1993-2001	-0.199	(-0.19895,-0.19806)	116.116	(115.237,117.096)
1994-2002	-0.188	(-0.18827,-0.18724)	117.163	(116.253,119.798)
1995-2003	-0.186	(-0.18648,-0.18570)	117.033	(115.429,118.539)

Table 5: Mom-estimates of the EVI (γ) and the right endpoint (x_{F_X}) with 95% confidence intervals for the HMD for $N = 5000$.

for the EVI in this application are thoroughly smaller than zero and even smaller than those of the CDS analysis in the main part of the article. The upper bounds of the 95% confidence intervals of these estimates are below zero in each window. This indicates the existence of a finite right endpoint for the HMD data. In this respect, the HMD analysis does not differ from the analysis of the CDS although the estimates here are considerably smaller.

The estimates for the right endpoint on only the HMD data are also smaller than those of the CDS analysis in each interval. Also the confidence intervals of this analysis are in lower age ranges than those of the CDS analysis. For many windows they not even contain the age of 119, while the confidence intervals for the right endpoint on the CDS data either contain this age or go beyond. Taking into account, that there was a proven case of a US female dying in the age of over 119 in 1999 (see discussion of the IDL data set in the main part of the article), the results of the CDS analysis appear more plausible than those illustrated in Table 5.

4 The Future of Mortality: Mortality Forecasting by Extrapolation of Deaths Curve Evolution Patterns

Source:

Börger, M., Genz, M., Ruß, J. (2019). The Future of Mortality: Mortality Forecasting by Extrapolation of Deaths Curve Evolution Patterns. Working Paper.

The Future of Mortality: Mortality Forecasting by Extrapolation of Deaths Curve Evolution Patterns

Matthias Börger*

Martin Genz[†]Jochen Ruß[‡]

Abstract

A variety of mortality models can be used to project future mortality. However, the parameters of most of these models lack a clear demographic interpretation. Hence, it may be hard to identify forecasts which are not consistent to the most recent observed trends in the mortality evolution or that are demographically implausible. On the other hand, demographers make predictions on future mortality but typically focus on single aspects only instead of comprehensive mortality forecasts. This article aims to close the gap between these forecasting approaches.

We establish a new best estimate mortality model which is based on the extrapolation of four statistics that have a clear demographic interpretation. The four statistics are taken from the classification framework of Börger et al. (2018). Our model yields forecasts for the deaths curve which are consistent with the most recent demographic trends in the deaths curve evolution. Moreover, expert opinions on future trends in the mortality evolution can easily be incorporated, and we illustrate how the model can be used for scenario analyses.

Keywords: Mortality scenario classification, Demographic mortality trends, Demographic scenario analyses, (consistent) Mortality forecasting

*Institut für Finanz- und Aktuarwissenschaften (ifa), Lise-Meitner-Straße 14, 89081 Ulm, Germany, email: m.boerger@ifa-ulm.de

[†]Institut für Finanz- und Aktuarwissenschaften (ifa) and Institut für Versicherungswissenschaften, Universität Ulm, Lise-Meitner-Straße 14, 89081 Ulm, Germany, email: m.genz@ifa-ulm.de (corresponding author)

[‡]Institut für Finanz- und Aktuarwissenschaften (ifa) and Institut für Versicherungswissenschaften, Universität Ulm, Lise-Meitner-Straße 14, 89081 Ulm, Germany, email: j.russ@ifa-ulm.de

1 Introduction

Estimates for future mortality are important in many areas, e.g. for projections of social security systems or risk management in the private pension and life insurance sector. To derive such estimates, a variety of mortality models has been developed most of which are statistical models in the sense that they extract certain patterns from historical mortality data and extrapolate those patterns into the future.

Some of the earliest mortality models (which are also called "laws of mortality") are quite simple, as they are based on only few parameters. One of the first models in this spirit is the law of mortality by Gompertz (1825). This model has often been extended. For example, Pollard (1987) gives a comprehensive review on mortality models which have been developed from the 19th century until the 1980s. Such models have often been used for deterministic mortality forecasts. During the most recent decades, stochastic mortality models have emerged, e.g. the Lee-Carter model (Lee and Carter, 1992) or the Cairns-Blake-Dowd model (Cairns et al., 2006). An overview is given for example by Booth and Tickle (2008) or (with a more actuarial focus) in Cairns et al. (2008). While the majority of the early "laws of mortality" only take into account age-dependent patterns of mortality, some of the more recent models also allow for period and even cohort effects. This allows for model fitting to longer data series and in particular for more flexible (e.g. stochastic) extrapolations.

However, in many cases, the parameters in these models lack a clear demographic interpretation. Therefore, it is not easy for the user of the model to assess if an extrapolation of these parameters will result in plausible mortality forecasts from a demographic point of view. We will provide examples for that later on.

In the literature we also find many approaches for forecasting (certain aspects of) future mortality from a demographic point of view. For example, Manton et al. (1991) divide the perspectives on the future life expectancy in three groups - the "traditional", the "visionary", and the "empiric" perspective - and give examples for each perspective. Olshansky et al. (1990) take the "traditional" point of view by suggesting that there might be a limit to the increase in human longevity. More than one decade later, Oeppen and Vaupel (2002) analyzed the world record life expectancy and found a linearly increasing trend which they extrapolate several decades into the future. In the spirit of Manton et al. (1991) this is rather a "visionary" perspective. Sometimes demographic forecasts are also made implicitly (i.e. without an explicit extrapolation of any statistic). For example, Dong et al. (2016) "strongly suggest that the maximum lifespan of humans is fixed" and build their claim on historical trends in demographic statistics measuring the evolution of old-age mortality. In the same spirit, but more than three decades earlier Fries (1980) derived future limits for the maximum human lifespan based on demographic statistics. So demographers typically do not only use statistical observations (which are often the only input for the statistical mortality modeling), but also combine knowledge from biology, medicine, sociology, and other fields of research in their forecast. However, their forecasts often focus on

single aspects of the mortality evolution only, e.g. the life expectancy at birth or the maximum human lifespan, and they typically do not make suggestions for the entire distribution of deaths over all ages in any future year. The latter, however, is provided by mortality models (but typically without taking demographers' forecasts into account). Therefore, it seems worthwhile to combine the purely statistical approach of mortality modeling with the demographic expertise. This paper aims to close this gap. To this end, we propose a methodology that can derive the entire mortality distribution for any future point in time based on forecasts for a set of statistics that come with a clear demographic interpretation.

We build our methodology on recent work of Börger et al. (2018). They introduce a new classification framework for (observed historic) changes in mortality over time which is based on the deaths curve, i.e. the density function of the age at death. They show that four statistics, each with a clear demographic interpretation, describe the key structural elements of a deaths curve - its support and the pattern it exhibits on this support. Thus, they express changes in a deaths curve between two points in time by changes in these four statistics. In their classification framework, a uniquely defined "mortality scenario" is then assigned to every possible evolution of a deaths curve over time. Such a scenario consists of four components and each component has one of three specifications: It can 1) be "right-shifting" or "left-shifting" (or neutral in respect of "shifting mortality"), it can 2) exhibit "extension" or "compression" (or be neutral in this dimension), it can 3) show signs of "compression" or "decompression" (or be neutral in that respect), and it can 4) show "concentration" or "diffusion" (or again be neutral in that respect). In Section 2.1 we will explain this in more detail.

This classification framework allows us to check if any given mortality forecast is consistent with most recent demographic trends. A mortality forecast can be considered demographically reasonable in case the historical patterns in the four statistics are extrapolated in a demographically plausible manner.

The other way round, it can also serve as a basis to create mortality forecasts for the whole deaths curve that match a given forecast for the four statistics from the classification framework. Such a forecast can either be an extrapolation of the four statistics based on their past evolution or a forecast based on demographic insights (or a combination thereof, e.g. an extrapolation that is adjusted based on demographic insights). This results in a new type of mortality model which will be derived in this paper.

The remainder of this paper is organized as follows: In Section 2 we briefly summarize the classification framework of Börger et al. (2018) and test frequently used mortality forecasting models for consistency with the past in terms of their projection of the four key statistics. This illustrates the need for mortality forecasts with better consistency to recently observed mortality scenarios. Consequently, in Section 3 we introduce a deterministic mortality forecasting model, which is based on the four components of the aforementioned classification framework. This model uses forecasts for the four statistics (e.g. extrapolation of recent trends or some expert opinion) to project the entire deaths curve and thus allows for consistency of the mortality

forecasts with recently observed mortality scenarios and/or the experts' expectations. In Section 4 we discuss examples and potential applications for this model. Finally, Section 5 concludes.

2 Consistency of Existing Mortality Models

2.1 A Unique Classification Framework for Mortality Evolution Patterns

This section briefly summarizes the classification framework introduced by Börger et al. (2018). They uniquely assign a mortality scenario to any change of a deaths curve between two points in time. More concretely, a mortality scenario is defined as a four dimensional vector with the following components:

1. *Right- or left-shifting mortality* is linked to an increase or decrease in the *modal age at death* (M). This statistic is defined as the age in which most people die, i.e. the maximum of the deaths curve. Thus M can intuitively be interpreted as the position of the center of the deaths curve. So shifting mortality is linked to a shift of the deaths curve's center to the right or to the left, respectively.¹
2. *Extension or contraction* is linked to an increase or decrease of the *upper bound of the deaths curve's support* (UB). This statistic is defined as the right endpoint of the deaths curve's support.² Intuitively it is the highest age that has been reached within the population of interest. Whenever this age moves to the right, the support of the deaths curve extends and whenever UB moves to the left the deaths curve's support contracts. Like the modal age at death, UB indicates changes in the deaths curve's position.
3. *Compression or decompression* is linked to an increase or a decrease of the *degree of inequality* (DoI). This statistic measures the size of the area between the observed deaths curve and a hypothetical deaths curve where the age at death is uniformly distributed. Intuitively this area (and hence the inequality of the distribution of deaths by age) will increase, if the observed deaths curve moves farther away from the uniformly distributed deaths curve. Consequently, there must be age segments, where the number of deaths increases, i.e. deaths become more compressed. If the observed deaths curve moves closer towards this uniform distribution, the area between these two curves will decrease and hence we observe decompression.

¹Börger et al. (2018) noted that the "peak might not be unique in only rather theoretical scenarios [...]. In such a case, one might use a suitable alternative to M or modify the framework to include additional statistics."

²In the literature we find an extensive discussion about the existence of this age. For example Gampe (2010) or more recently Barbi et al. (2018) find evidence for a plateau in the mortality hazard after an age beyond 100. In contrast, for example Gavrilov and Gavrilova (2011) find that this is an artefact. By use of methods from extreme value theory for example Feifel et al. (2018) found significant statistical evidence for the existence of a finite lifespan for US females. Thus, like Börger et al. (2018) we assume that the right endpoint of the deaths curve's support always exists.

4. *Concentration or diffusion* is linked to an increase or decrease in the *number of deaths in the modal age at death* ($d(M)$). This statistic measures the deaths curve's value at age M . If more/less people die at the ages around M , $d(M)$ will increase/decrease. Intuitively this statistic indicates the relative importance of the ages around M compared to the rest of the deaths curve. Like DoI , $d(M)$ indicates changes in the deaths curve's shape.

Each component of this vector can also assume the value neutral, which is the case if the respective statistic neither (significantly) increases nor decreases between two points in time. This results in a total of $3^4 = 81$ possible different scenarios, some of which might not be relevant in practice. It is noteworthy though that any material change of the deaths curve is detected by at least one of the components. Börger et al. (2018) consider a scenario to be a pure scenario if only one of four components is different from neutral. If at least two components indicate a change, a mixed scenario prevails. They find that in practice, mixed scenarios are the rule rather than the exception (see also Genz, 2017) and conclude that the presence of mixed scenarios is not considered sufficiently in the literature.

This classification framework for observed changes of the deaths curve can be applied to deaths curves with any starting age x_0 . For example, if one is interested in the evolution of old-age mortality, x_0 should be a rather old age and the deaths curve gives the age distribution at death conditional on survival to this (old) age.

We would like to emphasize once again that each of the aforementioned statistics has an intuitive interpretation. So either an extrapolation of past trends or an expert opinion on the future development can be used in forecasting. Also, it can be relatively easily assessed whether a simple statistical extrapolation of a past trend is (demographically) reasonable or not. This is much more difficult in models that are based on an extrapolation of statistics which do not allow for an intuitive interpretation. While an intuitive interpretation is not necessary for extrapolations of past trends, it is of course very helpful in case of extrapolations based on expert opinion.

An extrapolation of past trends can be based on the methodology that was developed by Börger et al. (2018) to determine (piecewise linear) trends in each of the four statistics in the past. In order to illustrate this concept, we use data from the Human Mortality Database (HMD, 2015) for Swiss females and determine the deaths curve for each calendar year between 1920 and 2014 and starting age $x_0 = 60$. For each calendar year, we estimate the four statistics of the classification framework and fit piecewise linear trends to the resulting time series (see Figure 1).³ For the extrapolation performed in this paper, the slope of most recent trend line is relevant, whereas for the classification in Börger et al. (2018), the sign of the trend was more

³Between any two trends, Börger et al. (2018) allow for jumps, which can be important whenever a time series has sudden changes in its level. In Figure 1 we can see an example for such a jump in UB .

relevant.⁴

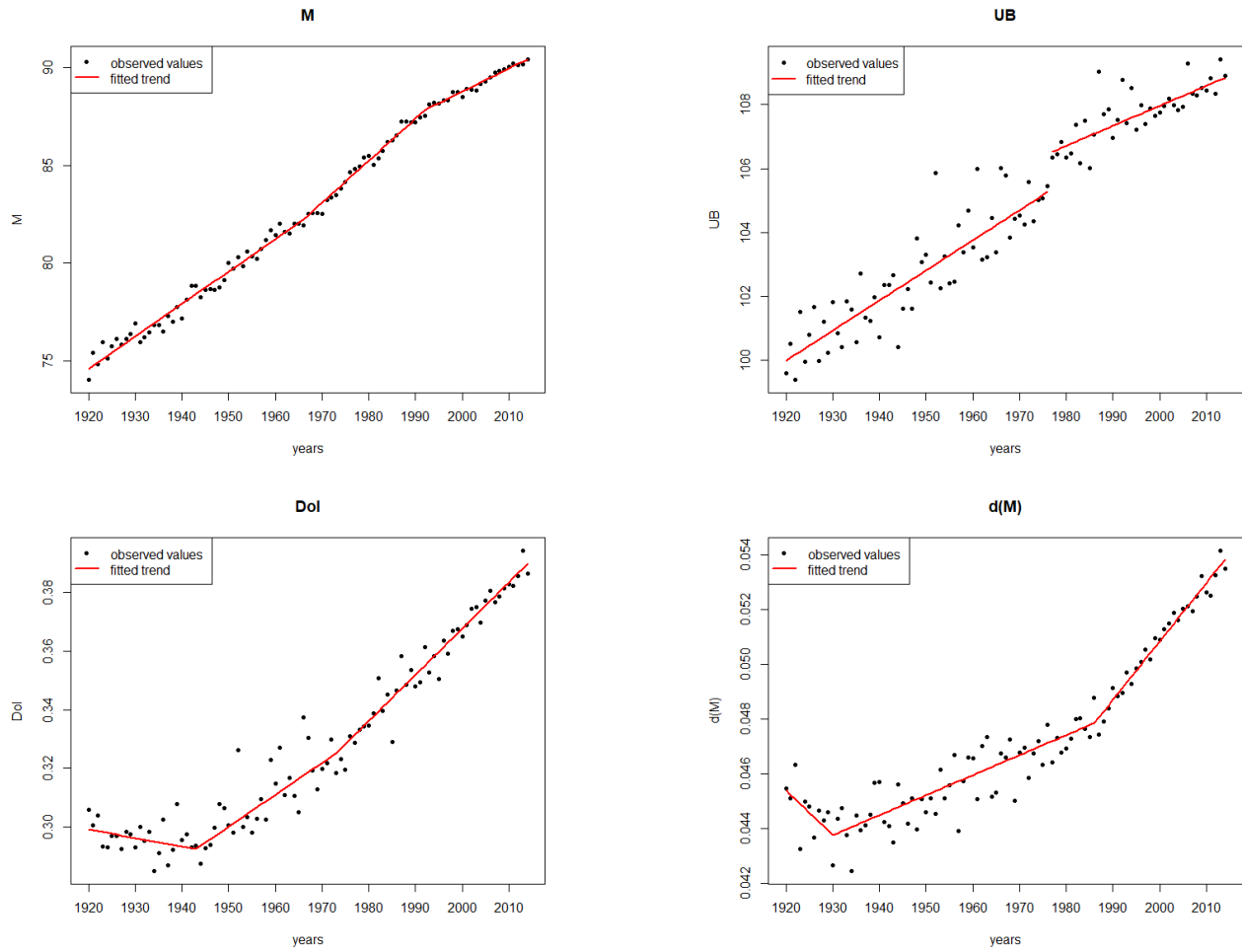


Figure 1: Observed trends in the four statistics of the classification framework for Swiss females between 1920 and 2014; top-left: M , top-right: UB , bottom-left: DoI , bottom-right: $d(M)$

2.2 Consistency Issues in Existing Forecasting Models

In this subsection, we check the demographic consistency of mortality forecasts. We fit two well-established mortality forecasting models (the Lee-Carter model (LC) and the Cairns-Blake-Dowd model (CBD)) to the mortality data from the previous subsection. From Figure 1 we can see, that the most recent trend change occurred in 1992 for M , in 1976 for UB , in 1963 for DoI , and in 1986 for $d(M)$. Since all statistics exhibit stable trends from 1992 onward, we fit both mortality models to data between 1992 and 2014 and for ages above 60. This gives both

⁴Hence, they developed a test whether any trend is significantly different from neutral (i.e. the slope of the line fitted to the time series is significantly different from zero). For any calendar year, each of the four components can then be classified as increasing, decreasing, or neutral. The resulting four-dimensional vector is the mortality scenario for this calendar year.

models the best possible chance to extrapolate historical trends in a demographically reasonable manner.⁵

Figure 2 shows the observed values of each statistic of the classification framework between 1992 and 2014 as well as extrapolated linear trends and the values forecast with the LC and the CBD model from 2015 to 2070.⁶

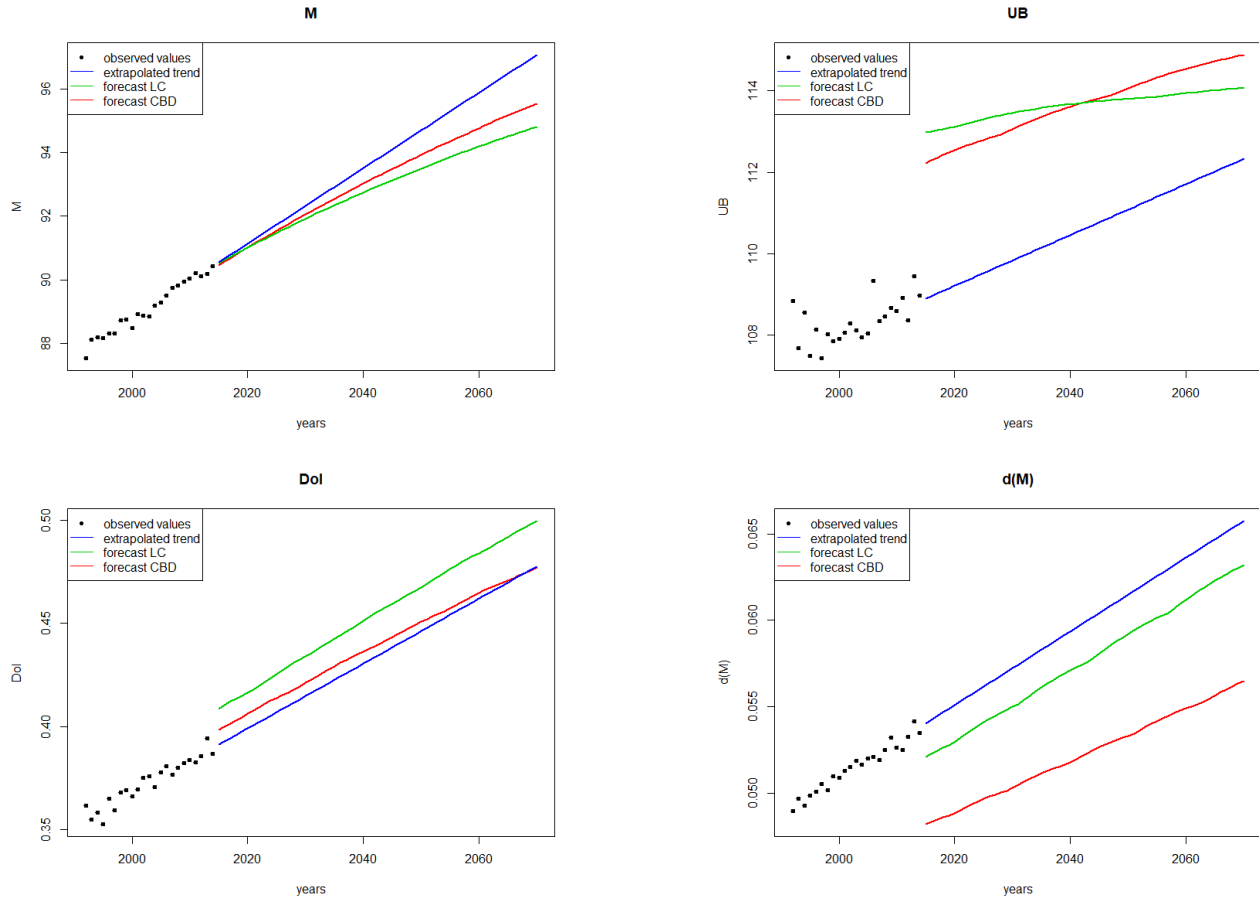


Figure 2: Observed values (black dots) between 1992 and 2014, extrapolated trends (blue lines), and model forecasts of the LC (green line) and the CBD model (red line) from 2015 to 2070; top-left: M ; top-right: UB ; bottom-left: DoI ; bottom-right: $d(M)$

Three of the four statistics exhibit significant jumps at the transition from the calibration to the forecasting period in both mortality models. This indicates that the extrapolated deaths curves

⁵When determining the calibration period for mortality models which do not allow for trend changes, the mortality evolution during the calibration period should not exhibit any trend change. Thus for such models the most recent trend change in the classification framework of Börger et al. (2018) provides a suitable criterion for the choice of the calibration period.

⁶To calibrate the parameters of both models we use the R package StMoMo (see Villegas et al., 2018). This package also gives best estimate forecasts from which we determine estimates of the four statistics of the classification framework. The kinks in the forecast values for UB , DoI , and $d(M)$ stem from a discretization issue which is due to the fact that the models only provide mortality rates for integer ages. We ignore this minor issue since it does not affect the interpretation of the graphs.

do not exhibit smooth changes at this transition, i.e. they are not consistent with the immediate past. Thus, for both mortality models, the forecast deaths curves appear to be implausible from a demographic point of view even for the immediate future.

The projection of M on the other hand seems more plausible than those for the other three statistics since the transition from the calibration to the forecasting period does not exhibit a jump.

3 A New Best Estimate Mortality Model

In this section, we present a deterministic mortality forecasting model which can yield deaths curves that are consistent with a specific given forecast for the four statistics defined in Section 2.1. In particular, consistency with the most recent evolution of the deaths curve can be ensured by extrapolating the most recent trends in the four statistics.

3.1 Theoretical Concept

Our mortality model forecasts (continuous) deaths curves which – if scaled such that it integrates to one – can be interpreted as the density function of the age distribution at death. From these deaths curves basically all other mortality statistics and in particular age dependent mortality rates can be deduced.

In order to develop the theoretical concept behind our model, we first assume that a forecast $(M_t, UB_t, DoI_t, d(M)_t)$ of the four statistics for any year t of the forecasting period is given. We then denote the space of all density functions on the (age) interval $[x_0, UB_t]$ which are differentiable at least three times by \mathcal{D}_{t,x_0} . It contains all functions $d_t : [x_0, UB_t] \rightarrow [0, 1]$ that fulfill the following two properties:

$$\int_{x_0}^{UB_t} d_t(x) dx = 1 \quad (1)$$

and

$$d_t(x) \geq 0 \text{ for all } x \in [x_0, UB_t]. \quad (2)$$

In order to restrict that space to deaths curves with "typical" shapes, we demand the following additional properties:

The deaths curve assumes the value of zero only at UB_t :

$$d_t(x) \begin{cases} > 0 \text{ for all } x \in [x_0, UB_t) \\ = 0, x = UB_t \end{cases} . \quad (3)$$

For a sufficiently large x_0 , the deaths curve is unimodal, i.e. it has a unique global maximum which is also the only local maximum.⁷ Consequently, we get:

$$d'_t(x) \begin{cases} > 0 \text{ for all } x \in (x_0, M_t) \\ < 0 \text{ for all } x \in (M_t, UB_t) \end{cases} . \quad (4)$$

For a sufficiently small choice of x_0 , a typical deaths curve has two (unknown) inflection points, one between x_0 and M_t and one between M_t and UB_t .⁸ Further, the deaths curve is concave between the inflection points and convex in both tails. Accordingly, the second derivative is positive in x_0 and UB_t , and negative in M_t :

$$d''_t(x_0) > 0, \quad (5)$$

$$d''_t(M_t) < 0,$$

$$d''_t(UB_t) > 0, \text{ and}$$

$$d'''_t(x) \begin{cases} < 0 \text{ for all } x \in [x_0, M_t) \\ > 0 \text{ for all } x \in (M_t, UB_t] \end{cases} .$$

We denote by D_{t,x_0} the subset of \mathcal{D}_{t,x_0} which contains all deaths curves that fulfill these constraints. It is the set of "density functions with reasonable shape".

In the practical application of this model there might be further requirements the deaths curve should fulfill to be plausible. Therefore the set D_{t,x_0} further can be restricted. In particular, for more stable results in the left tail, it is advisable to specify the number of deaths in the starting age $d(x_0)$. This is:

⁷Besides the expected maximum at M , local maxima can typically be observed at age 0 (infant mortality) and at the peak of the so-called accident hump (typically at some age between 18 and 25). Thus, for example a choice of $x_0 = 30$ would be sufficiently large in any relevant case. In this sense, this is not a limitation if one is interested in mature- and old-age mortality.

⁸If the distance between x_0 and M becomes too small, it may happen that the deaths curve is concave at x_0 and there is no inflection point between x_0 and M_t . In practical applications, however, this is irrelevant unless an x_0 is chosen that is very close to the initial value of M and/or a strongly decreasing trend for M is assumed.

Given a forecast for the number of deaths at the starting age at time t , $d(x_0)_t$, it must hold:

$$d_t(x_0) = d(x_0)_t \quad (6)$$

for pre-specified values $d(x_0)_t$.

Since typically, the logarithm of the number of deaths at the starting age $\log d(x_0)$ follows a piecewise linear trend, we linearly extrapolate the most recent trend of this statistic in order to get reasonable and demographically consistent forecasts for $d(x_0)_t$ at any future year t .

In the last step, we consider the subset $\overline{D}_{t,x_0} \subset D_{t,x_0}$ containing only deaths curves which also match the forecasts of the statistics $(M_t, UB_t, DoI_t, d(M)_t)$, i.e. for each element $d_t : [x_0, UB_t] \rightarrow [0, 1]$ of \overline{D}_{t,x_0} it must hold:⁹

$$\arg \max d_t(x) = M_t, \quad (7)$$

$$\sup_{x > x_0} (d_t(x) > 0) = UB_t, \quad (8)$$

$$\frac{UB_t - x_0}{2 \cdot (UB_t - x_0) - 1} \cdot \int_{x=x_0}^{UB_t} \left| d_t(x) - \frac{1}{UB_t - x_0} \right| dx = DoI_t, \text{ and} \quad (9)$$

$$\max_{x \geq x_0} (d_t(x)) = d(M)_t. \quad (10)$$

Of course, the set \overline{D}_{t,x_0} might be empty or there might be more than one curve in this set. In case this set is empty, we follow that the (long-term) extrapolations of the statistics are not plausible.¹⁰ In case there is more than one deaths curve in the set \overline{D}_{t,x_0} , a criterion is required to determine which deaths curve should serve as a forecast at time t . We discuss concrete examples for that in the next subsection.

The constraints from the Equations 1, 2, and 7 to 10 are mandatory for our model. A curve which does not fulfill all these constraints is either no deaths curve at all (if Equations 1 and/or 2 are not fulfilled) or not consistent with the forecasts of the future mortality scenarios (i.e.

⁹Börger et al. (2018) propose discrete estimators for the four statistics. Since we are in a continuous setting we had to adjust the estimators accordingly.

¹⁰Alternatively, this could also mean that the shape requirements (Equations 3 to 6) are too restrictive, i.e. a deaths curve that we regard as implausible from today's point of view because it violates one of these equations should nevertheless be considered feasible at a (probably far away) future point of time.

at least one of the Equations 7 to 10 is not fulfilled). The constraints from Equations 3 to 6 are helpful in choosing a plausible deaths curve but not necessary for an extrapolation of the deaths curve. In this sense these requirements can be extended or relaxed depending on the application of the model, the question at hand, or the user's view on what a plausible deaths curve should look like.

3.2 Basic Idea of Practical Implementation

In the previous subsection we have introduced the theoretical concept of our new model. Now we explain how it can be implemented in order to derive concrete mortality forecasts (i.e. for computing deaths curves for future years t , starting age x_0 , and statistics' forecasts $(M_t, UB_t, DoI_t, d(M)_t)$). First we need to specify a functional representation of the elements of the set of "typically shaped" deaths curves D_{t,x_0} . The functional representation needs to be chosen carefully since a restrictive and inflexible representation can imply that the set \overline{D}_{t,x_0} becomes empty too easily. This would for instance be the case if one considered only those deaths curve which correspond to a mortality curve according to the Gompertz law (Gompertz, 1825). In this case all deaths curves would be determined by only two parameters which makes it impossible in general to find deaths curves which fulfill Equations 7 to 10 simultaneously. Therefore, it appears more reasonable to use non-parametric representations, e.g. based on splines. In this case, the number and the degree of the splines as well as their positions obviously have an impact on whether the set \overline{D}_{t,x_0} is empty or not. Here, a reasonable trade-off between tractability of the representation and variability in the deaths curves it can generate is necessary. For the example applications in the subsequent section, we use a set of 21 B-Splines of polynomial degree five for the deaths curve representation. While the high polynomial degree ensures a reasonable degree of smoothness of the deaths curves, the large number of splines allows for sufficient flexibility in the deaths curves. This representation clearly outperforms the common parametric deaths or mortality curve representations in terms of flexibility in the generated deaths curves. In Appendix A we provide details on the spline representation and its estimation in forecasting future deaths curves.

Based on this representation we can determine whether the set \overline{D}_{t,x_0} is empty. In this case there is no reasonably shaped deaths curve for the year t . In the other case it usually contains infinitely many feasible deaths curves which differ only very slightly since they coincide with respect to the four statistics from Section 2.1. In order to determine a unique mortality forecast, one of the following criteria could be applied to pick a single deaths curve from \overline{D}_{t,x_0} :

- *Smoothness of the deaths curve*: One may pick the deaths curve from \overline{D}_{t,x_0} which leads to maximal smoothness according to some smoothness measure.
- *Minimum deaths curve changes over time*: One may choose the deaths curve which is "most similar" to a previous deaths curve. The similarity can be determined by any

distance measure.

In our algorithm we implicitly use the criterion of minimum deaths curve changes over time and choose the deaths curve with minimal distance to the deaths curve of the last year where we have observed data, i.e. the last year before the projection starts (see Appendix A for more details).

4 Applications of the New Model

In this section we apply the model to the data that had also been used in Section 2, i.e. Swiss females with starting age 60. For the first example, we extrapolate the most recent linear trends for each of the four statistics to the years between 2015 and 2070 and for each calendar year fit a deaths curve to the extrapolated statistics.

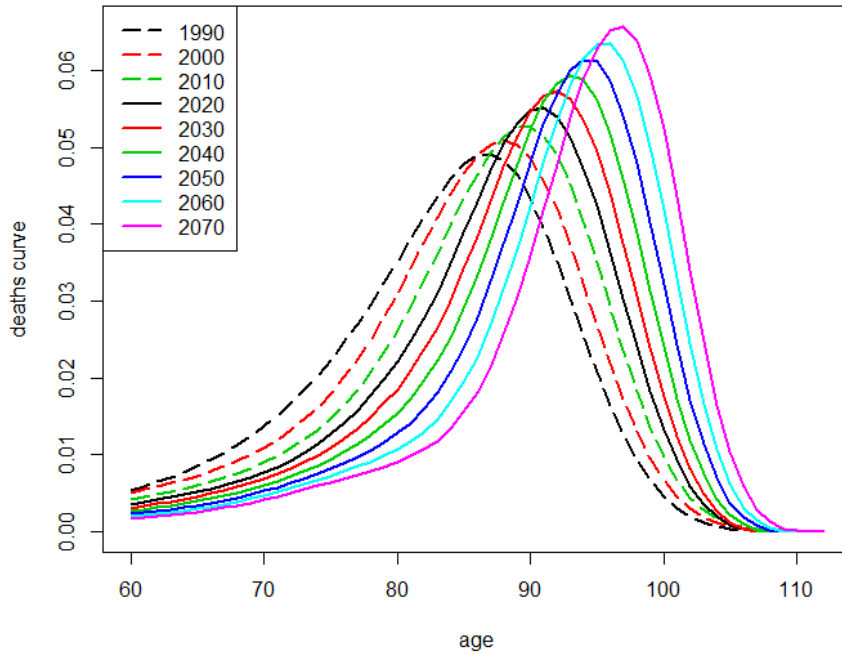


Figure 3: Deaths curves for Swiss females in the years 1990, 2000, 2010 (observed, dashed lines) and 2020, 2030, 2040, 2050, 2060, and 2070 (forecast with linear extrapolation of most recent trends, solid lines).

Figure 3 shows nine deaths curves for Swiss females between 1990 and 2070. Both for the historical deaths curves and for the extrapolated deaths curves we observe right-shifting mortality as the peaks of the deaths curves move more and more to the right. At the same time the upper bounds of the deaths curve's support also move to the right indicating extension. Also, the deaths curves from Figure 3 become more and more compressed between 1990 and 2070 since the extrapolation yields an increasing DoI implying compression. Finally, the number of deaths

in the modal age at death $d(M)$ increases over time as the peak of the deaths curves in this figure becomes ever higher.

It is particularly noteworthy that the forecast deaths curves between 2020 and 2070 smoothly continue the historical changes of the deaths curve between 1990 and 2010. The four trends (right-shifting mortality, extension, compression, concentration) are neither accelerated nor decelerated at or after the transition from the calibration to the forecasting period. Hence the mortality forecasts are consistent with the most recent observed demographic trends.

Of course, this new mortality model by definition extrapolates the four statistics it is based on in a reasonable manner. For the whole curve to be considered a reasonable forecast, also other quantities like probabilities of death should be plausibly forecast. Figure 4 shows forecasting results for $\log q_t(x)$ for selected ages from 60 to 100. We can see that these are reasonably extrapolated to the future.¹¹

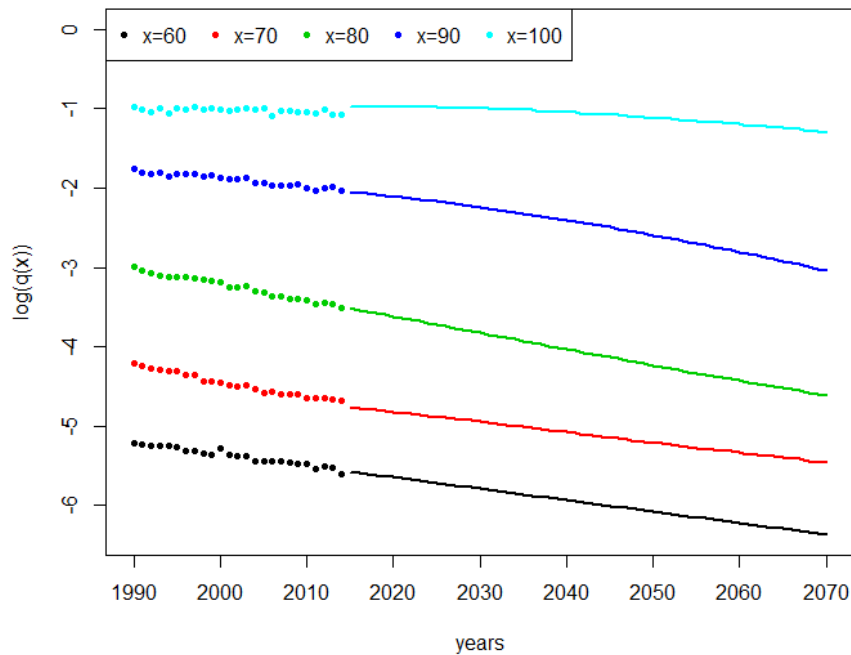


Figure 4: Logarithm of the probability of death for the ages $x = 60, 70, 80, 90$, and 100 between 1990 and 2014 (observed values) and between 2015 and 2070 (forecast values).

The demographic literature on mortality scenarios contains many different statistics which are designed to measure changes in the age distribution of deaths over time. In order to test our model for consistency also in terms of such other statistics we analyze their evolution between 1990 and 2070 (where from 2015 to 2070 we calculated the statistics based on our forecasts). As examples we considered the remaining period life expectancy at different ages ($e(60)$ in

¹¹We have performed similar analyses also for the force of mortality and the probability of survival. All were plausible. For the sake of brevity, we omit details.

particular), statistics of the C -family ($C50$ in particular; see Kannisto, 2000), the Inter-Quartile Range (IQR ; see Wilmoth and Horiuchi, 1999), graphical measures like the Moving Rectangle (MR ; see Wilmoth and Horiuchi, 1999), and the standard deviation of the age distribution at death above the modal age at death ($SD(M+)$; see e.g. Kannisto, 2001).

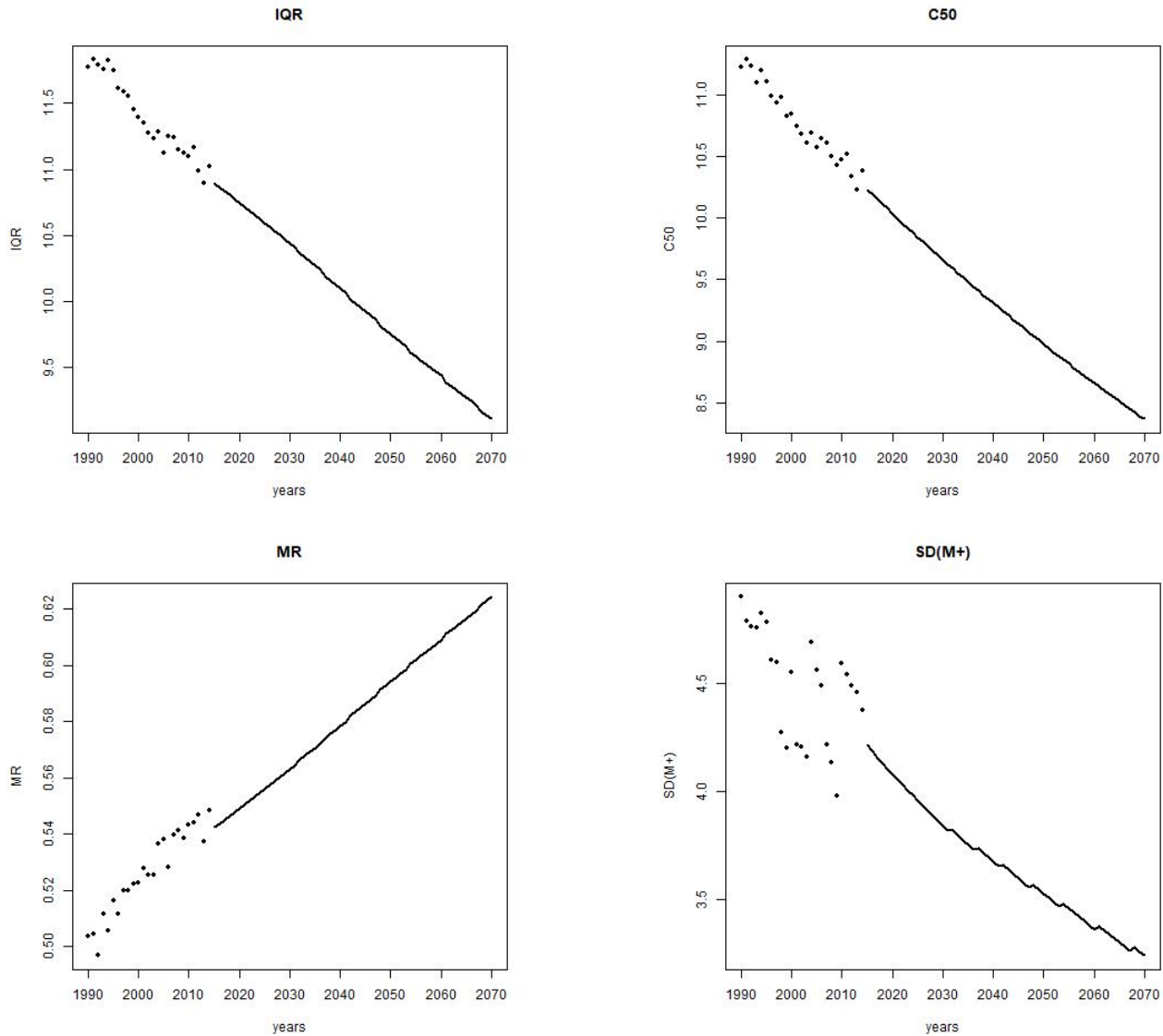


Figure 5: Alternative statistics between 1990 and 2014 (observed values) and between 2015 and 2070 (forecast values)

In Figures 5 (IQR , $C50$, MR , and $SD(M+)$) and 7 ($e(60)$ given by the black line) we illustrate some examples. Here we can see that the most recent trends in these statistics are reasonably

extrapolated until 2070.¹²

Since the four statistics used in our model allow for an intuitive interpretation, one might be interested in analyzing the effects resulting from the individual scenario components. For example the development displayed in Figure 3 is a "mixed scenario", where all four statistics of the classification framework of Börger et al. (2018) increase over time. We simultaneously observe right-shifting mortality, extension, compression, and concentration. In order to quantify the impact of each of the four scenario components separately, we consider so-called "pure scenarios" in which only one statistic changes with time (and the others remain constant).

Figure 6 displays deaths curves for each pure scenario in 2014 and the last future year where we still have reasonable deaths curves. Only in the pure right-shifting scenario we obtain deaths curves with reasonable shape until 2070. For all other scenarios the deaths curves' shapes in the pure scenarios become deformed such that before 2070 Equations 3 to 6 from Section 3.1 can not be fulfilled anymore. This is further evidence for the observation in Börger et al. (2018): pure scenarios may occur for a certain period of time, but, over a longer time period, typically only mixed scenarios can prevail.

Figure 7 shows the development of $e(60)$ in each of the pure scenarios as well as in the mixed scenario. In each pure scenario, the increase of the forecasts of $e(60)$ is slower than in the mixed scenario. The increase in $e(60)$ in the pure right-shifting scenario is faster than in any other pure scenario. This is an indication that in this example, the process of right-shifting mortality has a stronger impact on the increase of life expectancy at age 60 compared to the effects resulting from extension or compression.

As an alternative to the straightforward extrapolation of historic trends, the trends can easily be altered in our model. If, e.g., one has reason to believe that the trend in one or more of the four components will change at some point in time, one could either stop, intensify, reduce, or even reverse the increase / decrease in these components at one or more arbitrary points in time. Thus this model also allows for "what-if-analyses" for virtually any change of the mortality evolution in the future based on specifications of the four statistics. As an example for such an analysis, we use the base scenario from Figure 3 but double the intensity of right-shifting mortality and extension in 2015. Demographically this expert scenario means that the position of the deaths curve will change twice as fast in the future than recently observed, while trends in the deaths curve's shape remain unaltered. In this sense, this scenario is a stress scenario

¹²In total we considered 16 different statistics on the deaths curve: $C10$, $C50$, and $C90$ from the so-called C -family (see Kannisto, 2000), the remaining life expectancy at the ages 60, 70, 80, 90, and 100, the Fixed and Moving Rectangle, the Fastest Decline, the Sharpest Corner, the Quickest Plateau, as well as the Inter-Quartile Range (see Wilmoth and Horiuchi, 1999), the Prolate Index (see Eakin and Witten, 1995), and the standard deviation above the modal age at death (see Kannisto, 2001). We have obtained plausible results also for the statistics that are omitted in the figures for the sake of brevity.

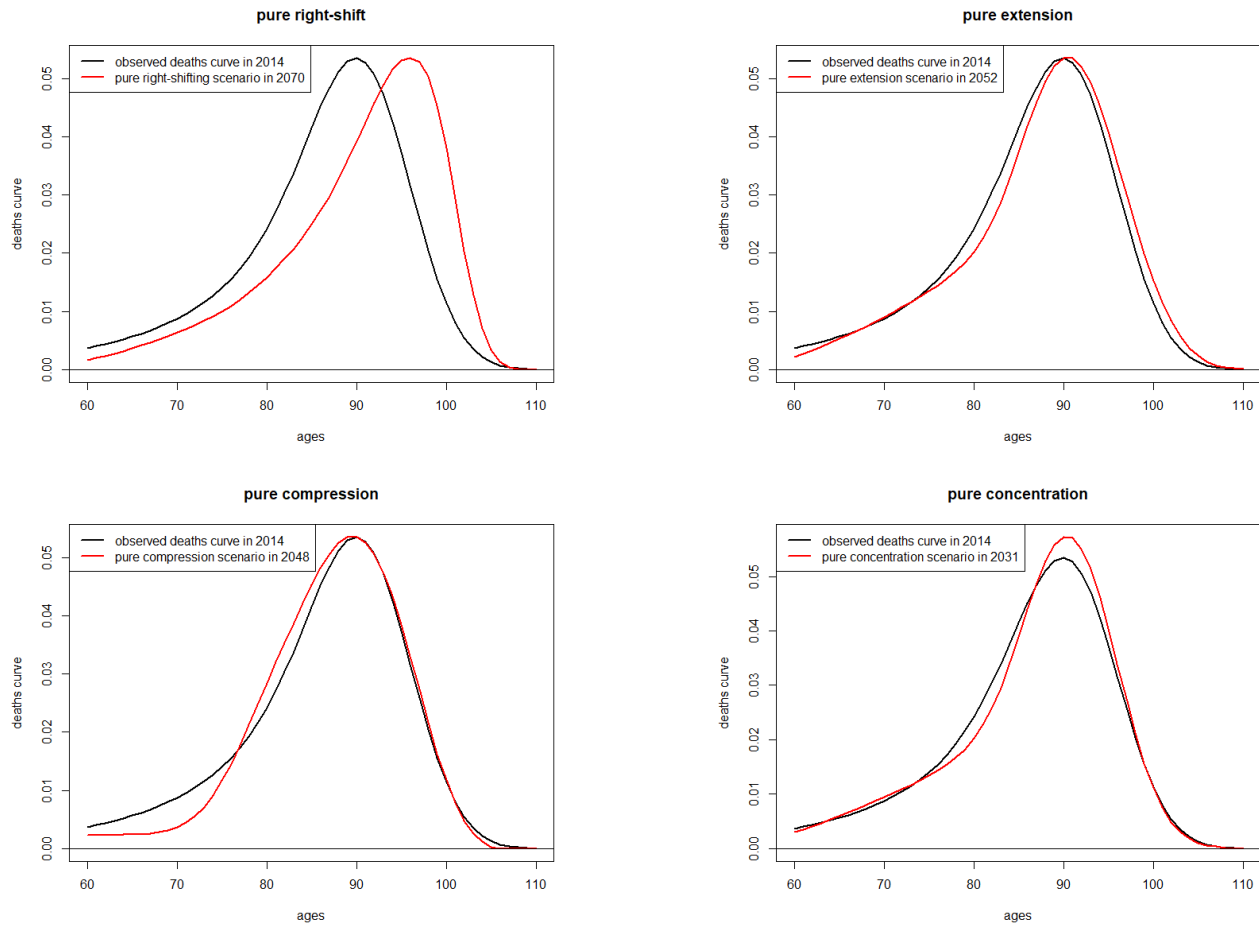


Figure 6: Observed deaths curve from 2014 and forecast deaths curve in each pure scenario; top-left: pure right-shifting scenario in 2070, top-right: pure extension scenario in 2052, bottom-left: pure compression scenario in 2048, bottom-right: pure concentration scenario in 2031.

for changes in the deaths curve's position.¹³ Note that for this scenario we obtain reasonable deaths curves until 2070. Also, in this scenario, the cohort life expectancy of a 60-year old Swiss female in 2015 (i.e. with year of birth 1954) would increase from 30.4 years to 32.1 years which is an increase of more than 5.4%.

5 Conclusion

This paper discusses the plausibility of mortality forecasts from a demographic perspective. In particular, trends in key demographic figures should be reasonably extrapolated into the future.

¹³It is noteworthy that such a stress scenario is not implausible. The observed trends for M and UB in Figure 1 have also exhibited strong changes in the past: While M increased by 0.21 years per calendar year between 1966 and 1992, its increase was only 0.12 years per calendar year thereafter. For UB we historically observed a stronger increase (0.09 years per calendar year from 1920 to 1976) than in the most recent period of stable trend (0.06 years per calendar year between 1976 and 2014).

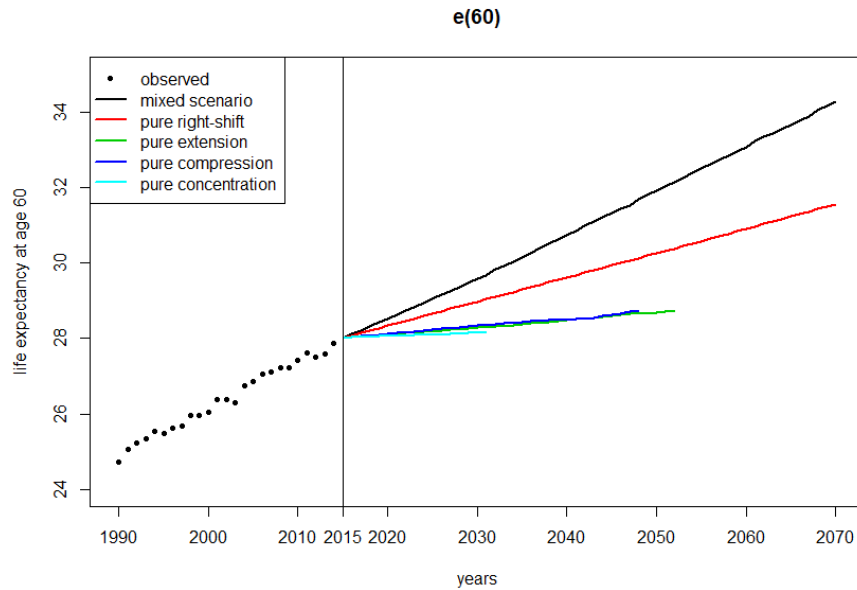


Figure 7: Observed and forecasted period life expectancy at starting age 60 in five different scenarios: mixed, pure right-shifting, pure extension, pure compression, and pure concentration scenario.

For purely statistical projection models whose parameters lack a clear demographic interpretation this is often not the case, and we provide examples for this. We apply the well-known models of Lee and Carter (LC) and Cairns, Blake, and Dowd (CBD) and analyze the resulting deaths curves in the classification framework of Börger et al. (2018). Here, deaths curves are classified based on four statistics which all have an intuitive demographic interpretation: the modal age at death M , the upper bound of the deaths curve's support UB , the degree of inequality DoI , and the number of deaths in the modal age at death $d(M)$. We find that both, the LC and the CBD model do not adequately extrapolate most recent historical trends in the four statistics. In most cases, we find jumps in the statistics at the transition from the historical to the forecast data in particular.

However, the classification framework of Börger et al. (2018) can not only be applied in order to analyze the demographic plausibility of existing mortality projections. Since in the framework two deaths curves are significantly different if and only if at least one of the four statistics differs, the framework can also be the basis for a projection model. Projections of future deaths curves (and thus essentially also any other mortality related quantity) can be derived from extrapolations of the four statistics. This ensures by construction that the mortality forecasts from this model are plausible from a demographic perspective. For the extrapolation of the statistics one may rely on their most recent observed trends or on experts' opinions on how the statistics may evolve in the future. Suitable additional constraints on the shape of forecast deaths curve need to be specified in order to ensure that the forecast deaths curves maintain a reasonable shape, and we show how this can be done in practice. We also provide a concrete

example forecast for Swiss females and illustrate its demographic plausibility by analyzing the evolution of additional statistics like the *IQR* over time.

The proposed projection model also allows for informative sensitivity analyses based on scenarios with a clear demographic interpretation. As (rather theoretical) examples, we consider the pure mortality scenarios in which only one of the four statistics evolves as in the most recent data and the other three statistics remain constant in the future. We find that such pure scenarios can only prevail for a certain period of time before the deaths curve's shape becomes implausible. As a possibly more realistic scenario, we also consider a 100% increase in the trends in M and UB compared to most recent historical trends (indicating increased right-shifting and extension of mortality). Here the remaining cohort life expectancy of 60-year olds increases by more than 5%. Such a scenario would be particularly critical from an annuity provider's perspective as the cohort life expectancy can be interpreted as an annuity present value with discount rate zero. Thus, the proposed projection model can be a useful risk management tool for quantifying the impact of demographic trends and its potential changes on the liabilities.

Furthermore, the proposed projection model can be a helpful addition to the toolkit of mortality projection models with respect to quantifying model risk as its structure and its forecasting approach are fundamentally different from those in established (statistical) models like the LC or the CBD model.

References

- Barbi, E., Lagona, F., Marsili, M., Vaupel, J.W., and Wachter, K.W. (2018). The plateau of human mortality: Demography of longevity pioneers. *Science* 360: 1459–1461. DOI: 10.1126/science.aat3119.
- Booth, H. and Tickle, L. (2008). Mortality Modelling and Forecasting: A Review of Methods. *Annals of Actuarial Science* 3(1-2): 3–43. DOI: 10.1017/S1748499500000440.
- Börger, M., Genz, M., and Ruß, J. (2018). Extension, Compression, and Beyond - A Unique Classification System for Mortality Evolution Patterns. *Demography* 55(4): 1343–1361. DOI: 10.1007/s13524-018-0694-3.
- Cairns, A.J.G., Blake, D., and Dowd, K. (2006). A Two-Factor Model for Stochastic Mortality with Parameter Uncertainty: Theory and Calibration. *The Journal of Risk and Insurance* 73(4): 687–718. DOI: 10.1111/j.1539-6975.2006.00195.x.
- Cairns, A.J.G., Blake, D., and Dowd, K. (2008). Modelling and Management of Mortality Risk: A Review. *Scandinavian Actuarial Journal* 2008(2-3): 79–113. DOI: 10.1080/03461230802173608.
- Dong, X., Milholland, B., and Vijg, J. (2016). Evidence for a limit to human lifespan. *Nature* 538(7624): 257–259.
- Eakin, T. and Witten, M. (1995). How Square is the Survival Curve of a Given Species? *Experimental Gerontology* 30(1): 33–64. DOI: 10.1016/0531-5565(94)00042-2.

- Feifel, J., Genz, M., and Pauly, M. (2018). *The Myth of Immortality: An Analysis of the Maximum Lifespan of US Females*. Preprint, Ulm University and Institute for Finance- and Actuarial Science, Ulm. URL: https://www.ifa-ulm.de/fileadmin/user_upload/download/forschung/2018_ifa_Feifel-et al_The-Myth-of-Immortality-An_Analysis-of-the-Maximum-Lifespan-of-US-Females.pdf.
- Fries, J.F. (1980). Aging, natural death, and the compression of morbidity. *New England Journal of Medicine* 303(3): 130–135.
- Gampe, J. (2010). Human Mortality Beyond Age 110. In: *Supercentenarians. Demographic Research Monographs (A series of the Max Planck Institute for Demographic Research)*. Ed. by H. Maier, J. Gampe, B. Jeune, Robine J.-M., and J.W. Vaupel. Berlin, Heidelberg: Springer: 219–230. DOI: 10.1007/978-3-642-11520-2_13.
- Gavrilov, L.A. and Gavrilova, N.S. (2011). Mortality Measurement at Advanced Ages: A Study of the Social Security Administration Death Master File. *North American Actuarial Journal* 15(3): 432–447. DOI: 10.1080/10920277.2011.10597629.
- Genz, M. (2017). *A Comprehensive Analysis of the Patterns of Worldwide Mortality Evolution*. Paper presented at the 2017 Living to 100 Society of Actuaries International Symposium, Orlando, FL. URL: <https://www.soa.org/essays-monographs/2017-living-to-100/2017-living-100-monograph-genz-paper.pdf>.
- Gompertz, B. (1825). On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Mode of Determining the Value of Life Contingencies. *Philosophical Transactions of the Royal Society of London* 115: 513–585.
- HMD (2015). *Human Mortality Database*. University of California, Berkeley, and Max Planck Institute for Demographic Research. URL: www.mortality.org.
- Kannisto, V. (2000). Measuring the compression of mortality. *Demographic Research* 3(6). DOI: 10.4054/DemRes.2000.3.6.
- Kannisto, V. (2001). Mode and Dispersion of the Length of Life. *Population: An English Selection* 13(1): 159–171. URL: <http://www.jstor.org/stable/3030264>.
- Lee, R.D. and Carter, L. (1992). Modelling and Forecasting U.S. Mortality. *Journal of the American Statistical Association* 87(419): 659–671. DOI: 10.1080/01621459.1992.10475265.
- Manton, K.G., Stallard, E., and Tolley, D. (1991). Limits to Human Life Expectancy: Evidence, Prospects, and Implications. *Population and Development Review* 17(4): 603–637. DOI: 10.2307/1973599.
- Oeppen, J. and Vaupel, J.W. (2002). Broken Limits to Life Expectancy. *Science* 296(5570): 1029–1031. DOI: 10.1126/science.1069675.
- Olshansky, S.J., Carnes, B.A., and Cassel, C. (1990). In Search of Methuselah: Estimating the Upper Limits to Human Longevity. *Science* 250(4981): 634–640. DOI: 10.1126/science.2237414.
- Pollard, J.H. (1987). Projection of Age-Specific Mortality Rates. *Population Bulletin of the United Nations* 21/22: 55–69.

- Schoenberg, I.J. (1946). Contributions to the problem of approximation of equidistant data by analytic functions. *Quarterly of Applied Mathematics* 4(1–2): 45–99, 112–141.
- Villegas, A.M., Kaishev, V.K., and Millossovich, P. (2018). StMoMo: An R Package for Stochastic Mortality Modeling. *Journal of Statistical Software* 84(3): 1–38. DOI: 10.18637/jss.v084.i03.
- Wilmoth, J.R. and Horiuchi, S. (1999). Rectangularization Revisited: Variability of Age at Death Within Human Populations. *Demography* 36(4): 475–495. DOI: 10.2307/2648085.

Appendices

A Aspects of the Practical Implementation of the Model

A.1 Deaths curve representation by B-splines

For the examples in this paper we use a B-Spline representation for the deaths curve (Schoenberg, 1946). The deaths curve in any year t is described as a linear combination of 21 polynomial spline functions $b_t^{(j)}(x)$, $j \in \{1, \dots, 21\}$ of degree five:

$$\hat{d}_t(x) = \sum_{j=1}^{21} a_t^{(j)} \cdot b_t^{(j)}(x) = B_t(x) * a_t, \quad (11)$$

where $a_t = (a_t^{(1)}, \dots, a_t^{(21)})^T$ is the vector of spline weights and the vector-valued function $B_t : \mathbb{R} \rightarrow \mathbb{R}^{21}$ gives the value of each spline at any age $x \in [x_0, UB_t]$. Each spline $b_t^{(j)}(x)$ is centered at its so-called knot $k_t^{(j)}$, $j \in \{1, \dots, 21\}$ and, since the polynomial degree is odd, symmetric around its knot. Furthermore, it is different from zero only on a certain interval:

$$b_t^{(j)}(x) \begin{cases} > 0 \text{ for all } x \in (\max \{x_0, k_t^{(j-3)}\}, \min \{UB_t, k_t^{(j+3)}\}) \\ = 0 \text{ for all } x \notin (\max \{x_0, k_t^{(j-3)}\}, \min \{UB_t, k_t^{(j+3)}\}) \end{cases} \quad (12)$$

In our forecasting model, deaths curves and thus also the spline functions need to be continuously differentiable at least three times (see Section 3.1), but differentiability in higher order is desirable. The polynomial degree of five therefore is a reasonable choice in our examples, but may of course be altered for other applications of the model.¹⁴

The number of splines essentially determines the flexibility of the deaths curve representation. The larger the number of splines, the more likely we are to find a deaths curve which matches a concrete forecast for the four statistics M_t , UB_t , DoI_t , and $d(M)_t$. The specific number of 21 splines is a result of the chosen positioning of the spline knots within the interval $[x_0, UB_t]$ (see below). We found that a number in that range is necessary to have a sufficiently flexible representation. Note that we fix the number of splines independently of the length of the deaths curve's support. Thus in case of extension or contraction (i.e. when UB increases or decreases), we reposition the spline knots over time, but do not change the number of splines.

In order to specify a concrete spline representation of a deaths curve, the spline knot positions $k_t^{(j)}$, $j \in \{1, \dots, 21\}$ and the spline weights $a_t = (a_t^{(1)}, \dots, a_t^{(21)})^T$ need to be determined such that Equations 1 to 10 are fulfilled. In what follows, we explain these two steps in more detail.

¹⁴We also considered the more commonly applied third degree polynomials, but found that deaths curves can become slightly "wavy".

A.2 Positioning of spline knots

Given a forecast of the four statistics $(M_t, UB_t, DoI_t, d(M)_t)$ for any year t , we determine the positions of the spline knots as follows: First, we set one knot at each endpoint of the deaths curve's support, i.e. at x_0 and at UB_t , and at the modal age at death M_t . Between x_0 and M_t we add nine equidistant knots and between M_t and UB_t we add five equidistant knots.¹⁵ For sufficient flexibility also at the boundaries of the deaths curve's support we need two additional knots left of x_0 and right of UB_t . These knots are positioned such that all knots left or right of M_t , respectively, are equidistant. Thus, the total number of splines is $3 + 9 + 5 + 2 \cdot 2 = 21$. Figure 8 illustrates the 21 splines before weighting (see next paragraph) and their knot positions. The two splines at each boundary whose maxima are not yet visible have their knots outside the support $[x_0, UB_t]$.

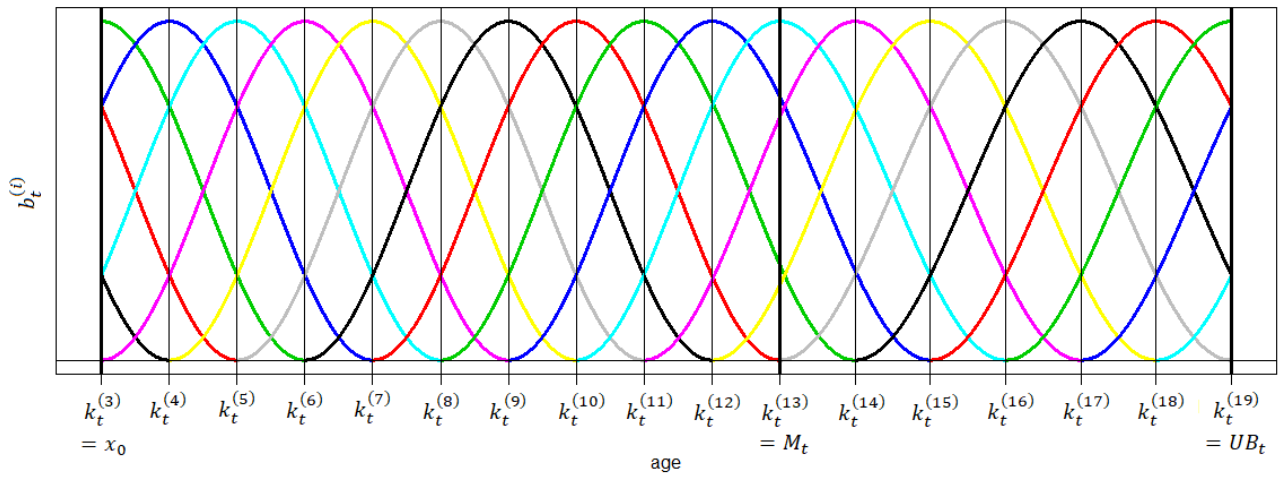


Figure 8: Knots and polynomial functions of degree five for 21 splines.

A.3 Determination of spline weights

Now, the spline weights need to be chosen for any future year t such that the resulting linear combination of the 21 polynomial spline functions (see Equation 11) results in a deaths curve that matches the given four statistics $(M_t, UB_t, DoI_t, d(M)_t)$. For the determination of the vector of spline weights a_t we use numerical methods: For any t the Equations 1 and 6 to 10 from Section 3.1 can be translated into a system of non-linear, but convex equalities:

$$A_t^{(1)}(a_t) = y_t^{(1)}, \quad (13)$$

where $A_t^{(1)} : \mathbb{R}^{21} \rightarrow \mathbb{R}^6$ is a corresponding function in the spline weights and $y_t^{(1)}$ is a vector of the target values. Analogously, the Equations 2 to 5 from Section 3.1 define a system of linear

¹⁵It might be reasonable to modify the respective number of knots (nine or five) if the distances between x_0 and M_t or between M_t and UB_t are rather small or extremely large.

(and thus also convex) inequalities:

$$A_t^{(2)} * a_t < y_t^{(2)}. \quad (14)$$

These constraints are continuous in the sense that they refer to each (also non-integer) age in certain intervals of the support $[x_0, UB_t]$. For the numerical implementation we need to "discretize" the constraints by evaluating them only at a finite number of ages. It turned out to be sufficient to only evaluate the constraints at the spline knots.

In order to obtain starting values for the algorithm described below, we fit a spline representation to the observed deaths curve of the most recent year t_0 for which data is available. To this end we perform a least square estimation where we require the constraints from Equations 1 to 5 to be fulfilled.

The basic idea behind the algorithm is that – for any future year t – we search for spline weights a_t which solve the Equations 13 and 14 and differ as little as possible from the weights a_{t_0} . Thus, in terms of picking a single deaths curve from \bar{D}_{t,x_0} we opt for minimum deaths curve changes over time (see Section 3).¹⁶ The algorithm proceeds as follows:

1. In a first step we ensure that the curve is an element of the space \mathcal{D}_{t,x_0} . To this end, we scale the vector of spline weights, i.e.

$$\tilde{a}_t = \frac{1}{\int_{x_0}^{UB_t} B_t(x) * a_{t_0} dx} \cdot a_{t_0},$$

such that Equation 1 holds.

2. In the second step, we alter the vector of spline weights such that the deaths curve is an element of the set D_{t,x_0} . To this end, we check if \tilde{a}_t solves Equation 14. If it does, we proceed with the third step. Otherwise, we carry out the following steps a to c which build on the method of gradient descent:

- (a) Based on Equation 14 we define a function f_t of a vector of weights a :

$$f_t(a) = \sum_{j=1}^m w_j \cdot A_t^{(2,j)} * a,$$

where m is the number of rows of the matrix $A_t^{(2)}$, $A_t^{(2,j)}$ is the j^{th} row of the matrix

¹⁶By this algorithm we find deaths curves with minimal distance from the observed deaths curve in the year t_0 . The algorithm would also work iteratively, i.e. by using the weights a_{t-1} as starting values for year t . However, if one was only interested in a forecast for year t , one would have to forecast deaths curves for all years up to year t in that case.

$A_t^{(2)}$, and the weights w_j are given by¹⁷

$$w_j = \begin{cases} 1 & \text{if } A_t^{(2,j)} * a < 0 \\ 10^6 & \text{if } A_t^{(2,j)} * a \geq 0 \end{cases}.$$

- (b) We determine the gradient $\nabla f_t(a)$ of the function $f_t(a)$ and, following the method of gradient descent, update the vector of weights $\tilde{a}_t^{(\nu)}$ via

$$\tilde{a}_t^{(\nu)} = \tilde{a}_t^{(\nu-1)} - \theta \cdot \nabla f_t(\tilde{a}_t^{(\nu-1)}).$$

The step size θ is determined by minimizing the difference $f_t(\tilde{a}_t^{(\nu-1)}) - f_t(\tilde{a}_t^{(\nu)})$ under the constraint that no additional component of the vector $A_t^{(2)} * \tilde{a}_t^{(\nu)}$ becomes positive. In the first iteration we set $\tilde{a}_t^{(0)} = \tilde{a}_t$ and $\nu = 1$.

- (c) We update \tilde{a}_t as long as there are positive entries in the vector $A_t^{(2)} * \tilde{a}_t$. After the final update we set $\tilde{a}_t = \tilde{a}_t^{(\nu)}$.

3. In the third step, we modify the vector of spline weights such that the deaths curve is an element of the set \overline{D}_{t,x_0} . Similarly to step 2.a, we define a function g_t of the vector of weights a ,

$$g_t(a) = \sum_{j=1}^k \omega_j \cdot (A_t^{(1,j)}(a) - y_t^{(1,j)})^2,$$

where k is the number of equality constraints, ω_j is a scalar weight for the j^{th} equality constraint, $A_t^{(1,j)}(a)$ is the left hand side value of Equation 13 for the j^{th} equality constraint, and $y_t^{(1,j)}$ is the right hand side value of Equation 13 for the j^{th} equality constraint. In order to determine the root of the function g_t , we apply the method of gradient descent analogously to the optimization of the function f_t . Starting with the vector of spline weights \tilde{a}_t from Step 2, we search for the vector of spline weights a_t for which the value of the function g_t assumes its minimum value zero.¹⁸

The resulting deaths curve $\hat{d}_t(x) = B_t(x) * a_t$ is an element of the set \overline{D}_{t,x_0} and thus a valid forecast of our model.

¹⁷The constant 10^6 in the following formula stresses the constraints which are not yet fulfilled and therefore accelerates convergence in the subsequent steps. We found 10^6 to be in a reasonable range.

¹⁸For numerical reasons the algorithm terminates as soon as $g_t(\tilde{a}_t) < e^{-7}$.

Curriculum Vitae

Martin Genz
born 1987 in Ulm

Higher Education

Since 09/2013	Ulm University , Ulm Doctoral candidate
10/2011 - 08/2013	Ulm University , Ulm Master of Science in Mathematics and Management
10/2008 - 09/2011	Ulm University , Ulm Bachelor of Science in Mathematics and Management
09/1998 - 06/2007	Robert-Bosch-Gymnasium , Langenau Abitur

Positions and Practical Experience

Since 09/2013	Institut für Finanz- und Aktuarwissenschaften , Ulm Actuarial Consultant
06/2012 - 08/2013	Lebensversicherung von 1871 a.G. , München Student assistant
03/2012 - 04/2012	Lebensversicherung von 1871 a.G. , München Internship (product development)
08/2011 - 08/2013	Institut für Finanz- und Aktuarwissenschaften , Ulm Student assistant
03/2011 - 04/2011	Allianz Deutschland AG , Stuttgart Internship (actuarial office)
04/2010 and 10/2010 - 07/2011	Ulm University , Ulm Student assistant

Scientific Publications and Working Papers

- 2017 **A Comprehensive Analysis of the Patterns of Worldwide Mortality Evolution**
Published in the monograph to the 2017 Living to 100 Society of Actuaries International Symposium, Orlando, FL. Available at <https://www.soa.org/essays-monographs/2017-living-to-100/2017-living-100-monograph-genz-paper.pdf>.
- 2018 **Extension, Compression, and Beyond – A Unique Classification System for Mortality Evolution Patterns** (with Matthias Börger and Jochen Ruß)
Demography 55(2): 1343-1361.
- 2018 **The Myth of Immortality: An Analysis of the Maximum Lifespan of US Females** (with Jan Feifel and Markus Pauly)
Working Paper.
- 2019 **The Future of Mortality – Mortality Forecasting by Extrapolation of Deaths Curve Evolution Patterns** (with Matthias Börger and Jochen Ruß)
Working Paper.

Research Conferences (with Contributed Talks)

- 06/2015 2nd Annual International Conference on Demography and Population, Athens (Greece)
- 09/2015 11th International Longevity Risk and Capital Market Solutions Conference, Lyon (France)
- 02/2016 7th Demographic Conference of “Young Demographers”, Prague (Czech Republic)
- 01/2017 2017 Living to 100 Society of Actuaries International Symposium, Orlando, FL (USA)
- 10/2017 IAALS Colloquium, Barcelona (Spain)
- 03/2018 Research Workshop IVW/ifa, Ulm (Germany)
- 06/2018 31st International Congress of Actuaries (ICA) 2018, Berlin (Germany)
- 03/2019 Research Workshop IVW/ifa, Ulm (Germany)

Ulm, Juli 2019

Martin Genz