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

Matthias Börger
Institut für Finanz- und Aktuarwissenschaften (ifa)
Lise-Meitner-Straße 14
89081 Ulm, Germany

Martin Genz (corresponding author)
Institut für Finanz- und Aktuarwissenschaften (ifa) & Institut für Versicherungswissenschaften, Universität Ulm
Lise-Meitner-Straße 14
89081 Ulm, Germany
phone: +49 731 20 644 264
fax: +49 731 20 644 299
email: m.genz@ifa-ulm.de

Jochen Ruß
Institut für Finanz- und Aktuarwissenschaften (ifa) & Institut für Versicherungswissenschaften, Universität Ulm
Lise-Meitner-Straße 14
89081 Ulm, Germany

Abstract

There exists a variety of literature on the question how the age distribution of deaths changes over time as life expectancy increases. However, corresponding terms like extension, compression, or rectangularization are sometimes defined only vaguely, and statistics used to detect certain scenarios can be misleading. The matter is further complicated since often mixed scenarios prevail and the considered age range can have an impact on observed mortality patterns.

In this paper, we establish a unique classification framework for realized mortality scenarios building on four statistics which describe changes of the deaths curve over time. The statistics determine whether elements of extension or contraction, compression or decompression, left or right shifting mortality, and what we call concentration or diffusion are present. We can identify not only pure, but also mixed scenarios where two or more statistics change at the same time. Furthermore, the framework can not only test the presence of a particular scenario, but also assign a unique scenario to any observed mortality evolution. It can detect different mortality scenarios for different age ranges in the same population. We present a methodology for the implementation of our classification framework and apply it to mortality data for US females.

Key words

Mortality scenario classification, longevity, rectangularization, shifting mortality

1. Introduction

Mortality evolutions, i.e. realized changes in mortality rates, have been analyzed extensively in the last decades. These analyses go far beyond determining trends in the evolution of life expectancy. For instance, for social security systems, pension funds, and life insurers, it is not sufficient to know people's average lifetime. They also need to know the distribution of lifetimes in order to assess how much uncertainty is involved in their planning and reserving. In this sense, changes in aggregated statistics like life expectancy are simply a consequence of the underlying change of the age distribution of deaths.

There is a wide range of literature on the question how realized mortality 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, e.g. rectangularization, compression, extension, expansion, and shifting mortality to mention only the most important ones. These terms have been helpful in the analysis of historical mortality evolution patterns. Their definitions are, however, mostly intuitive which can lead to ambiguity in practical applications. For instance, Fries (1980) defines 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 Section 2, this definition can be misleading. Similarly intuitive but also 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 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 use certain statistics of the deaths curve, i.e. the age distribution of deaths, to define compression. According to Kannisto (2001), 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, decreases at the same time. This definition is (at least implicitly) applied by many other authors, e.g. Cheung and Robine (2007) or Ouellette and Bourbeau (2010). Wilmoth and Horiuchi (1999), on the other hand, identify compression by a shrinking inter-quartile range (IQR), i.e. the length of the age range between the 25th and the 75th percentile of the distribution of deaths. Analogously, Kannisto (2000) uses the so called $C\alpha$-statistics, i.e. the shortest age range in which $\alpha\%$ of all deaths occur. Thatcher et al. (2010) observe compression if the slope parameter in a logistic mortality model increases with time. Although these different definitions of compression yield the same results in many cases, this is not always true (see Section 2).

Scenario definitions can also be critical when they only rely on observations for a rather small age range. For instance, when analyzing the evolutions of M and $SD(M+)$, one completely ignores the mortality evolution for all ages below M. As we show in an example in Section 2, 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) states that rectangularization should be "best thought of as 'compression of mortality'". Also for Myers and Manton (1984), compression and rectangularization 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) use the terms expansion and shifting mortality but do not explain the differences between them. Others define the three terms in different ways: Wilmoth and Horiuchi (1999) use the term expansion if the force of mortality decreases faster for older ages than for younger ages; Bongaarts (2005), on the other hand, explains the scenario of shifting mortality as a result of “delays in the timing of deaths”, i.e. the force of mortality curve exhibits simply a shift in age; in Cheung et al. (2005) the term “longevity extension” is used for a scenario, where 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) regards shifting mortality as a scenario “where the compression of mortality has stopped”. Obviously, such a definition implies a distinction between the scenarios, i.e. shifting mortality and compression in this case. However, this also rules out mixed scenarios by definition. As we will see later, elements of different mortality scenarios can often be identified at the same time. Therefore, analyses which solely focus on testing for one particular scenario, e.g. compression, can never provide a comprehensive insight into the mortality evolution.

In this paper, 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 provide precise definitions of scenarios and show how they can be identified. Our methodology can deal with mixed scenarios and 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 suits best 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 where more and more deaths get shifted from younger to older ages, but where deaths become more and more evenly spread within 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 Section 2).

The remainder of this paper is organized as follows: Section 2 illustrates different issues in identifying mortality scenarios and sets out the requirements for a new classification framework. We establish a framework that allows for a unique classification of mortality scenarios and a clear definition of the different patterns in Section 3. The framework is based on changes in the shape of the deaths curve over time. Section 4 discusses the implementation of the framework. In particular, statistics for measuring relevant changes in the shape of the deaths curve are discussed, and a method for detecting trend changes is introduced. For illustrative purposes, we then present an application to the mortality evolution of US females in Section 5. Finally, Section 6 concludes.

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 that will be developed in Section 3.

2.1. Imprecise Mortality Scenario Definitions

Mortality scenarios describe patterns in the evolution of mortality over time, i.e. a process of change. However, in the literature we find several attempts to define this process solely by some (often only theoretical and hence unreachable) final stage that is approached. A classical example is the definition of rectangularization as a process where the survival curve approaches a rectangular form. However, a rectangular form can be reached “on different routes”.

This is illustrated by the left panel of Fig. 1 which shows a hypothetical, but not unrealistic evolution of deaths curves $d(x)$ over time.[1] Assume that at some point in time, mortality in some population follows the curve labeled "State 1". At some later point in time, it follows "State 2", etc., 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 which is somewhat the opposite of compression can be observed.

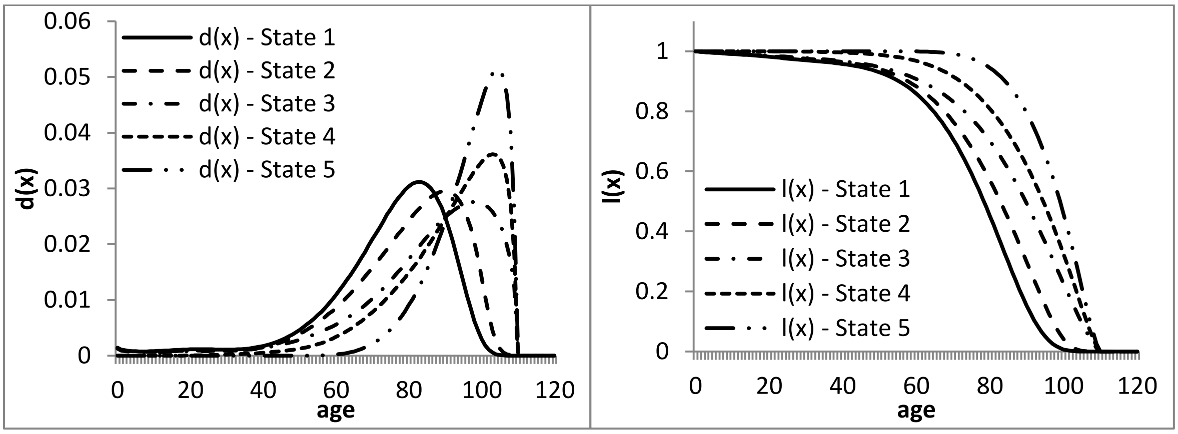

Fig. 1 Mortality evolution in a hypothetical example. Left: deaths curves; right: survival curves

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, 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.

For the evolution in Fig. 1, we can 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, Ouellette and Bourbeau (2010), this would mean that rectangularization prevails throughout the process which is somewhat consistent with the intuition from looking at the right panel of Fig. 1. Other authors propose different statistics. The so-called Prolate index PI as used e.g. by Wilmoth and Horiuchi (1999) decreases between States 1 and 3 (from 0.958 in State 1 to 0.946 in State 3) and then increases between States 3 and 5 (to 0.983 in State 5) indicating that the pattern in the first phase is different from the second phase, which fits the intuition from the left panel of Fig. 1.

So, one can conclude that the definition of a mortality scenario by some theoretical final state that is being approached will not always lead to a unique result. Also, even statistics that are frequently used to describe a mortality evolution can lead to contradictory results.

In a second example, we now show that it may not be appropriate to define a certain mortality scenario as the opposite of some other scenario. A classical example is the relationship between shifting mortality[2] (or alternatively extension) and compression. The left panel of Fig. 2 shows a "mixed scenario", where (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

[1] All deaths curves in this paper 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 Sections 2 and 3 are hypothetical, but reasonable since overall mortality improves and life expectancy increases.

[2] In the introduction we have pointed out that the terms expansion, extension, and shifting mortality co-exist 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 (see below).

other. Analogously, in the right panel of Fig. 2 neither shifting mortality nor compression can be observed. Thus, rejection of one scenario does not imply that the other scenario prevails. So clearly it is not possible to classify all possible mortality evolutions using disjoint categories such as compression and shifting mortality. This again shows the need for a more sophisticated classification system.

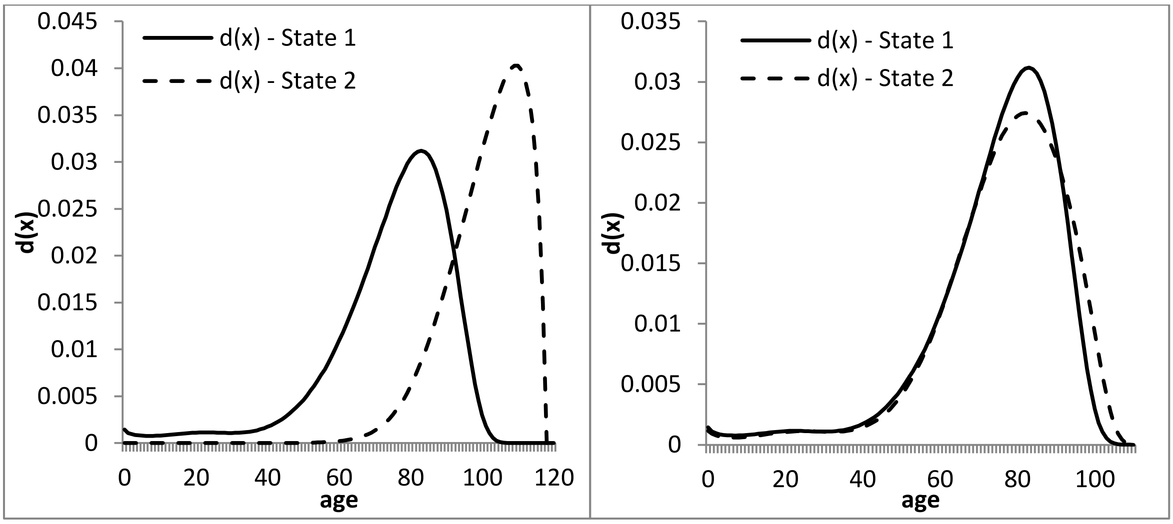

Fig. 2 Two hypothetical examples. Left: shifting mortality and compression coexist; right: neither shifting mortality nor compression exists

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 always leads to a loss of information. Therefore, one should very carefully identify statistics that preserve that part of the information one is interested in. 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.

First, we would like to illustrate that the presence or absence of compression cannot always be derived from the statistics M and $SD(M+)$ alone. Figure 3 shows a mortality evolution from State 1 to State 2 that would be classified as compression or rectangularization by looking at these two statistics alone since the modal age at death M increases (from 83 to 95) and $SD(M+)$ decreases (from 7.6 to 4.5). Fries (1980) even proposed that the process of rectangularization has reached its limit once the standard deviation of ”natural deaths” has reached a value of 4. Although this is almost the case for State 2, the deaths curve does not look more dense as we move from State 1 to State 2 (the number of deaths at the modal age even declines).

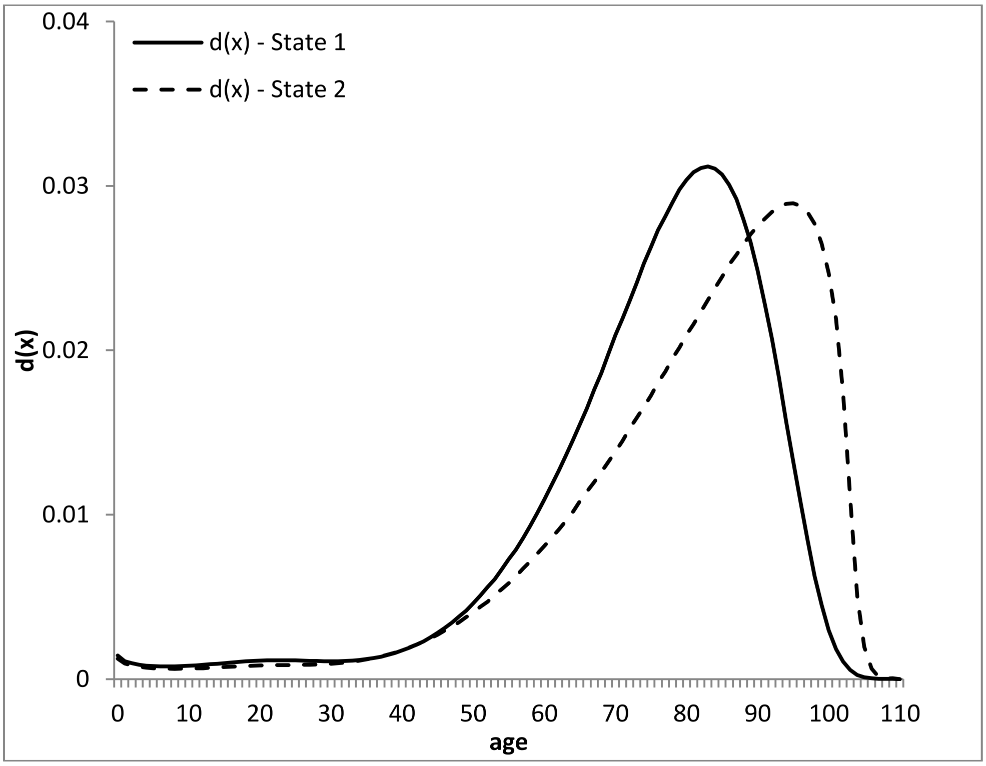

Fig. 3 Mortality evolution without compression where frequently used statistics would identify compression

In the next examples, we deal with statistics that are based on percentiles of the mortality distribution: the inter-quartile range IQR (see e.g. Wilmoth and Horiuchi 1999) and the so-called $C\alpha$-statistics (see Kannisto 2000).

Often, compression is defined by a reduction of the IQR and/or a $C\alpha$-statistic (see Kannisto 2000; Wilmoth and Horiuchi 1999). Figure 4 shows two scenarios of mortality evolution where the structure of the mortality distribution has changed considerably from State 1 to State 2, with clear characteristics of mortality improvement and compression. However, in the left panel the IQR remains unchanged, whereas in the right panel, $C50$ remains at the same value. Thus, neither IQR nor $C50$ alone are able to identify compression.

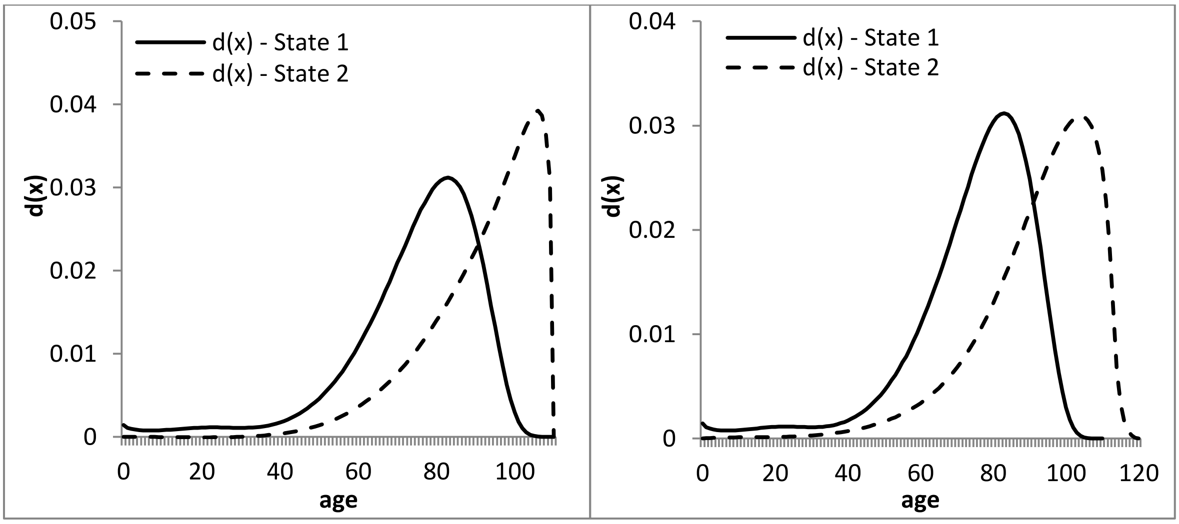

Fig. 4 Mortality evolutions with compression. Left: unchanged IQR; right: unchanged $C50$

2.3. Impact of Age Range

Sometimes, different types of mortality evolution occur in different age ranges. Myers and Manton (1984) analyze the survival curve starting at age 0 and compare it to the survival curve starting at age 65 for US females and males between 1962 and 1979. For the entire age range, they observe a clear tendency toward rectangularization, while they do not observe any indication for rectangularization in the older ages. If one is interested primarily in a certain age range (e.g. old age mortality) one should therefore only consider the mortality evolution in the corresponding age range (e.g. from age 65 onward). This might be of interest to social security systems, pension funds, or annuity providers since they are affected primarily by mortality changes beyond the retirement age.

However, when restricting the age range, undesired effects may occur whenever statistics are being used which depend on the number of people being alive at the beginning of the considered age range, for example $d(M)$, i.e. the number of deaths at age M. Assume 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 that one is interested in and a change in younger age mortality that one is not interested in. 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. 5 shows some mortality evolution over the entire age range. Here, clearly compression towards higher ages can be observed. If one is only interested in the age range 65+, one might intuitively look at the respective age range of the left panel of Fig. 5 (i.e. without normalizing) which displays signs of compression. However, in the normalized curves (right panel of Fig. 5) the deaths curve of State 2 looks less dense than for State 1 which is an indication against compression.

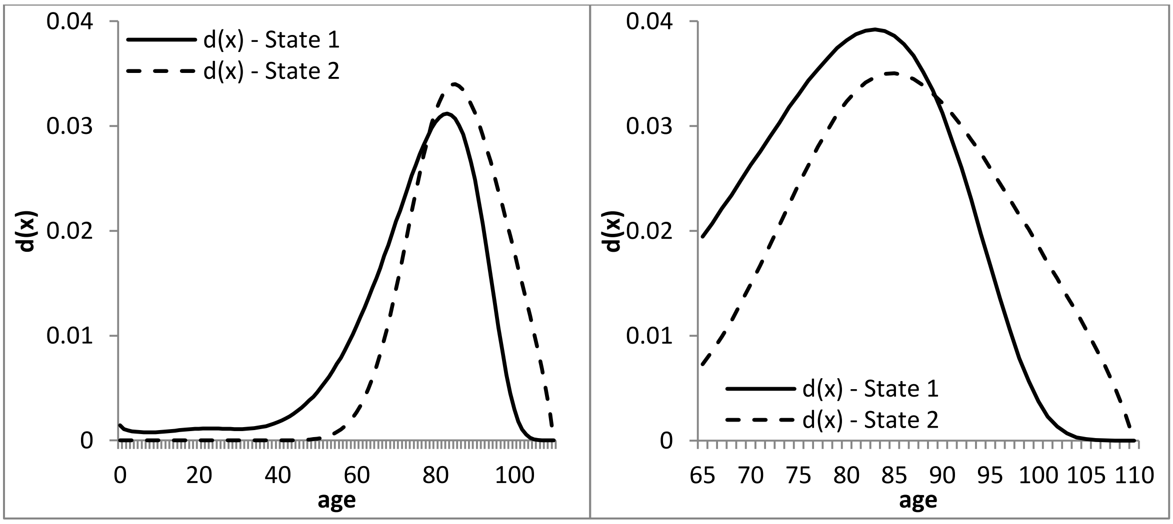

Fig. 5 Mortality evolution with increasing number of survivors to age 65. Left: complete age range; right: starting at age 65 with normalized l(65)

3. A New Classification Framework for Mortality Scenarios

In the previous section, we have identified shortcomings of existing approaches for the classification of mortality evolutions. We will now propose a new framework where unique mortality scenarios are defined based on observable changes in the shape of the deaths curve. Note that in this section, we introduce the "intellectual concept" of the framework whereas 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 classification framework can be applied for any age range which includes the right tail of the deaths curve. Depending on the question at hand, the age range could start e.g. at zero, some juvenile age, or the retirement age. In particular, it is possible that the classification framework identifies different mortality scenarios for different age ranges (see Section 5 for an example).

For any given age range, we will use four key characteristics of the deaths curve: the position of the deaths curve's peak, the support of the deaths curve, the "degree of inequality" in the distribution of deaths between ages, and the height of the peak of the deaths curve. Significant changes in one or several characteristics over time mean that the position and/or the shape of a deaths curve have 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 different deaths curves and to uniquely classify mortality scenarios.

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.[3] Thus, an increase in M indicates *right shifting mortality,* while a decrease in M implies *left shifting mortality*.

The support of a deaths curve is determined by its upper bound, which we refer to as UB. We assume that UB always exists. Typical deaths curves clearly convey the impression that it does, although this might be debatable from a theoretical point of view. In theory, UB can only exist if the probability of death reaches one for some age. If the probability of death remains below one for all ages, any age could be reached in principle.[4] Estimating UB in practice certainly involves some ambiguity, and we refer to Section 4 for more details. For now, we assume that UB exists and that it can be determined or at least approximated. We denote the respective changes of UB as *extension* (if UB increases over time) and *contraction* (if UB decreases over time).

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. 6 shows two deaths curves which are significantly different although the other three statistics of our framework coincide. Therefore, an additional statistic is required which is related to the shape of the curve. The deaths curve of State 2 is almost zero 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. Intuitively, DoI is the "distance" between a given deaths curve and a hypothetical deaths curve where the number of deaths is the same for all ages. If this distance becomes larger, the distribution of deaths over all ages becomes less uniform (and hence DoI increases). As for UB, we assume in this section that DoI can be determined and refer to Section 4 for an exact definition.

If DoI increases, there must be at least one part of the age range where deaths get more compressed. Such a *compression* typically occurs around M, but it might occur at other ages as well. If DoI decreases over time, this means that the distribution of deaths becomes more balanced and that a *decompression* can be observed. Note though, that compression and

[3] Since the shape of a deaths curve can and typically does change over time, it is hardly possible to determine a pure shift of the entire deaths curve. Therefore, we consider the deaths curve's "center" M.

[4] Research by several authors, see e.g. Gampe (2010), indicates that probabilities of death typically flatten out at very old ages, possibly somewhere around 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 at all.

decompression in this sense relate to the entire age range under consideration. Locally, e.g. around M, deaths can still get more concentrated. This once again shows how important it is to restrict the age range to the part one is actually interested in.

Finally, the height of the peak of a deaths curve is given by $d(M)$. In this section, we assume that the modal age at death can be determined uniquely.[5] This component addresses the evolution of a deaths curve at and around its "center" M. Note that 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 local measure for a small region around M, whereas DoI describes the equality/inequality of the distribution of deaths over the entire age range under consideration. An increase in $d(M)$ indicates that the distribution of deaths becomes more concentrated around M. This *concentration*, as just outlined, can also occur in combination with a decompression over the whole age range under consideration. The counterpart to *concentration* is what we refer to as *diffusion*, and it is observed if $d(M)$ decreases.

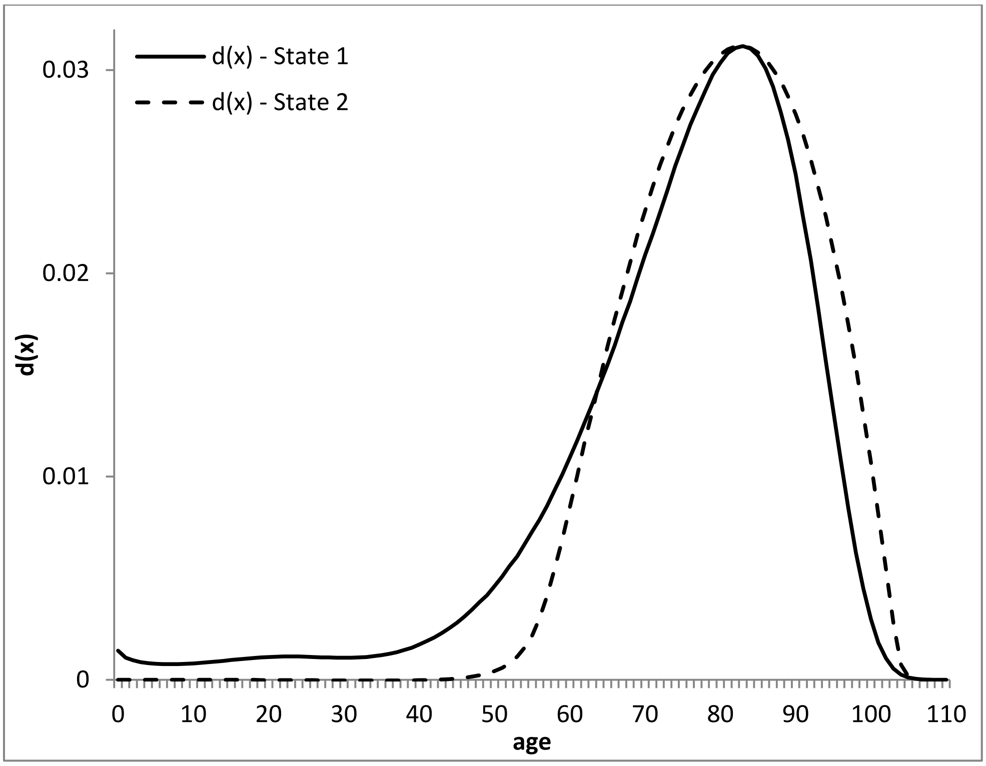

Fig. 6 Mortality evolution with constant M, $d(M)$, and UB, but changing DoI

Of course each of the four components mentioned above can remain unchanged over time. In this case, the respective component is referred to as *neutral*. Thus, every component can attain three states.[6]

Note that two of the four statistics explained above (UB and M) determine the "position" of the deaths curve, while the other two ($d(M)$ and DoI) describe its shape. We believe that these four characteristics provide a good trade-off between granularity and complexity. The four

[5] Only in rather theoretical scenarios, the peak might not be unique, e.g. 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.

[6] Of course, if 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 Section 4) can be added.

components are summarized in Table 1. In principle, every combination of the three different states for each component is possible. This implies that we can classify both pure and mixed scenarios, which was one of the requirements from Section 2. 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 like (*neutral, extension, compression, neutral*) describes a mixed scenario which contains elements of both extension and compression. In total, there are $3^4 = 81$ possible mortality scenarios which might seem unfeasible at first glance. However, many scenarios will hardly be observed in practice, e.g. (*left shifting mortality, extension, compression, diffusion*). Those scenarios are nevertheless part of our classification framework as they complete the set of possible scenarios and thus make the framework applicable in a very general setting. Note that there are no unclassifiable evolutions and that classifications are unique.

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

Table 1 Scenario components, attainable states, and criteria

4. Methodology for the Implementation of the Classification Framework

Application of the classification framework introduced in Section 3 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 in order to determine the prevailing states in each of the four scenario components. This analysis is addressed in Subsection 4.2. Obviously, various different estimators and methods can be used to implement both steps, and thus the specific estimators and methods described in this section are only one possible implementation.

4.1. Estimation of Statistics

For the first component, i.e. the position of a deaths curve's peak measured by M, we use the following estimator by Kannisto (2001):

$$M = x_{d\_max} + \frac{d(x_{d\_max}) - d(x_{d\_max} - 1)}{\left(d(x_{d\_max}) - d(x_{d\_max} - 1)\right) + \left(d(x_{d\_max}) - d(x_{d\_max} + 1)\right)},$$

where x_{d_max} is the age for which the largest number of deaths is observed. As a byproduct, the last scenario component, i.e. the height of a deaths curve's peak, can then be estimated by the number of deaths at age x_{d_max}:

$$d(M) = d(x_{d\_max}).$$

For the second scenario component, i.e. the upper bound UB of a deaths curve, we use the age at the 99th percentile of the distribution of deaths, $x_{99^{th}}$, plus an estimate for the remaining life expectancy at that age. Thus, the estimator for UB is

$$UB = x_{99^{th}} + e_{x_{99^{th}}}.$$

This approach builds on Rossi et al. (2013) who propose using the 90th percentile of the distribution of deaths as an approximation for the highest attainable ages. We prefer our combined estimator as it is considerably less biased. For the populations we have looked at, it provides a reasonable compromise between only cutting off a small part of the distribution of deaths and stability in the statistic's evolution over time.

The degree of inequality DoI in the distribution of deaths is intuitively given by the area between the actual deaths curve and a hypothetical flat deaths curve $d_{flat}(x)$ as illustrated in Fig. 7. Using discrete data, this area can be approximated by adding up the absolute differences in the numbers of deaths between the two deaths curves. Thus, we estimate DoI as

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

Note that DoI assumes its maximum value $l_{x_0} - l_{x_0}/(\lfloor UB \rfloor - x_0 + 1)$ in case all people die at the same age, which would then be M and UB at the same time. The statistic's minimum value is 0 in case deaths are uniformly distributed over all ages, i.e. $d(x) = d_{flat}(x)$ holds for all x.

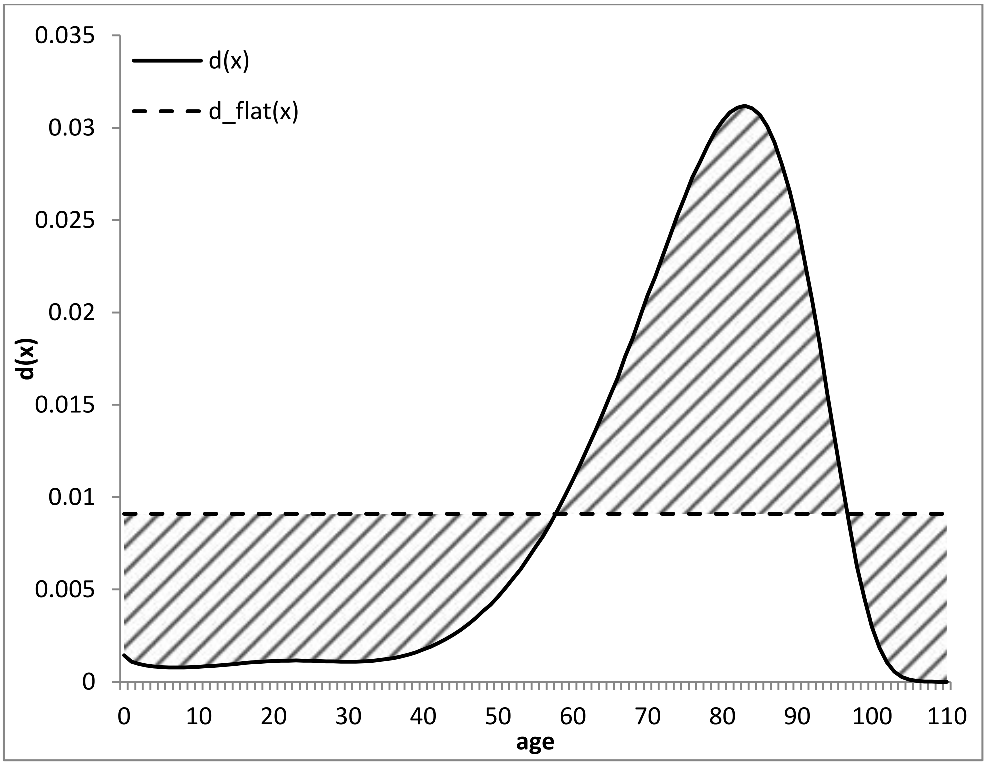

Fig. 7 DoI as the area between observed deaths curve $d(x)$ and hypothetical flat deaths curve $d_{flat}(x)$

Note that the dependence of DoI on UB is uncritical in our framework since we are only interested in changes of DoI over time. A potential misestimation 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 both $d(x)$ and $d_{flat}(x)$ would basically cancel each other.

As mentioned above, alternative estimators could be used for the four statistics. In particular, there is an extensive literature on measuring the upper bound of a deaths curve 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, amongst others, be found in Cheung and Robine (2007), Fries (1980), or 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 the entropy as originally proposed by Demetrius (1974) and adopted by Keyfitz (1985) and Wilmoth and Horiuchi (1999). However, the Gini-Index and the 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 estimating the four statistics for each year in the observation period (an example of the resulting time series is shown in Fig. 8), the trends prevailing at each point in time need to be determined. We will introduce a possible methodology in the remainder of this section.

Eliminating Outliers

First, potential outliers should be eliminated as they are irrelevant with respect to long-term trends, but can significantly blur the trend analysis. Such outliers are typically caused by extreme events like the Spanish Flu. In order 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% confidence interval for the data point under consideration. If it lies outside the confidence interval, it is considered an outlier.

Determining Trends, Trend Changes, and Jumps

In order 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, e.g. the fall of the Soviet Union or wars like WW II, 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). Intuitively, the following methodology first determines which of the three possibilities is the most likely one 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.

In the first step, we identify data points that are "candidates" for a trend change and/or jump. For this, we carry out three fits for every set of three data points:[7] a straight regression line from the first to the third data point, a continuous line with a change in slope at the second data point, and two straight lines (from the first to the second and from the second to the third data point) 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, given that the adjacent changes in slopes, adjacent jumps or the start and/or end point of the time series are located at the first and third data point. In the first Chow test (significance level of 99%), the null hypothesis of one persistent trend, i.e. no change at the second data point, is tested against a continuous

[7] Of course, when performing the fits, we fit to all original data points that lie between the first and the third point of the selected set of three points.

change in slope. The result of the test (the new null hypothesis) is then tested versus a jump in a second Chow test.

In the second step, we identify the number and locations of trend changes that result in an optimal fit. We commence with one straight regression line, i.e. the case of no trend change at all. Then we compare this with the case of one trend change. From the previous analysis we know which data points are candidates for this trend change (given that the first and third data point are the very beginning and the very end of the time series) and whether it would be a change in slope with or without a jump.[8] In order to determine the optimal position (and thus also the type) of the one trend change, we compare the resulting fits for all candidate data points by the Akaike Information Criterion (AIC; Akaike 1973).[9] The number of parameters is four in case of a change in slope (position of the trend change, one intercept, and two slopes) and five in case of a jump (one additional intercept). Moreover, since we do not know the residuals' variances which are required in the likelihood part of the AIC, we use the sample variance of the residuals from the fit without trend change as variance estimator.

Having determined the best fit with exactly one trend change, we use another Chow test (again with significance level 99%) to test it against the case without trend change (the null hypothesis). Thus, the additional trend change is only accepted if it significantly improves the fit, which is in line with our intention of determining long-term trends. If the null hypothesis of no trend change stands, we have found that the time series can be adequately described by a straight line.

Otherwise, we successively increase the number of possible trend changes, say from n to $n+1$. Again, fitted curves for all feasible combinations of locations for $n+1$ trend changes are compared by the AIC, and variance estimators are derived from the previous optimal fit, i.e. with n trend changes. Note that we determine constant variance estimates separately for each period with constant trend. We use a regime switch argument to justify that the variance can change when the trend changes and thereby allow for heteroscedasticity as it can e.g. be observed in Fig. 8.

Having determined the optimal fit with $n+1$ trend changes (if it exists), we perform a Chow test versus the optimal fit with n trend changes.[10] More precisely, since the original test by Chow only considers 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).

The algorithm terminates if either no candidate for an additional trend change exists or an additional trend change does not significantly improve the fit.

Testing for Increasing, Decreasing, or Neutral Statistics

Finally, we have to determine if the resulting trend curve (see the lines in Fig. 8) should be considered increasing, decreasing, or neutral in the context of our framework. For each period with constant trend, we use an F-test to analyze whether the slope is significantly different from

[8] If the previous analysis identifies no candidate at all, a single straight regression line would obviously be the most adequate fit to the time series and the algorithm would stop here.

[9] Obviously, alternative information criteria like the Bayesian Information Criterion (BIC; Schwarz 1978) could also be applied. However, we prefer the AIC since its penalty term does not depend on the length of the time series. Thus, the addition of new data points does not change the penalty for possible trend changes far in the past.

[10] Note that beginning with the Chow test for two versus three trend changes, we can account for heteroscedasticity also in this test. Here, we derive variance estimators from the optimal fit with $n-1$ trend changes.

zero.[11] If it is, for a positive (negative) slope we consider the statistic as *increasing* (*decreasing*) during the corresponding period of time. If the slope is not significantly different from zero, the state *neutral* is assigned.

5. Application of the Classification Framework

In this section, we apply our classification framework to the historical mortality evolution of females in the USA.[12] 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) for years 1947 to 2013. For each calendar year, these log mortality rates are then smoothed and extrapolated using P-splines. This allows us to derive normalized (see Subsection 2.3) and smoothed deaths curves. We prefer this approach over using the raw deaths curves from the HMD for several reasons: Potential disturbing effects resulting from birth cohorts of different sizes are eliminated; random effects in the data which might, e.g., lead to double peaks in the deaths curve are significantly reduced; the potential impact of age misspecifications in the raw data, in particular with respect to estimating UB, is reduced; and the time series for the four statistics exhibit less random fluctuations and are thus easier to analyze.

We consider deaths curves covering different age ranges as discussed in Subsection 2.3: 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 8 displays the four components of our classification framework for both starting ages along with the respective piecewise linear trend lines.[13]

[11] We use a significance level of 90% in the F-test and found this to be a reasonable choice in general. However, the sensitivity of results to changes in this parameter should be checked carefully when applying the framework.

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

[13] We have also considered the starting ages 0, i.e. the complete age range, and 30 in order to exclude effects of young adult's mortality like accidents, etc. It turned out that the observed scenarios for starting ages 0, 10, and 30 are quite similar. Thus, we decided to only display the results for starting age 10.

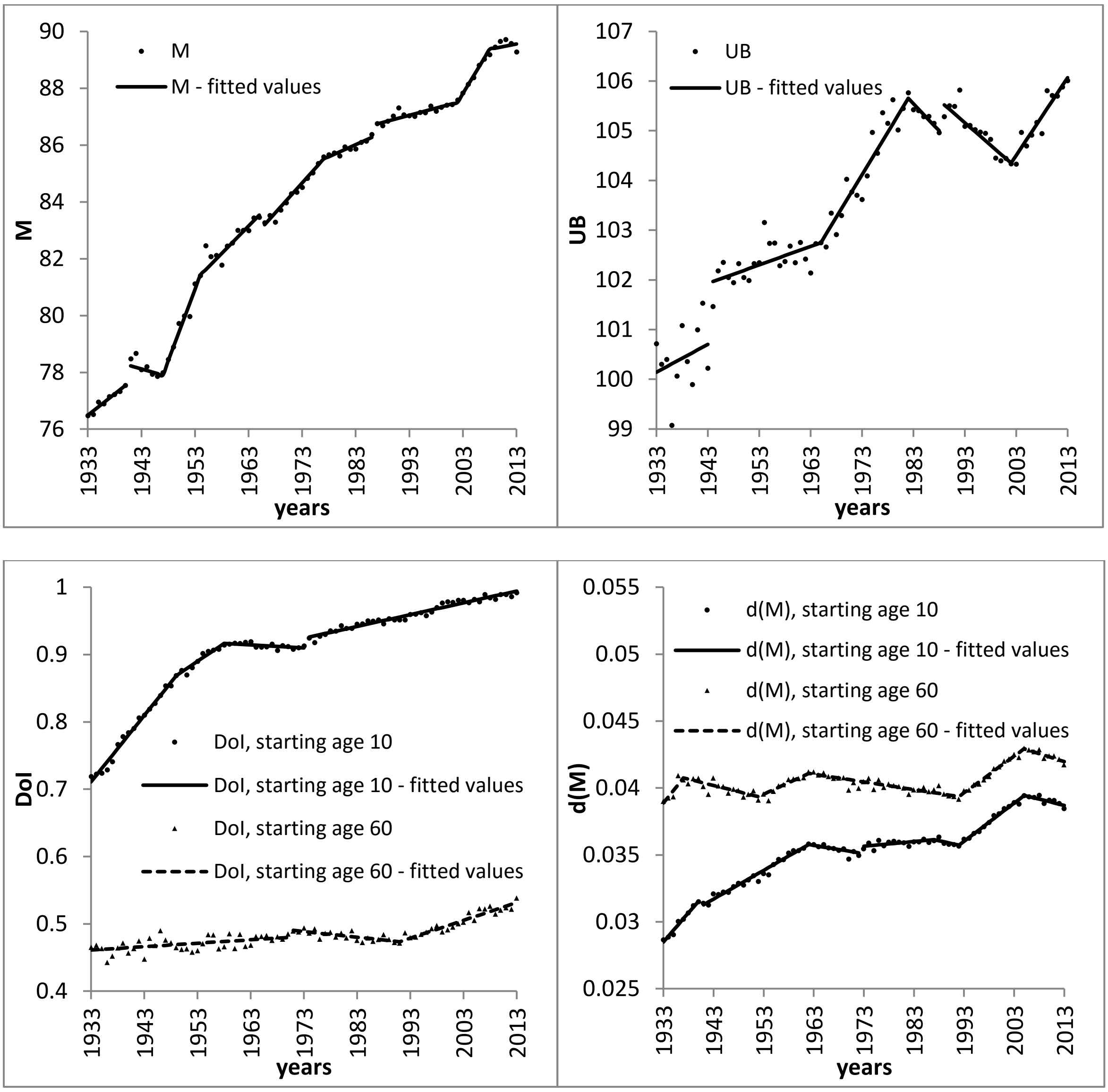

Fig. 8 Development of the four components of our classification framework for US females from 1947 to 2013. Upper left panel: M; upper right panel: UB; lower left panel: DoI; lower right panel: $d(M)$

Due to the definition of the modal age at death M, the curves for this statistic 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. Since 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 only display UB for starting age 10.

From Fig. 8 we can see that our framework identifies several trend changes for each of the statistics and both starting ages. Such trend changes can mean that either the trend changes its direction (e.g. from increasing to neutral or decreasing, etc.), or only the intensity of the trend (i.e. the slope of the trend line) changes significantly while the direction of the trend 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) only indicate changes in 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 in Section 4, trends can change with or without a jump in the absolute level of the statistic. In our example, such jumps occur for almost every statistic ($d(M)_{60}$ being the only exception).

The direction of each trend as well as the position of the trend changes and their kinds (i.e. a change in slope with or without an upward/downward jump) is summarized in Fig. 9. 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, +, +, -), i.e. the scenario is neutral with respect to shifting mortality and exhibits extension, compression, and diffusion at the same time.

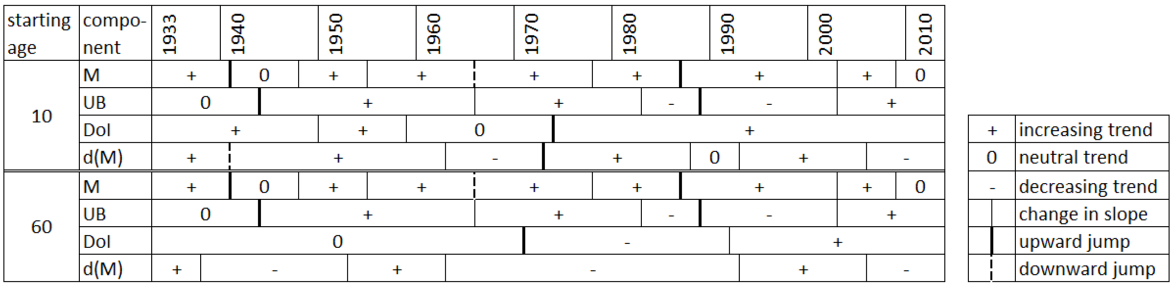

Fig. 9 Time bars of mortality evolution for US females, each statistic, and both starting ages

By comparing Fig. 8 to Fig. 9 we find some periods with apparently increasing (decreasing) trends in Fig. 8, but a classification as "neutral" in Fig. 9. One such example is the first trend for UB. Here the underlying data has relatively strong variance, and therefore the seemingly increasing trend is not significant.

The results of our analysis particularly 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, i.e. we observe right shifting mortality and extension. However, particularly for UB there are some periods (see for example the 1990s) where we observe the opposite trend, i.e. contraction, 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, i.e. 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 time periods where the trends differ by starting age. For instance, between 1974 and 1991 we observe compression for starting age 10, but decompression for starting age 60.

Furthermore, we find that pure scenarios where only one component is non-neutral are the exception rather than the rule and cannot be observed at all in the example. In contrast, there are even periods where all four indicators change, e.g. between 1974 and 1982 for the 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.

6. Conclusion

In this article we explain why many existing approaches to classify patterns of mortality evolution have three major shortcomings: Mortality scenario definitions are often imprecise and intuitive rather than rigorous; some frequently used statistics are not sufficient to identify scenarios, in particular mixed scenarios; often, the impact of the considered age range is being 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; (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; 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 statistic and determine trends and trend changes in the data. Also, we apply our approach to data for US 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.

Note that the purpose of our framework is a classification of realized mortality evolutions. In this sense it is purely descriptive, i.e. 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 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 purposes of mortality extrapolation. For instance, existing mortality projections can be tested for consistency with observed trends in the most recent history. Also, best estimates for the future mortality evolution can be derived by extrapolating recently observed trends. Such applications could be beneficial for governments and life insurers/pension funds, alike, e.g. for projecting obligations of social security systems and annuity/pension portfolios, respectively.

Literature

Akaike, H. (1973). Information Theory and an Extension of the Maximum Likelihood Principle. *2nd International Symposium of Information Theory*, B.N. Petrov and F. Csaki (eds.), Akademiai Kiado, Budapest, 267-281.

Bongaarts, J. (2005). Long-Range Trends in Adult Mortality: Models and Projection Methods. *Demography , 42* (1), 23-49.

Canudas-Romo, V. (2008). The Modal Age at Death and the Shifting Mortality Hypothesis. *Demographic Research , 19* (30), 1179-1204.

Cheung, S. L., Robine, J.-M., Tu, E. J.-C., & Caselli, G. (2005). Three Dimensions of the Survival Curve: Horizontalization, Verticalization, and Longevity Extension. *Demography , 42* (2), 243-258.

Cheung, S. L., & Robine, J.-M. (2007). Increase in Common Longevity and the Compression of Mortality: The Case of Japan. *Population Studies , 61* (1), 85-97.

Chow, G.C. (1960). Tests of Equality Between Sets of Coefficients in Two Linear Regressions. *Econometrica, 28(*3), 591-605.

Debón, A., Martínez-Ruiz, F., & Montes, F. (2011). *Temporal evolution of some mortality indicators. Application to Spanish data.* Society of Actuaries. Life Monographs. https://soa.org/library/monographs/life/living-to-100/2011/mono-li11-1b-debon.pdf. Retrieved January, 16 2015.

Demetrius, L. (1974). Demographic Parameters and Natural Selection. *Proceedings of the National Academy of Sciences of the United States of America,* 12, 4645-4647

Finch, C. E., & Pike, M. C. (1996). Maximum Life Span Predictions From the Gompertz Mortality Model. *Journal of Gerontology: Biological Sciences , 51A* (3), B183-B194.

Fries, J. F. (1980). Aging, Natural Death, and the Compression of Morbidity. *The New England Journal of Medicine , 303* (3), 130-135.

Gampe, J. (2010). Human Mortality Beyond Age 110. In H. Maier, J. Gampe, B. Jeune, J.-M. Robine, & J. W. Vaupel, *Demographic Research Monographs – Supercentenarians,* Heidelberg: Springer Verlag, 219-230.

Genz, M. (2017). *A Comprehensive Analysis of the Patterns of Worldwide Mortality Evolution.* Society of Actuaries. Life Monographs. *In press.*

Human Mortality Database. (n.d.). *Univerity of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany).* www.mortality.org. Retrieved August 16, 2014.

Kannisto, V. (2000). Measuring the Compression of Mortality. *Demographic Research , 3* (6).

Kannisto, V. (2001). Mode and Dispersion of the Length of Life. *Population: An English Selection, 13* (1), 159-172.

Keyfitz, N. (1985). *Applied Mathematical Demography,* New York: Springer Verlag.

Manton, K. G., & Tolley, H. D. (1991). Rectangularization of the Survival Curve - Implications of an Ill-Posed Question. *Journal of Ageing and Health , 3* (2), 172-193.

Myers, G. C., & Manton, K. G. (1984). Compression of Mortality: Myth or Reality? *The Gerontologist , 24* (4), 346-353.

Nusselder, W. J., & Mackenbach, J. P. (1996). Rectangularization of the Survival Curve in the Netherlands, 1950-1992. *The Gerontologist , 36* (6), 773-782.

Ouellette, N., & Bourbeau, R. (2010). *Recent adult mortality trends in Canada, the United States and other low mortality countries.* Society of Actuaries. Life Monographs. https://soa.org/library/monographs/life/living-to-100/2011/mono-li11-4b-ouellette.pdf. Retrieved January 16, 2015.

Rossi, I. A., Rousson, V., & Paccaud, F. (2013). The Contribution of Rectangularization to the Secular Increase of Life Expectancy: An Empirical Study. *International Journal of Epidemiology , 42*, 250-258.

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics, 6* (2), 461-464.

Thatcher, A. R., Cheung, S. L., Horiuchi, S., & Robine, J.-M. (2010). The Compression of Deaths Above the Mode. *Demographic Research , 22* (17), 505-538.

Wilmoth, J. R. (1997). In Search of Limits. In K. W. Wachter, & C. E. Finch, *Between Zeus and the Salmon - The Biodemography of Longevity* (pp. 38-64). Washington D.C.: The National Academies Press.

Wilmoth, J. R., & Horiuchi, S. (1999). Rectangularization Revisited: Variability of Age at Death within Human Populations. *Demography , 36* (4), 475-495.

Wilmoth, J. R. (2000). Demography of Longevity: Past, Present, and Future Trends. *Experimental Gerontology , 35*, 1111-1129.