Mortality and Longevity

Aging and Retirement

A Method for Mortality Rate Projection: A Five-Step Approach

NOVEMBER | 2023

LIVING TO 100

SOA INTERNATIONAL SYMPOSIUM

SOA Research INSTITUTE

A Method for Mortality Rate Projection: A Five-Step Approach

AUTHOR Sam Gutterman, FSA, FCAS, MAAA, CERA, FCA, HonFIA
Consulting Actuary

SPONSORS Mortality and Longevity Strategic Research Program Steering Committee

Aging and Retirement Strategic Research Program Steering Committee

CONTENTS

A Method for Mortality Rate Projection: A Five-Step Approach

Executive Summary

This paper describes a general methodology that can be applied to estimate and project mortality rates and mortality improvement rates, including various considerations that may be useful in implementing this methodology in different situations. The method consists of a systematic decomposition of the process into periods of description (estimates of past and current rates) and projection (forecasts of future rates).

The five-step process described can provide a rigorous and constructive approach to the mortality improvement projection process and decision support for an application for which future mortality is an important assumption. The periods analyzed correspond to a structure for conducting a typical analysis:

1. A historical period during which reliable and reasonably relevant experience data are available.
2. A historical period during which limited or no data are available, that is, immediately prior to the starting date of the projection.
3. The period over which penultimate mortality improvement rates apply.
4. The transition period between steps (2) and (3).
5. The period over which final (ultimate) improvement rates apply.

For each of these periods of estimation and projection, factors that an actuary can consider in developing estimates and projecting mortality rates are discussed. The considerations and sources of information underlying the methodology used will depend on the particular application. The importance of understanding the context and objectives application is emphasized, partly because of mortality heterogeneity among and within population segments, such as those reflecting the range of socioeconomic status in the applicable population. Thus, results generated from the aggregate national population may not be appropriate for a specific population segment and application.

Several alternative approaches to specific situations are assessed, for instance, during a temporary (e.g., COVID-19) or permanent (e.g., a change in smoking prevalence) mortality disruption and a period whose pattern over time looks more like a wave (e.g., AIDS). Although the effects of certain mortality drivers will be reviewed (e.g., obesity and climate change), the effects of specific causes will not be dealt with in an in-depth manner. Cause-of-death modeling and possible treatment of mortality disruptions will be assessed.

Several approaches to considering uncertainty are also described, in particular, through stochastic modeling and scenario analysis. In addition, the interrelationships with other assumptions will be briefly covered, for example, voluntary lapsation (a driver of select-and-ultimate mortality tables) in individual life insurance, and mortality and disability for long-term care.

The application of this methodology can be made in any field of actuarial practice that involves a mortality assumption. The basis for its development and description is the actuarial experience gained by the author over several decades of involvement in developing and updating mortality assumptions for a wide range of practice applications.

1. Introduction

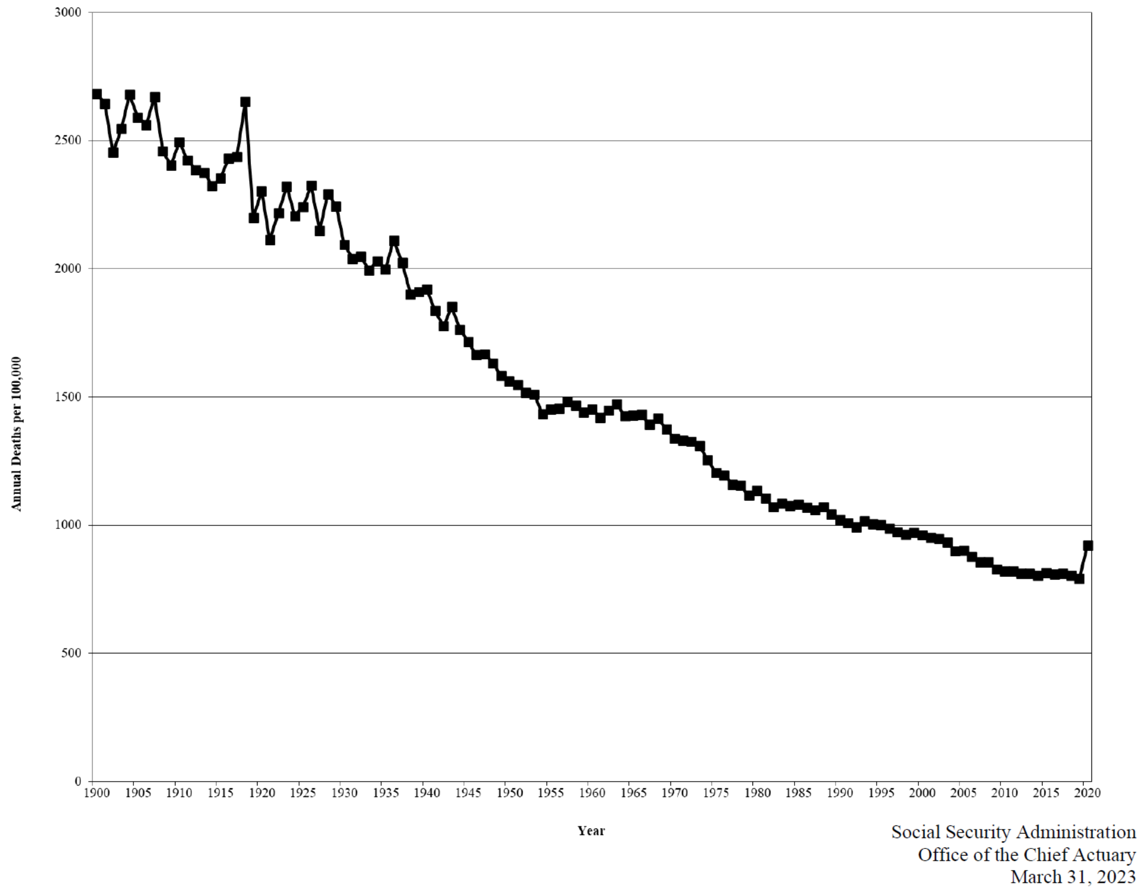
Overall reductions in mortality rates and corresponding increases in mortality improvement rates and life expectancies have been one of the great achievements of the last two centuries (see Figure 1 for the U.S.). Corresponding improvements have occurred in almost all countries, albeit at an uneven pace in some areas and for some groups. Reasons for these improvements have included improved public sanitation, public health actions, infectious disease management, and medical understanding, diagnostics, prevention and treatment of noncommunicable diseases. An important transition in premature deaths for many countries, except for certain epidemics and pandemics, has been (and will continue to be in some less-developed countries) a significant change from the dominance of deaths from infectious diseases to those from noncommunicable diseases.

Developing sound mortality estimates and forecasts is important and has many applications. For example, outputs from mortality models have become an integral input to policymaking (in some European countries, they are directly used to set statutory retirement ages). For many actuarial practice areas, fundamental components in fulfilling such functions are pricing and reserving for life insurance and annuity coverages and setting contribution and soundness targets for pension programs.

A key issue confronting society, as well as for actuaries, is the extent that these favorable trends will continue, what headwinds might emerge, and what are the most affordable approaches to manage the risks involved. They are also key inputs to health-care-related policy decision making. Although Figure 1 might lead one to believe that the regular and seemingly continued increase in longevity (decrease in mortality) should be straightforward to forecast, albeit possibly at a somewhat reduced rate, many factors are involved, some of which are not easy to observe. Note the adverse shock shown for 2020 is because of the COVID-19 pandemic. Opinions with respect to the speed and duration of future improvements have and will likely continue to differ considerably–that is, will future improvement look more like that in 2000 to 2009, 2010 to 2019 or another period? The one thing that is certain is that the future will not look like the past.

Figure 1

HISTORICAL U.S. AGE-GENDER ADJUSTED CENTRAL MORTALITY RATES: 1900–2020



Mortality tables have been developed for more than three centuries, with actuaries involved ever since there were actuaries. Although the original objective was to provide annuities or pensions on a sound basis, a great deal of effort has been the development of sets of historical mortality rates for nationwide or segments of populations. There are now many more uses and users of this information—longevity risk is an issue of social, economic and financial importance.

The primary objective of this paper is to describe a systematic approach to the analysis of mortality and its projection. The first forecast published in the U.S. regulatory area incorporating mortality improvement was in 1949, the Group Annuity Table of 1949 (Jenkins and Lew 1949). Projections incorporating mortality improvement for life insurance were not published until the 2010s because of concern for conservative regulatory reporting.

Too often, actuarial literature regarding mortality over the last several decades has gotten bogged down in the question of which formula or what refinement should be used to represent the all-cause mortality rate curve, with too little in the literature addressing the practical process of developing a forward-looking mortality assumption. Both the Continuous Mortality Investigation (CMI in the United Kingdom) and the Society of Actuaries' Retirement Plan Experience Committee (RPEC) have developed approaches to project mortality for actuarial practice areas over the last 20 years.

The basic objective in constructing a set of useable mortality rates is to convert the number of raw deaths and corresponding population data into a "smooth"[1] table of annual mortality rates and improvement factors that differ by, at least, age and gender,[2] extrapolated[3] in some form to various years in the future. Particularly when a "life table" is being constructed, it is important to have smooth rates.

In this paper, "estimation" refers to the quantification of an estimate of values in a historical period, whether observed or not, whereas "projection" refers to the quantification of values in a future period—what has been in contrast with what will be. Both sets of values are, by necessity, estimates of the underlying values.

Conditions and risk characteristics of the population included in the relied upon historical period are not necessarily consistent with those to which projected mortality is to be applied, that is, it is not an experiment involving an urn with replaced marbles as done in an introductory probability course. As a result, it is expected that historical experience cannot be applied without the application of expert judgment or adjustment in the estimation or projection process. Although this estimation is a process rather than just the selection of a set of numbers, it contains elements of a science and an art.

The mortality of a group of people is the end result of a multitude of life processes and external forces. It can be a byproduct of aging, with the body's resilience to disease decreasing over time, resulting in mortality rates that increase with age (there is a wide range of theories of aging, but space in this paper does not permit their discussion[4] and they are not directly relevant to the objectives of this paper). However, certain diseases and causes of death, including conditions that primarily strike children and external causes such as accidents, homicides, war and drug overdose, do not follow the "normal" aging pattern.

If mortality simply followed the effects of normal aging, a mathematical representation, such as a statistical log-linear curve such as Gompertz's (1825) or Makeham's (1860), could be used in a straightforward manner, especially over a specified age range, or if future average mortality could be expected to continue to improve in the same manner it has since 1950. However, health and mortality are by their nature messy, often defeating the simple laws of nature with no unique formulaic curve properly fitting all ages. Even age is not straightforward, as the risk characteristics of an individual or group of individuals are not necessarily directly related to calendar age; rather, they depend on numerous genetic and environmental factors. Many disruptions (e.g., pandemics, wars, changes in behavior and the nature of work, and medical developments) to extrapolated mortality can be subject to heavy tail risks that could dramatically transform its results. All of these contribute to the complexity of this subject.

Mortality projections fundamentally consist of two basic elements: (1) a base mortality level, often based on one or more sets of annual mortality rates, involving certain smoothed estimates, and (2) a set of usually multiplicative changes (improvement factors). The factors from (2) are applied to the base rates

[1] To avoid volatility, unexplained discontinuities that may not be warranted and for which no fundamental reasons are known.
[2] Early on, because of observations regarding different patterns of reported mortality experience, it was realized that age was a central driver of mortality, with gender being next to be recognized as being important.
[3] Extrapolation is a process in which a number is estimated outside the range of a data set. Interpolation, by contrast, is a process in which a number is estimated between two data values.
[4] These include mutation accumulation theory, in which aging is seen as an inevitable result of the declining force of natural selection with age, with deleterious mutations that tend to accumulate, leading to an increase in mortality rates late in life; antagonistic pleiotropy theory in which late-acting harmful genes may be favored by selection and be actively accumulated in populations if they have any beneficial effects early in life (Gavrilov and Gavrilova 2002); evolutionary senescence theory of aging using dynamic models, relational models and other mathematical representations, evolution/population genetic theory, reliability models and characteristics of the biological aging process; frailty development, which is based on individual risk-factor development; and decomposition approaches.

from (1) to derive mortality rates expected at future periods, applied on a cumulative basis. An assessment of the sources of and understanding of the uncertainties associated with the projection usually accompanies the resulting projections.

Many methods of projection have been developed and used over the years (for further discussion, see Booth and Tickle 2008). They group these methods into three categories:

- Expectations, including expert opinion (sometimes referred to as "professional judgment")
- Extrapolation, based on statistical or actuarial modeling, applying expected patterns by age and other variables, and past trends and
- Explanation, based on an analysis of causes or risk drivers[5] of death.

Each of these approaches has its advantages and limitations. Over-reliance on a single approach has the potential of producing biased, if not inaccurate, projections. I believe that the application of key elements of each in combination may constitute the best methodology in many applications, as will be discussed later in this paper, the relative importance of which may be based on the application. In any case it is important to search for the reasons for and to enhance the understanding of the observed mortality patterns (Gutterman and Vanderhoof 1999). Nevertheless, they all begin by developing a historical set of mortality rates as a baseline.

Although the underlying contributors to change in mortality often act in a linear manner, this is not always the case. Often the risk factors or mortality outcomes behave in a nonlinear or discontinuous fashion. This requires choices to be made regarding how to manage their development, either in understanding historical data or in developing estimates or projections of mortality.

This paper is organized in the following manner. After initial discussions of groups, cohort and period effects, and possible applications of the analysis of mortality, the paper introduces the topics of mortality data and metrics. It is followed by a discussion of a methodology using five historical estimation periods and future forecast periods. It closes with the important topics of mortality modeling, uncertainty, validation and governance.

2. Groupings of Individuals

Since the mortality experience of an individual does not provide an appropriate base for the analysis of mortality patterns, all historical or forecast analysis includes some form of aggregation into groups, for example, a country's total population or participants in a pension plan.

Although using one level of aggregation or categorization in a particular situation or application may be appropriate, it may not be in another. Groupings for this purpose can be categorized by time and/or space. For example, if the estimate or forecast is to apply to the most recent quarter or subsequent years, alternative approaches may be appropriate, as indicated later in this paper. If conducted in the middle of a

[5] A mortality risk driver (sometimes referred to in this paper as a "driver") generally refers to the primary or important secondary contributor to a death. A driver is distinguished from a cause of death in that the latter usually consists of the primary medical reason, internal or external condition (sickness or injury), that directly results in a death. An example is that the excess quantity and poor quality of nutrition can be drivers of death, while cardiovascular diseases may be the direct medical cause, the latter of which would be recorded on death certificates and in many cases is available to the analyst, depending on the database used. In some cases attribution issues arise, especially when more than one driver or cause of death is involved in the trajectory of death.

pandemic, then weekly, monthly or quarterly forecasts may be more appropriate, even though actuaries normally deal with mortality expressed in annual intervals.

Constraints on groupings used to analyze historical experience or the extent of granularity of projections often arise because of factors such as the lack of availability of reliable data or privacy concerns, while the purpose and use of the projection can establish the population to which the mortality rates will be applied.

The population segment studied could be a subset of an overall population or exclusionary (e.g., other than those who are disabled or unhealthy, such as in the development of life expectancy for the healthy, referred to as "healthy life expectancy"). An approach similar to that described in this paper can be used to assess healthy life expectancy, although the data can be more difficult to obtain and its scope can be more subjective. An unhealthy life expectancy, in contrast, is the period a group of unhealthy or disabled individuals can be expected to live. However, the unhealthy population could include those newly unhealthy every period, which results in a hybrid set of future mortality rates. In addition, the population segment studied may depend on its relevance (that is, sufficiently close to the expected population that under similar conditions would be expected to experience similar mortality) to the population to which the mortality projection will be applied.

Significant differences in mortality rates for consecutive individual ages in two adjoining quinquennial age groups (e.g., between ages 74 and 75 in age groups 70–74 and 75–79) can accumulate when mortality improvement rates for the two age groups differ. Even small differences in these improvement rates between edges of consecutive age groups might result in discontinuities in future rates, especially in the steep part of the mortality curve. If this occurs, a more granular age assumption (e.g., annual in contrast to quinquennial) or smoothing of the future rates could be performed.

Every group consists of individuals with heterogeneous risk characteristics and different risk profiles. If differential experience is identified or obtainable by major subgroup, the experience of the subgroups should be assessed to determine whether they can provide insight into trends or other patterns.

The change in the composition of a population segment over time may require a dynamic evaluation of experience. To ascertain whether mix risk[6] has or is expected to occur, the most recent information regarding the heterogeneity, concentration and change in the risk profile of each group in each step needs to be assessed. If significant changes have or are expected to arise between periods, a determination should be made as to whether an adjustment is needed. Such a change could result from a substantial number of population entrants or exits (such as births or migrants in examining a certain population segment, new entrants or voluntary terminations in the case of an insurance group, or new workers, terminators prior to vesting and retirees for pension plans).

If a mortality disruption has occurred (see section 8.1 for further discussion) during the historical period, especially if expected to be permanent (e.g., a structural transformation of some kind) or has a large marginal effect, an adjustment may be needed to provide a sound base for estimating or projecting future experience. In any case the application period can influence the choice of population segment used as a base and can require adjustment in the experience data used or the projections made.

[6] The risk that the mix of subpopulations changes so as to result in a significant difference in its overall expected level of mortality.

Population segments can experience different levels or patterns of mortality. Arguably, they can also experience different levels of mortality improvement. The following are several age-related categories that have had and will likely continue to experience unique mortality patterns:

- *Youngest ages*. Although mortality for children has tended to improve at a faster rate than that at adult ages, there is likely to be a minimum below which mortality cannot breach because many accidental deaths are impossible to eliminate. As a result, in some countries, projecting significant improvement into the future at the rate experienced over the last several decades may eventually overstate the rates that mortality can achieve at those ages.
- *Working adult ages*. Mortality trends and disruptions will inevitably affect the aggregate mortality results. An example is in the 2010s when a wide range of adult ages in the United States experienced mortality deterioration (negative mortality improvement rates). Included in these adverse[7] mortality trends were the so-called "deaths of despair," particularly with an increase in deaths due to excessive drug overdoses, firearms and alcohol consumption.
- *Older ages*.
    - I. *Age gradient.* A common actuarial assumption is that the annual rate of mortality improvement at older ages will be lower than for those at younger adult ages, expressed on a percentage basis. Although this has been true overall in the U.S. over a long period, this relationship has not held in each period.
    - II. *Multimorbidities and data.* As people age, they tend to accumulate multiple sets of morbidities, referred to as multimorbidities. After, say, age 65, it is unusual to find a person who has none or only one significant ailment or disease. And with improved treatment and reduced mortality, more people will have multiple adverse physical or mental conditions. Although not all of them are life-threatening, their accumulation can make it difficult to attribute deaths to a single primary cause, sometimes making cause of death analysis problematic at older ages. Over time at a specific attained age, especially in higher-income countries or population segments, the number of outstanding morbidities per person has increased, restraining mortality improvement. In addition, this also may limit the possible favorable effect on mortality of a medical breakthrough that favorably affects a specific cause or population segment. Also, as the number of pharmaceuticals taken for these conditions increases, the risk of harmful interactions may increase.
    - III. *At very old ages (the oldest old).* A debate has occurred with respect to the trajectory of mortality rates at very old ages, that is, at ages 100 and older. Some have expressed the view that these rates tend to increase by age, some that they plateau, while others even say they may decrease after a certain age. In some cases, because of the limited number of people at these ages in most populations, a formula is used—possibly starting at age 95, 100 or 105, depending on the data sources used. This also avoids the common problem of inaccurate, incomplete or inadequate birth date data at those ages.
- *Population age and risk profile*. As populations age, the mix of people at a specific age will also change, possibly reflecting new entrants, exits or changes in risk characteristics in those who remain. In any case, because of the aging process, the experience of a group at a specific age cannot be the same as the population at that same age in the future. As a result, in many cases

[7] For pension and annuity products, experiencing adverse mortality means a decrease in mortality, whereas for life insurance and public health, being adverse means an increase in mortality.

confirmation of the risk characteristic profile of a population segment at a similar attained age group may be needed.

Any group or population segment consists of individuals with heterogeneous risk characteristics or profiles (mortality risk inequalities). Over time, those who are more vulnerable (or unlucky) to die prematurely at an earlier age will leave the group, often resulting in a healthier group of survivors, all else being equal. This may lead to a convergence of disparate risks (that is, earlier differences by such characteristics as educational attainment or socioeconomic status) as a group ages. This is one reason mortality differences between these groups diminish at older ages.

3. Cohort and Period Effects

Attained age in a period is often the key driver of mortality. A common framework related to attained age is the use of cohorts of individuals of a certain age with a common entry point, often based on year (or grouping of years) of birth. In some contexts, the use of birth cohorts can be utilized in an age-period-cohort (APC) approach if it is felt that mortality drivers have a differential effect by birth year.

Whether period (the drivers of rates or change are effective in a calendar period) or cohort mortality rates are used depends upon historical or projected experience patterns, the nature of key mortality drivers and the available data. In some cases the effective time span of a cohort is clear—it may constitute a year or a period demarcated by a single event (e.g., the beginning of the AIDS epidemic) or trend (e.g., a period of nutritional deprivation). The period of the cohort (e.g., starting and ending year or date) should be validated. If determined to be a range of years, whether its effects are uniform over the period or more intense in a particular part of the cohort, cohort experience over the period may differ between younger or older cohorts. Note that differences in results may arise, including the possibility that the effects of a cohort may get larger or diminish over time.

Other cohorts might be measured from initial risk exposure, such as a life or health insurance policy issue date, the time of joining the population analyzed such as immigrants, or the date of disability or adverse health condition onset.

Each external factor that can influence mortality tends to have more of a cohort or period effect, although some are of a hybrid nature, reflecting both types of effects. Others can have a hybrid effect of a temporary or permanent nature; that is, although the external influence or driver can affect a certain population segment in a given period, their experience will be subject to a different level of mortality than others who did not share that experience. Some may be easy to observe in retrospect but difficult to predict. The ebbs and flows of period and cohort effects of drivers and causes of death and their interrelationships can make it difficult to isolate the most important drivers—some form of a multivariate or predictive analytic method might be needed to tease out the dominant or primary effect(s).

These possibilities are more fully, yet certainly not comprehensively, described below:

- *Period*. Many mortality risk drivers operate on a calendar year period basis. The most obvious examples are a pandemic, war or medical breakthrough, although in some cases each may have a delayed effect that is spread over succeeding calendar years. In Figure 2 for males (the corresponding heat map for females is fairly similar, except the bottom middle blue is not evident and the bottom right blue is not as intense, indicating the smaller impact of the AIDS epidemic and drug overdoses on females), there appears to have been a period in the 1950s and 1960s (varying by age) where there was a plateau in mortality, with no obvious single external factor involved.

Then there was a favorable period immediately after 1965 primarily at ages older than 65 that resulted from the effects of the introduction of Medicare and Medicaid in 1965, in addition to the reversal of some of the unfavorable mortality outcomes of the prior period. The favorable period beginning in the 1970s was the introduction of cardiovascular risk mitigation factors (against high blood pressure and cholesterol).

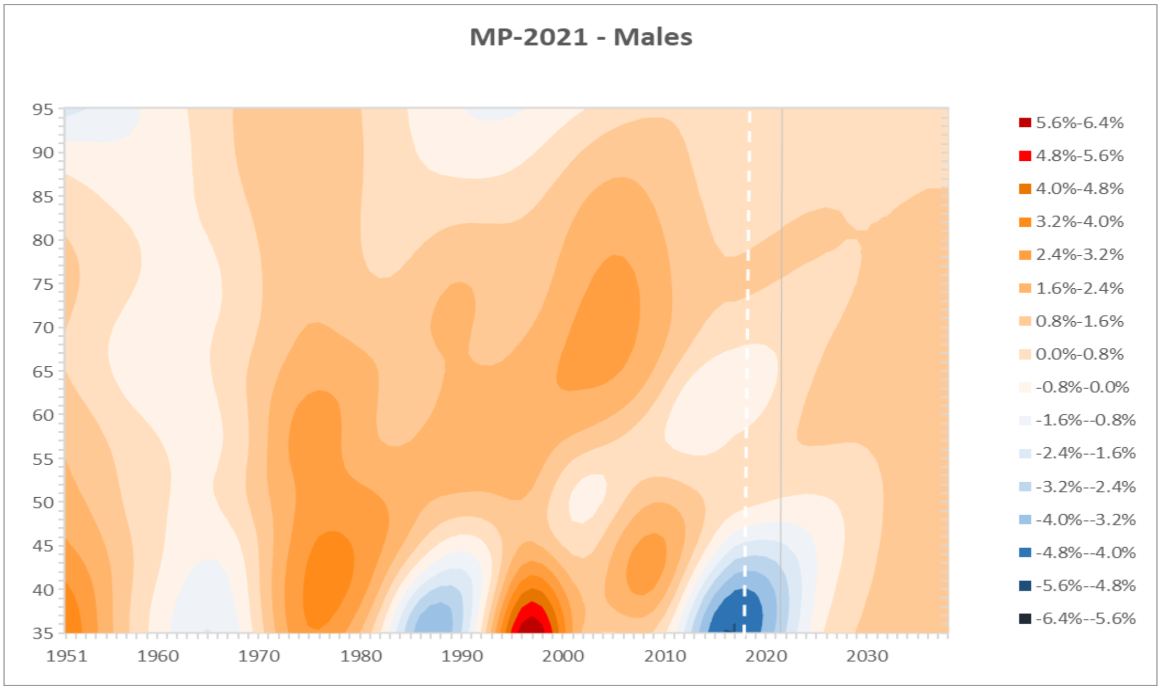
- *Cohort*. As shown in Figure 2, those adults attaining ages 35 through 40 in the mid-1970s (the long diagonal orange area) improved their mortality experience in a favorable cohort. This could be attributed to several causes, including the favorable impact resulting from the reduction in the very high prevalence of smoking for males that began in the 1940s and 1950s, and the introduction of cardiovascular risk mitigation factors. It is important to note that just because a cohort has more or less mortality improvement in the past than neighboring cohorts does not guarantee a continuation of such a relationship in the future. The introduction of a set of more restrictive underwriting rules will affect the average policyholder for a life insurance company. In addition, a limiting case occurs at the very oldest ages, as the effects of many socioeconomic and sociodemographic differences tend to decrease in the run-up to advanced ages, possibly in their 80s and 90s.
- *Hybrid*. A combination of period and cohort effects can result from a driver that operates over a specific period but can also affect an exposed group for a considerable amount of time differently from those born (or entering into the applicable population) before or after. The exposed group might be concentrated in a given population segment, such as in middle adult ages. An example is the HIV/AIDS epidemic: its adverse effects on mortality can be seen in Figure 2 in the blue period at ages 35–40 beginning in the late 1980s. Also, important to note is that where there is such a temporary shock, a reversal or rebound (the red area) in the 1990s can bring the mortality in that age group close to the prior trend.

The Society of Actuaries Research Institute's MIM-2021v2 (2021) offers a facility to incorporate either or both period and cohort trends. Note that this approach requires additional data analysis and may require care, because cohort trends may be difficult to determine over a long period.

Two examples are the change in an educational requirement and the introduction of a new public health care program. In both cases a hybrid approach might be useful. Although the situation that created a change was introduced in a particular period, it can affect the group affected for a long time. Also, some drivers are set, such as educational attainment in the mid-20s, if not earlier, whereas there are others that are dynamic, such as certain behaviors. For a new health care program, although the event becomes effective at a certain date, it may also have a longer-term effect on a cohort or period combined basis. The shape of a common period mortality curve is often affected by both varying age and cohort effects (Yang 2008).

To determine how and the extent to which they are important, the reasons and drivers for both period- and cohort-related effects need to be recognized. This not only includes the effects of past, current and possible future drivers, but when temporary, there will likely be a reversal or rebound of their effects that may be needed to revert to the "normal" level for the cohort. An example is cigarette smoking, which had a huge effect, particularly in the U.S., both adversely following its increased prevalence and favorably as its prevalence decreased. The blue periods in Figure 2 were due to AIDS and drug overdoses in the 1980s and 2010 and 2020s, respectively. It may be useful to aggregate the effect of both the risk driver and its corresponding reversal to determine the future mortality rates, if applicable.

Figure 2
HEAT MAP FOR U.S. MALES



Source: Society of Actuaries Research Institute (2021). Historical mortality improvement rates were calculated from smoothed Social Security Administration historical and projected population mortality data for the years 1951 through 2018. Attained ages are the vertical axis, and calendar years are the horizontal axis.

Another way of showing this type of data is through traditional data tables, such as Table 1, which shows the trend in rates of mortality improvement (negative values represent mortality deterioration) by age group and gender in the U.S.; people differ as to what form of presentation can best enhance their understanding. From this table, a great deal of variation is seen in mortality improvement by period and age group. Over the long period, overall annual improvement rates have averaged about 1.0% a year, with an age gradient by age, although, at least looking at historical trends by the periods presented, there appears to have been a wave in these rates over the 75 years illustrated.

Table 1
RATES OF MORTALITY IMPROVEMENT FOR THE UNITED STATES BY AGE GROUP AND GENDER

	Age/Period	1954–1968	1968–1982	1982–1999	1999–2009	2009–2019	1900–2019
Females	0–14	1.70%	4.26%	2.76%	1.58%	1.49%	2.94%
	15–49	0.40	2.13	1.09	0.84	−0.74	1.23
	50–64	−0.22	2.21	1.87	1.14	0.03	0.84
	65–84	−0.32	1.49	1.04	2.44	0.92	0.79
	85 +	−1.07	1.81	−0.53	1.50	0.38	0.49
	Total	−0.41	1.83	0.81	1.80	0.45	0.90
Males	0–14	1.78%	4.06%	2.56%	1.49	1.42%	3.00%
	15–49	0.30	2.74	0.72	0.14	−0.21	1.72
	50–64	0.68	1.65	1.01	1.33	0.02	1.17
	65–84	0.87	2.00	0.23	1.70	0.93	1.02
	85+	0.04	2.29	−0.52	1.21	0.16	0.63
	Total	0.59	2.14	0.20	1.37	0.45	1.13

Source: Social Security Administration, Office of the Chief Actuary, June 2, 2022, “The Long-Range Demographic Assumptions for the 2022 Trustees Report,” Mortality, p. 18.

4. The Application or Use

In developing a strategy to analyze mortality, it is important to identify or confirm its objectives and the application(s) for which the final results and recommendations will be used. The importance of this cannot be overemphasized. The methodology used usually consists of a combination of data analysis, risk analysis, expert opinion and a form of extrapolation or interpolation, ranging from the simple to a highly refined actuarial model. Approaches taken may differ by the scope, scenario, needed extent of refinement and relative importance of time and risk dimensions.

Its objectives can either be descriptive or predictive, with the former focusing on understanding the past. The understanding gained in the analysis of the past can also be used to project rates into the future.

In constructing a table of mortality rates, it is important to weigh the relative importance of stability and responsiveness to recent experience patterns. The importance of stability is particularly important in assessing trends and when the application could result in unwarranted fluctuations from period to period (e.g., in the valuation of an insurance company's liabilities or pension plan contribution levels), especially when the fluctuations cannot be effectively explained. Not only will the actuary's credibility suffer, but there may be little justification for such volatility, especially when the long term is more important than the short term. Thus, the application of smoothing and grouping techniques, as well as weighing the relative importance of responsiveness in a particular situation, may be warranted.

Some applications require only short-term projections, for which a different type of modeling or forecasting may be more appropriate than the methodology that is the focus of Section 9. The degree of uncertainty involved can represent a fundamental difference between short-term versus long-term projections. Short-term projections (e.g., one week to three years) usually take the form of simple extrapolation or time series analysis, possibly adjusting for known drivers of change over the period. However, such extrapolation techniques may have poor predictive power over a long-term period.

Consistent with actuarial principles, the development of mortality rates needs to consider the population to which they are to be applied. If a set of mortality rates is derived from a population different from that to which they will be applied (e.g., dissimilar socioeconomic groups), the differences should be assessed and adjusted, as appropriate.

In deciding whether or how to make this adjustment, the analysis of these differences (possibly age-adjusted) and their patterns are assessed. For example, available historical experience might indicate that the overall population of a country or a continent (e.g., Europe), which was initially presumed to form a reasonable basis for the mortality experience of a segment, e.g., an insurer's life insurance portfolio or a pension plan risk portfolio, is not similar after all. In addition, the differences in mortality between age ranges might be assessed, especially those of particular relevance to the application.

This is particularly relevant where the population segment studied consists of a significantly higher percentage of those of a higher income and socioeconomic group than that of the general population. This issue arises since, in many analyses, it has been shown that overall mortality levels have been lower than the general population and rates of mortality improvement have, over the last several decades, been greater for those of the higher socioeconomic groups compared with those of lower socioeconomics. Such differentials are a prime reason the mortality experience for life insurance and private pensions has been lower than that of the general population. Explanations for this difference—and there are many—include differential access to health care services, standards of living, smoking prevalence, nutrition and educational attainment.

Each population segment tends to differ with respect to mortality risk characteristics. For public policy purposes, applications can include the assessment of public health hypotheses and the evaluation of past, existing or proposed public policy. For the actuary, possible applications include setting underwriting guidelines, developing insurer and pension program benchmarks/goals, pricing, reserving, valuation and capital adequacy studies of products, books of business and company or pension plan performance. The results or recommendations emanating from an analysis for another application need to be thoroughly vetted before application to a different application or population.

A recent approach used to study U.S. mortality is to analyze differences and heterogeneity in mortality by county and their average but disparate socioeconomic status (e.g., Barbieri 2022). As this type of data has become available, it has also been used for other purposes, including the use of its quantile values to estimate the mortality of certain population segments. Each county is assigned a socioeconomic index based on data from the U.S. Census Bureau that can be used to assess the heterogeneity of the overall population. However, it has to be remembered that the mortality for a given quantile is only a proxy for mortality experience for a given population segment, because each county usually consists of people with different risk characteristics. Nevertheless, this source split into quartiles or deciles can be used as a practical indicator of mortality differentials and a benchmark for the derivation of the mortality of population segments, such as those covered by life insurance or pensions.

Another approach to assess the expected financial effect of lower mortality of a group of insured individuals often used by actuaries is to study mortality weighted by the amount of insurance inforce, rather than weighted by the number of individuals or the number of insurance policies the insureds have. Similarly in a study of the financial effect of the mortality of a pension plan's population, pension benefits could be used. In effect, these weightings apply a greater weight to the financial metric at-risk—both the numerator and denominator of the mortality rate for the group. Thus, for example, the actual mortality rate expected by the same group of individuals might be 94% of an expected mortality table based on face amount of insurance, in contrast to 102% of the same table based on number of inforce policies. Recognition of the smaller percentage weight considers not only the expected financial effect of mortality, but also the expected mortality of those at a higher socioeconomic level.

Another issue that may arise is the potential bias due to the method used to apply the mortality estimates or projections. The mortality rates used to project the financial impact of a program using the average current amount of a stream of payments to a group of individuals are an example. If, as is often the case, those with a higher average payment (benefit) experience lower mortality, the future average benefit for that payment cohort will tend to increase over time; if the current average benefit size is used, the resulting financial value of the stream of payments will be underestimated. Although one method to correct this bias is to adjust future mortality rates accordingly, a better approach may be to apply a factor (modeled to reflect the effect of this differential mortality) to the future benefits. Another is to not use current average benefits, but rather to apply alternative mortality rates applicable to average benefits for those in different benefit size groups (or each individual's benefit separately).

In some cases the user might prefer either as low or as high a valuation as possible, corresponding to an acceptable or reasonable range of mortality and mortality improvement rates. For example, in a pension plan, a sponsor (contributor of funds) may desire to minimize its contributions to the plan: this would mean that assumptions used for those ages regarding postretirement mortality would be as large or mortality improvement would be as small as possible. Another example is that, to assess the financial condition of an insurer, a supervisor of an insurance company offering life insurance may wish to use mortality rates at the high end of a range of mortality or the low end of a range of mortality improvement for a somewhat stressed scenario in a conservative manner. For a tax authority, the opposite view might be taken.

In certain applications, the mortality rates used are dictated by an external entity, such as by an applicable tax or regulatory authority. In this case the actuary may have limited or no options. Despite this, if this external constraint provides a misleading or an egregiously erroneous result, the actuary may wish (depending upon the authority's rules and standards of practice applicable to the situation) to disclose or otherwise quantify the amount of the difference between the use of the dictated rates and those the actuaries would have used absent the rule. In any case an understanding of such a bias should be understood and appropriately communicated to the user.

Who bears the mortality or longevity risk can be relevant in certain cases. For example, for participating insurance policies it may be the group of policyholders themselves rather than the insurance company. As a nonguaranteed element in a Universal Life Insurance policy, it may be a combination of the beneficiaries and the insurer, depending on the level of guarantees provided. For a private pension plan, it could be the plan sponsor for a defined benefit plan or the pensioner or family for a defined contribution plan.

Examples of items for actuaries to consider in instances where mortality levels or rates of improvement based on one segment may not be appropriate to apply to another segment, such as the population of a country and an insurance company, follow, which may represent examples of asymmetric information by the insurer and applicant or insured:

1. *Population segment scope*. Many private insurance programs are marketed or distributed to a specific population segment. This target market(s) may, for example, consist of a group of those with better-than-average socioeconomic risk characteristics with respect to mortality.
2. *Underwriting selection/rules/manual/guidelines* regarding risk classification. The insurer with a high degree of intensity in its underwriting selection (e.g., determined based on historical or expert medical opinion) regarding applicants who have certain existing medical conditions or indications will tend to experience lower average mortality rates. Often premiums will be offered in a graduated schedule, reflecting expected mortality outcomes, for example, as a result of current adverse smoking habits or a surgical/disease history. The extent or conservatism of the underwriting criteria applied will affect the speed and period over which lower mortality would expect. Such an underwriting screen can never be "perfect," as other risk factors may exist or accidents can occur after underwriting. The select period (to develop select and ultimate mortality rates)[8] can differ, also depending on the intensity of underwriting selection—possibly 5 or 10 years if a simple underwriting screen is used (e.g., being actively at work at the time of application) or 20 or 25 years depending on the extent or effectiveness of underwriting selection and age. Mortality rates in the period immediately after underwriting can be less than half that experienced in the ultimate period (after the effects of underwriting have worn off).
3. *Policyholder (anti-)selection*. Offsetting to some extent underwriting selection is the extent that insureds accept offered insurance coverage. These insureds include those with lower expected future mortality who may be offered lower premiums by another insurer. Those offered insurance at higher premium rates by the insurer, particularly in a competitive market, may also not take (accept) the insurance offered, either because of a concern with affordability or because a competitor offered insurance to the applicant at a less expensive rate classification. In addition, after policy issue, those with better expected mortality may decide they no longer need insurance

[8] For individually underwritten life insurance policies, the use of a select period to reflect the extent that the selection or underwriting can reduce or eliminate (or charge a higher premium for) those applicants with known reported mortality risk factors. The period is the number of years over which mortality for the policy cohort is expected to be lower as a result of the selection or underwriting. The ultimate (after the select period is over) mortality rates are not expected to differ by period since policy issue.

or may be able to obtain insurance at a lower rate, while at the same time, other insureds who perceive that they are a higher-than-average mortality risk will tend to stay in the program. Thus, those who voluntarily terminate from the policy cohort tend to have better mortality than those who remain with the insurer. Experience has indicated that after initial selection or underwriting, mortality tends to increase compared to that at the same risk classification in the earlier policy years (that is, at the same attained age), usually converging toward general population segment mortality rates over time.

4. *Premium change*. If the insurance premium substantially increases (e.g., 10 years after a 10-year renewable term life insurance product is issued), the mortality for a specific attained age for the block of remaining insureds will increase, largely because the remaining population cannot obtain continuing insurance coverage at a more affordable rate or is no longer interest in this coverage.

Risk categorization and mortality estimation and projection are dynamic processes involving changes in participation and their recategorization over time. For example, in a population study, births and immigrants are new entrants to the population, while deaths and emigrants are leavers. In life insurance, there will be new sales and increases in coverage, as well as voluntary lapses, surrenders or reductions in coverage, exchanges or deaths as leavers.

Even where the same population grouping is used for both observation and application, changes in key risk drivers (the nature and number of the risk drivers can also change) in the environment (both external to the population and changes in the risk profile of the population itself) and the effects of survival (or new entrants or voluntary leavers) on the mix of population in the projection period (mix risk, e.g., if the composition of the population or population segments changes over time) should be considered or adjusted for.

This does not, however, suggest that complex models of mortality are needed to capture every nuance or risk driver—that would be impossible. Remember that a model is only a representation of reality and should not be overly granular or refined relative to the application being addressed. Nevertheless, it also does not mean that relevant refinements should be avoided for the sole purpose of simplicity.

If a sizable portion of the population consists of new entrants or leavers, their experience should theoretically be studied where possible, although that is rarely practical to do so. The significance of this dynamism depends on the situation.

5. The Metrics

Before describing the estimation or projection approach, the metrics used need to be discussed. The two primary ones of interest are the mortality rate (in actuarial literature, q_x, the number of deaths divided by the corresponding exposure) and the rate of mortality improvement (or deterioration) of the mortality rate, that is, the percent change in the mortality rate from one year to the next.

Because of the almost universal increase in mortality by age after the teenage years, most all-cause mortality studies are based on age, either individual age or age groups, such as every five years[9] (usually

[9] Five-year groupings are referred to as *quinquennial*. This age range is often used if there is an insufficient number of deaths or number of deaths for credibility purposes. If five-year data groupings are used, an interpolation formula is needed to allocate the average rate over the quinquennial ages where individual age rates are desired and special formulas are needed for data, for example, for at least the first year of

treating the first year of life and ages above 100 in a separate manner). Also because of the almost universal differences involved, gender is the second-most differentiated variable used. These two variables are also used because, in general, they have been reported consistently and accurately.

If comparisons are being made over time or across populations covering an other-than-narrow age range, mortality rates are often “age-adjusted,” that is, weighted by a consistent population at each age (or age group) of a representative population at a single point in time. This composite metric is also commonly used for summary purposes. This concept can also be used to adjust for any characteristic that proportionally changes over time, such as being age and gender adjusted.

Age can be measured on either an age-last-birthday or age-nearest-birthday basis. The former is the commonly used day-to-day reference of a person’s age. In contrast, in gathering data over an annual period, age-nearest-birthday or equivalent may be used. In any case the numerator and denominator used in calculating a mortality rate need to be determined consistently. In addition, the population to which the mortality rate is to be applied also needs to be categorized in a consistent manner. This is particularly important at older ages when the difference between rates at consecutive ages can differ significantly.

Mortality, normally determined on an annual basis, can be expressed in terms of either a mortality rate (probability from the beginning of a year; for example, measured by calendar period or common point of reference, such as date of disability within a calendar year) or a force of mortality (also referred to as the instantaneous rate of mortality or central death rate) centered in the middle of the year. Care is needed when rates are rapidly changing, for example during the first year of life, at the oldest old ages or when a spike occurs during a year. An annual mortality improvement rate is simply expressed as a ratio of one year’s mortality rate divided by that of the prior year minus one, which can be either a positive or negative percentage change. Mortality rates can also be studied in logarithmic form (e.g., Gompertz or Lee-Carter).

Metrics can be expressed as the expected (mean) value, median or mode. Among the units that can be used include the number of individuals, policies/employee headcounts, lives or face amount, the last reflecting the financial effect of mortality for a life insurer.

The forecaster can develop a best estimate or a probability distribution of the metric, the latter being useful to develop stochastic projections or to directly study the uncertainties involved. As discussed later in this paper, a range of alternative scenarios can also be developed.

To assess data reliability, both the numerator and denominator of the mortality rate have to be evaluated, usually expressed in terms of number of lives (but they can also be expressed in terms of number of policies or amount of insurance or pension benefit):

- Numerator: number of *deaths* at a particular age during a specified period.
- Denominator: *Population* measure. Often a midyear amount of population exposure is used. However, if there are new entrants (e.g., immigrants, new sales, or births) or nondeath leavers (e.g., emigrants, voluntary terminators or policy lapses), an assumption can be made regarding when, on average, they occur. In some cases the average exposure over the calendar year period

life, policy or the first year of disability, as applicable. For most of the middle adult ages (e.g., those aged in their 20s through 50s), the mortality rates within the group usually do not differ greatly (that is, the mortality curve increases only gradually at these ages), so that the average rate within such a group of ages may be sufficiently accurate, representing about the average age within the group. However, if the population exposed within the group is uneven or has a significant bias, then more refined interpolation methods may be needed to estimate rates for individual ages.

> or the value at the midpoint of the year is used. Depending on the formula used, those who died during the year are given full weight for the year—otherwise, in a population of one person, the annual mortality rate would be greater than 100%, which does not make sense. Special treatment may be needed at age 0 or the first policy year if a calendar year study is used. Care is especially needed in periods during which the mortality rate is rapidly changing.

One of the most widely used summary measures of longevity is the life expectancy of an individual or group of a certain age, such as at birth or age 65. This metric, particularly when measured at birth, in some cases may mislead a user. This is partly because life expectancy at birth is far more sensitive to changes in mortality at younger ages than changes at older ages. As a result, trends using this metric can prove misleading or hide key components of changes in mortality that may move in the opposite direction of the aggregate metric and thus may not be a good indicator of underlying mortality trends. Nevertheless, life expectancy can be useful for communicating aggregate results and, with appropriate caveats, in comparing mortality across populations or population segments.

Life expectancy at an age such as 65 can serve as a communication metric in applications such as for pensions or retirement studies.

It was common decades ago for actuaries, particularly in the study of mortality experience for life insurance, to study mortality on a policy year basis, that is, deaths and their corresponding populations covering the period from the anniversary of the issuance of an insurance policy to the next, and so on. This meant that one policy year overlapped two calendar years except where all policies were issued on a particular date, such as New Year's (which can happen in some group life insurance contracts). A disadvantage of this approach is that it takes longer to fully examine the results of a policy year's mortality, thus not being as responsive to recent trends because it didn't include, on average, the last half a year, and may not be consistent with the company's internal management reports, external financial statements or general population data. Because of these disadvantages, most mortality studies of life insurance and annuities are now performed on a calendar year basis. Nevertheless, studying mortality by policy year (or disability year if measured from the date of a disability) can be useful, because mortality can be sensitive to the period reported on.

Reporting lags have to be considered. It can take a considerable amount of time to completely report on and process deaths. For example, for internal management purposes, private sector life insurance insureds can take two or three months; for intercompany studies it can take more than six months, depending on the resources applied to such reporting. For national reporting, the time lag is only as fast as the reporting process, which can take a longer period. Details of deaths (e.g., cause of death) or population (e.g., ethnic origin) may take more time to validate before these data become available to users.

Actuarial tabulations can be developed (incurred but not reported [IBNR]) through actuarial triangles, studying historical reported data through a consistent number of weeks, months or another period measured from the end of the reporting period.

6. Data

The initial step in the analysis of mortality is the identification of reliable and relevant sources of data and information. Particularly if the objective of the study is to describe and understand historical mortality experience, the historical experience of the population being studied would be most relevant. However, if that population is of insufficient size or has structurally changed between the time when data are available and the projection period, reasonably comparable sources with a larger amount of reliable data would be used.

If the objective of the study focuses on the future, a population as similar as practical to the population for which a forecast is being made is also desirable to attempt to optimize relevance. However, in many cases the amount of data either is of insufficient size or may not be sufficiently accurate or comparable to rely upon. As a result, a national or regional database, or that with a large roughly similar population segment, may be looked to as a base data source. In some cases especially for the most recent period, only aggregate outcomes may be available, with appropriate adjustments necessary.

Some of the more important publicly available mortality-related databases include the following:

- Multinational populations
    - Human Mortality Database (HMDB). This validated source includes national data sets by year, with separate rates (and life tables) by cause for U.S. states and selected countries for which reliable data are available. For the countries, mortality rates by individual (and quinquennial and decennial) ages and gender, population and deaths, and life tables are available, on both a period and cohort basis. The historical period covered differs by country.
    - Population Division of the United Nations. This source provides mortality (quinquennial) rates and life expectancies by gender.
    - World Health Organization, WHO Mortality Database. Available by *International Statistical Classification of Diseases* coding (currently tenth revision, ICD code) from 1950.
- United States population
    - Social Security Administration, Office of the Chief Actuary. Its website includes mortality rates, life tables, population and deaths by individual age. More detailed studies are available on an irregular basis.
    - Centers for Disease Control, National Center for Health Statistics. Its Wonder database is available by cause by year for the last two or three years. Also, its Rapid Release data are available by quarter by gender, cause of death and decennial age groups.
    - Society of Actuaries. Their website includes experience gathered from life insurance companies for individual life and group life insurance, individual and group annuities, and public pension programs.
- United Kingdom
    - Office of National Statistics (ONS). Various data regarding deaths and population.
    - Continuous Mortality Investigation (CMI). Various tabulations of mortality rates for life insurance companies and pension schemes.
    - Biobank. A large database of genetic and environmental risk factors regarding chronic diseases, from more than 500,000 lives of those recruited at ages between 40 and 69 during the 2006 to 2010 period.
- Canada
    - Statistics Canada. Annual quinquennial mortality data by gender.
    - Canadian Institute of Actuaries. Various tabulations of experience for individual life insurance.
    - The Office of the Superintendent of Financial Institutions (OSFI). Detailed population mortality studies are published every three years.

Mortality rate data should always be assessed with a degree of skepticism. Reasons for this skepticism include the following:

- *Age reporting*. Especially at older ages, age assignments may include inaccuracies. In some cases particularly at the time of reporting, people can identify themselves in round number terms (e.g., ending in 0 and 5 or date of birth as of January 1). This is especially the case for some immigrants whose personal documents are not readily available. The use of quinquennial age groupings can reduce the significance of this source of error.
- *Racial, ethnic, gender and marital reporting*. In some cases, given that racial and ethnic groups are in some cases loosely defined and population and death data may be incomplete or inconsistently reported (e.g., particularly in situations where parents or grandparents are of different origins or where a substantial number of other-than-legal immigrants are present), accurate distinctions may be impossible to obtain. In recent years, people in some countries are able to self-identify as multiple classes or none of the above. If this type of reporting expands, racial, ethnic or categories that used to be relied upon may no longer be consistent over time and may prove less useful. In a period where the mix of married, partnered or solo agers and types of family structure is changing, care is needed when studying the mortality of these groups over time.
- *Assignments or attribution of deaths*. This is of particular concern in a study of by-cause data, due in some cases to weak recording processes or lack of training of those who record this data. The existence of multimorbidities, especially at older ages when multiple disabilities arise for a person increase, also can make unique and consistent attribution problematic. In some cases attribution of a primary cause of death is difficult, as is seen with reported COVID-19 deaths for which COVID-19 may be either the primary, joint or secondary cause. Because ICD coding in some cases has been inconsistently reported or has been implemented beginning at different dates, resulting in discontinuities in cause-related data, the data may have to be adjusted if the aggregator of the data did not do it.
- *Lack of data*. In some cases data, especially relating to risk characteristics, are not available or are limited. Important examples include drivers of death, such as smoking or obesity, where information may have to be gleaned from supplementary or independent sources.
- *Self-reported data*. In some cases the only source of data is that reported by the individual. This is of special concern when significant differences arise from professionally measured or identified information. In some cases the results of studies of bias introduced by this type of source can be used to make approximate adjustments, especially for mortality drivers such as weight and height.
- *Insufficient number of deaths* (credibility concerns).[10] Because of the sparsity of deaths at younger ages, statistical fluctuations may make patterns or trends difficult to discern. Some statistical bureaus do not provide data for categories with a small number of deaths where privacy concerns may arise.

Methods have been developed to offset or adjust for some of these data concerns. These range from grouping similar but broader categories (e.g., use of quinquennial age groups or groupings of causes of death), using a greater amount of data or smoothing of mortality rates, being somewhat conservative (depending on the application), using supplementary data, or relying more on external or similar but larger groups for use of their patterns by age or other variables.

Limited or no reliable data may be available for a population segment. A similar population or an aggregation of similar population segments may have to be used. If, for example, a small country, state or

[10] Statistical credibility in this context is the number of people in a population and number of deaths that provide a given/desired level of statistical confidence in or stability of the results, avoiding an undue amount of fluctuations. It is also a measure of the reliability or confidence that can be given to a set of data in the aggregate or according to one or more variables.

employer is being assessed, the experience of a larger aggregation or group of countries may have to be used. In some cases aggregate relativities and comparisons between populations may be made: in this case an aggregate adjustment to a set of reasonably shaped age and gender rate structures could be made, often applied in a multiplicative manner, for example, in a situation in which there are significant differences in gender mix in an age group or the average income of the two populations. The more that is known about the other group or groupings of groups will facilitate determining whether or what type of adjustment is appropriate.

A source of potential bias exists in using an inter-insurance company or multiple pension plan experience is that the mix of contributing entities may change over time, thus affecting risk profile trends over time, or the company or plan being assessed may be sufficiently different that the patterns or trends of the larger group may not be sufficiently relevant or reliable for the analysis being made.

7. The General Method of Periods

The framework within which the methodology described here is presented is based on five periods of time measured from the assessment date, in the past and the future. The following five-step process can provide a rigorous and constructive approach to follow in the mortality improvement projection process and for use in supporting decisions in an application for which future mortality is an important assumption.

Separate considerations and techniques are associated with each step. There are two major segments in the estimation or projection timeline: (1) *estimates* of mortality rates and mortality improvement rates during the historical period (prior to the assessment date), using one or more data models, the .XXXXs in Table 2, and (2) *projections* of mortality during the future period over which a projection is made, using one or more forecasting models, the .YYYYs in Table 2. These can be characterized as estimation and projection processes, respectively. Although they do share certain similarities, they also illustrate differences.

Table 2
PERIODS

	Historical Period						**Projection Period**						
	Period 1–Historical with Experience				Period 2–Historical with Limited Experience		Period 4–Transition Period			Period 3–Penultimate Period		Period 5–Final Ultimate Period	
↑	.XXXX	.XXXX	.XXXX	.XXXX	.XXXX	.XXXX	.YYYY	.YYYY	.YYYY	.YYYY	.YYYY	.YYYY	.YYYY
	.XXXX	.XXXX	.XXXX	.XXXX	.XXXX	.XXXX	.YYYY	.YYYY	.YYYY	.YYYY	.YYYY	.YYYY	.YYYY
Age	.XXXX	.XXXX	.XXXX	.XXXX	.XXXX	.XXXX	.YYYY	.YYYY	.YYYY	.YYYY	.YYYY	.YYYY	.YYYY
	.XXXX	.XXXX	.XXXX	.XXXX	.XXXX	.XXXX	.YYYY	.YYYY	.YYYY	.YYYY	.YYYY	.YYYY	.YYYY
↓	.XXXX	.XXXX	.XXXX	.XXXX	.XXXX	.XXXX	.YYYY	.YYYY	.YYYY	.YYYY	.YYYY	.YYYY	.YYYY
	.XXXX	.XXXX	.XXXX	.XXXX	.XXXX	.XXXX	.YYYY	.YYYY	.YYYY	.YYYY	.YYYY	.YYYY	.YYYY
	.XXXX	.XXXX	.XXXX	.XXXX	.XXXX	.XXXX	.YYYY	.YYYY	.YYYY	.YYYY	.YYYY	.YYYY	.YYYY
	.XXXX	.XXXX	.XXXX	.XXXX	.XXXX	.XXXX	.YYYY	.YYYY	.YYYY	.YYYY	.YYYY	.YYYY	.YYYY
	.XXXX	.XXXX	.XXXX	.XXXX	.XXXX	.XXXX	.YYYY	.YYYY	.YYYY	.YYYY	.YYYY	.YYYY	.YYYY
	.XXXX	.XXXX	.XXXX	.XXXX	.XXXX	.XXXX	.YYYY	.YYYY	.YYYY	.YYYY	.YYYY	.YYYY	.YYYY
	.XXXX	.XXXX	.XXXX	.XXXX	.XXXX	.XXXX	.YYYY	.YYYY	.YYYY	.YYYY	.YYYY	.YYYY	.YYYY
						← YEAR	YEAR →						

The steps applied in this methodology correspond to these five periods:

1. *Historical period over which relevant and robust experience data are available*. This period usually ends one to three years prior to the assessment date. These estimates may be directly derived from publicly available information or proprietary information more relevant to the application.
2. *Historical period over which limited or no relevant rate experience data are available*. The estimates for this period are usually for a short period, from one to three years. In some cases only aggregate data are available from which to derive these estimates. Statistically, it could be looked at as a set of missing or incomplete data. It can serve as the starting point for mortality projections.
3. *Penultimate mortality rate improvement[11] period*. This period usually begins anywhere between 10 and 30 years from the assessment date.
4. *Transition mortality period* between periods 2 and 3. Although for this period the rates might be determined using a simple interpolation or curve-fitting technique, the period can also be split into a more refined initial period, possibly influenced by individual causes or drivers of death, followed by a simple extrapolation until period 3 is reached. It can begin either at the end of periods 1 or 2; if at the end of period 1, there would be no period 2.
5. *Final ultimate period* of mortality improvement rates. The improvement rate (often a single annual percent change) effective after a specific date may be the same rate as is used for period 3. The beginning of this period may be expressed in terms of a specified number of years from the assessment date, possibly 40 years, or at a certain age, possibly between ages 95 and 110. The annual improvement rate for this period is usually quite small or even 0%, which is applied after that period or age. It is used to close off the projected mortality table.

The demarcation between the historical and projection periods, referred to here as the assessment date, can either be the analysis date or a date in the immediate future, such as the reporting date of an entity or another date when the projection begins. This date can depend on the application but is usually close to when the mortality study is conducted.

The following two sections discuss each of these periods.

8. Historical Periods

After obtaining an understanding of the objective and context of the context and objectives of the application, the next steps in the estimation or projection process are to determine the best process to follow and to obtain a set of historical data or information that can form the basis of the forecast. These data are normally organized by age and gender.

A study or development of estimates of historical mortality rates for a well-defined population can enhance the understanding of the experience or history of that population, compare outcomes among groups or risk categories, or if an objective of the study is predictive, can serve as the baseline from which mortality rates can be projected. It should be kept in mind that the population studied may not be appropriate for use in the projection process for a different population.

The depth and type of analysis of historical data may differ depending on whether the intent of the analysis is purely descriptive, that is, to develop an enhanced understanding of the mortality risk processes or

[11] Although often described in terms of "improvement," this can either constitute a decrease or increase from the prior period.

patterns, or whether the objective is for it to be predictive, focusing on the use of historical rates to form the baseline from which the projection of the trajectory of morality rates will be based.

Rates derived from a population or population segment represent a sample or estimate of the underlying experience of the population over the period studied, because any data derived from or about that population are not necessarily indicative of its intrinsic nature. There are several sources of volatility in (and reasons the reported mortality data differ from the inherent) mortality rates for a given population, including statistical (random fluctuations for which the frequency and percentage size of the fluctuation tends to be larger for smaller populations), changes in underlying exposure in time or space, reporting anomalies and errors, seasonality or other regular patterns and permanent or temporary disruptions (in the form of either a short-term spike or longer-term wave).

If possible, the effects of random statistical fluctuations (i.e., not due to an identifiable cause) should be eliminated or reduced. Often this is done by the application of a smoothing method (sometimes referred to in actuarial literature as graduation or interpolation or extrapolation). A tension between stability and responsiveness to underlying experience, especially nonrandom volatility, often arises. How this is addressed will depend on the application and desires of the analyst and the user of the results.

A regression, interpolation or other numerical analysis method is often used for this purpose. Since trends can differ by age, a form of two-dimensional smoothing can be applied. North American actuaries are familiar with the Whitaker-Henderson B method (order 2 or 3), in which the user can control for a combination of fit and smoothness. Other methods include the use of P-spines. Average rates covering five-year intervals can be used, with interpolation performed to develop rates for annual ages.

Available data may be “presmoothed” by the developer of the data, eliminating mortality rate outliers, especially at or near the mortality table’s boundary edges. Appropriate inquiries regarding the extent and form of such smoothing should be made to help determine whether any inadvertent patterns were created or destroyed as a result of the smoothing (or lack of smoothing). This relates to not only data but also visuals. For example, regarding heat maps, it may be useful to understand the number of consecutive years combined for visualization purposes.

The need and desirability for, as well as the extent of, smoothing must be determined for all historical periods, partly as a recognition that mortality is in many respects a stochastic process that is affected by a myriad set of factors, both as input as well as output. This is particularly appropriate because the source of the data may not be of sufficient size to directly produce desired output in the form of smooth mortality values over time or space.

Section 8.1 describes methodologies that can be used for period 1 for which historical mortality rates are available, and section 8.2 discusses the recent or current period for which limited or no such data are available, as in period 2.

8.1. PERIOD 1: HISTORICAL PERIOD OVER WHICH RELEVANT AND ROBUST EXPERIENCE RATE DATA ARE AVAILABLE

In assessing whether or how to use a given set of historical experience data, several factors need to be considered, including the relevance of the population whose mortality data are available, length of the period and age range used, limitations of the experience data and any significant disruptions that have occurred during the experience period.

The population on which historical experience data are based should be as relevant to the application as possible, considering its risk profile and the reliability of the available data. Although it could be the population to which the forecasts are being developed, this may not be practical.

The length of period from which data are obtained is chosen may be based on the following:

- *Period over which reliable data are available*. In some cases the data in the initial part of the period for which data are available should not be included in the study period, because the data may be of questionable quality or not relevant to a future period.
- *Length of projection period*. A common rule of thumb sometimes used in long-term mortality forecasting has been that the duration of the projection period should be at least the length of the historical period from which mortality improvement factors are developed. Nevertheless, the application of this rule of thumb can be problematic because of its lack of relevance to the projection period; that is, it may not bear any relation to the conditions of the period to which the results are applied. If the period covers a wave or shock, both the up and down parts of the pattern should be included.
- *Relevance of the early part of the period*. Experience 40 or 60 years ago may bear little if any relevance to the projection period. For example, the then key drivers of U.S. mortality change might have consisted of the Vietnam War experience, high smoking prevalence, low water quality and high air pollution levels, limited treatment of cardiovascular risk factors or an environment quite different than that of the expected future period, which will produce biased or misleading results.
- *Importance of recent period*. Responsiveness to recent trends can be quite important and relevant to the expected experience in the projection period. For example, in the 2010s, the rate of mortality improvement was lower than that of the prior decades, as suggested in Figures 1 and 2.
- *Avoidance of mortality shocks*. A temporary period of unfavorable or favorable conditions, as an outlier, may be avoided, at least to the extent that they do not represent permanent shifts or are unlikely to reoccur.

In sum, the analyst should understand the source(s) and period of mortality data being assessed, both regarding accuracy and reliability and relevance to the population assessed and the period of study.

The smoothing of mortality rates can reduce the effect of significant fluctuations in the pattern of mortality rates. Nevertheless, some will remain, and these are evident in population heat maps (as in Figure 2, the data of which had already been smoothed). Of course, some volatility should be expected, whereas other fluctuations may suggest or even hide underlying trends and drivers of changes in mortality.

Types of mortality patterns can be categorized in terms of the following:

- *Level*. The level of expected mortality can be specific to a population segment or an individual based on the applicable risk profile. It can be affected by many factors, some of which may be discernable given appropriate data and information but may be difficult to recognize because of data or information limitations. Where credible, it is usually assessed through current observations, although an average over a recent experience period is sometimes used.
- *Trend*. Actuaries have often assumed a small regular annual mortality improvement (possibly between 0.75% and 1.5%, usually graded downward at older ages), based on the average improvement of mortality over the last several decades or even a century. However, it is difficult to see why mortality will continue to increase at about that pace over a long future period, given what may become drastically different conditions. In some cases, especially evident in the recent decade-long slowdown (other than as a result of COVID-19) in mortality improvement in many

countries (mostly, but not exclusively, because of a reduced level of improvements in cardiovascular mortality), the assumption that the rate of historical trends will continue may produce biased results.

- *Cycle*. It is easy to conclude that mortality will inevitably continue to improve on average at the same rate as in the past, even though this has not been and may not continue. As seen in Figure 2, trends in certain age groups appear to have exhibited cycles, although the risk driver(s) of a cycle may be difficult to discern. The mortality of a group may periodically offset one trend with a subsequent trend in the opposite direction that then results in multiple waves (or cycles). The drivers of some cycles are clear, such as those driven by periodic technology improvements. But, for example, understanding is limited of the longer-term mortality cycles for those aged 90 and older over the last several decades, suggested by the heat map in Figure 2. Although Figure 2 relates to the entire U.S. male population, these cycles may occur more frequently in certain population segments.
    - Seasonal fluctuations are short-term examples of a mortality cycle. For example, influenza is a mortality risk every winter. Some of the fluctuations in the frequency of such deaths or illnesses between years will occur as a result of the severity of the flu strains each year, as well as the effectiveness and coverage of flu vaccines and other mitigation techniques (e.g., effective masking used against COVID-19 has also proven effective against influenza). In addition, people tend to spend more time indoors in the winter, with implications for nutrition and physical activity. Analytical approaches to reflect this volatility include separating such deaths from nonseasonal ones and observing mortality on an annual or 12-month rolling basis. Where partial-year data are available, a comparison to the experience of previous seasons can be made. Since seasonal influenza is common, an adjustment might arise only in the case of an extraordinarily bad or good influenza year.
- *Temporary discontinuity or disruption*. Temporary mortality discontinuities and disruptions usually do not affect longer-term mortality. This is the case because a reversal (rebound) of a temporary disruption in mortality levels also often occurs (whether in the same year as the initial discontinuity or in subsequent periods, depending on the nature and duration of the discontinuity and type of recovery), although residual effects sometimes continue.
    - *Spike*. A spike is a sudden and significant change in mortality of a limited duration. There are several types of spikes, that is, temporary discontinuities or disruptions. A spike can result in a sudden, even catastrophic (if the spike is an adverse one) temporary shift from the immediately prior level or trend. Its effects can arise in a brief period such as a day, month or year. Examples are the results of a terrorist attack, armed conflict, earthquake or a short-lived epidemic or pandemic).[12] Note that health after-effects of a longer-term nature can include excess morbidities, such as long COVID, mental illnesses or exacerbation of an existing illness.
    - *Wave*. In contrast to a spike, a wave can take place over a relatively long period, characterized by a significant shift in mortality in excess of (or reduction from) a mortality benchmark over a longer period than a spike, followed by an opposite shift in the direction of the benchmark or trended level of mortality. The period involved can be long, depending on the risk drivers and reasons for or nature of the wave. An extreme case has been the adverse health effects of smoking combustible tobacco cigarettes. In many countries smoking has been common for a long time, exacerbated by a long lag time between the period of smoking and its consequential adverse health impacts (possibly 30 or more years). Another example is the

[12] The distinction between an epidemic and a pandemic is primarily due to the extent of its spread and severity. Its outbreak may be assigned such a label by the World Health Organization (WHO).

AIDS epidemic that has affected certain population segments for 40 years, although without pharmaceutical intervention the period between infection and adverse health conditions is relatively short. The actuary should recognize both the initial wave period, as well as the level or trend that it reverts to and the speed of its reversion (which can differ by demographic characteristic, mitigation effectiveness, cultural and geographic context, and period). Note that it can be difficult, especially in the wave's early stage, to project its trajectory.

- *Permanent discontinuity*. A permanent shift in the level or its trend can occur, possibly because of a medical breakthrough, a pandemic's endemic period, or a permanent change in culture, behavior or law. Such a change may be structural, for example, if the health care infrastructure is reformed or a residual part of a pandemic becomes endemic. As just noted, it can be difficult to distinguish a permanent discontinuity from a temporary one; this has to be determined on a case-by-case basis. For example, a slowdown in mortality improvement was experienced in many high-income countries in the 2010s; even a dozen years later, it is uncertain whether this slowdown has only been temporary or will be permanent.
- *Chaotic element* (including random noise). Often these consist of unrecognized, offsetting and numerous minor causes. In aggregate, they have often been referred to as the residual or unexplained portion of a change. In developing tables, graphs or heat maps, these fluctuations are often smoothed. Although, individually, each of these elements may not be directly attributable to a single cause or its effect is minor, it can be valuable to examine any significant outlier to determine whether it is truly a random fluctuation or an indicator of a future disruption.

Multiple processes and risk drivers may act simultaneously or on a delayed basis, with either a favorable or adverse effect. Cause-and-effect (direct consequences), association (with indirect or uncertain relationships), confounding (unrecognized factors), feedback (outputs of a process affect the process itself) and cascading (a string of a chained set of) risks can result in an increased level of risk, uncertainty or even catastrophe. Time lags between the risk driver and unexpected (or expected) adverse (or favorable) mortality outcomes, such as in the case of smoking or obesity, can be considerable, possibly decades. Thus, a lack of early recognition (early warning) of an adverse disruption can lead to a significant drag on mortality improvement. Even the identification of their underlying risk drivers can take a long time, for example, the adverse effect of air pollution, climate change, smoking and weight gain from excessive intake of high-fructose-sugar drinks.

Estimates of mortality or mortality improvement rates can be derived from a combination of relevant and available sources, possibly weighted in importance based on the aggregate number of lives or deaths. An unweighted average might be used if detailed population values are not available or relevant. One approach to develop mortality rates for a population segment is to use a set of rates of a national population as a base or *reference population* to enhance stability, together with a set of rates from a relevant population segment, for example, for all pension plans or even a specific group of people. These could be credibility-weighted based on size, supplemented with an aggregate age-sex-adjusted or other selected ratio of the particular population of attention to the reference population. For a small country, state or population segment or multiple neighboring or similar population groups, the reference population might be based on weighted or unweighted values. If the period of the most recently available data is not the same for the sources used, appropriate trend adjustments would then be needed to obtain a consistent set of aggregated data.

In any case a thorough analysis of differences between the reference population and the population to which the mortality rates will apply should be conducted to determine whether the reference population is sufficiently relevant and timely and what adjustments, if needed, should be applied. An example is the application of national rates to an insurance company: an adjustment may be needed if the socioeconomic characteristics of the insurer's risk portfolio are not sufficiently consistent with those of the entire country. An insurer's portfolio may consist primarily of higher-income people, who may experience, especially considering underwriting selection, a different level of mortality or mortality improvement than the population as a whole.

Special treatment may be warranted for particular age groups. For example, because of the larger number of deaths at birth or at least within the first month measured from birth compared with those during the remainder of the first year of life, special exposure formulas may be needed for the first year of life. Although it is common for most ages to assume that deaths are uniformly spread throughout the calendar year, this may also not be the case at the oldest ages, during which there is a rapid increase or a high level of mortality rates. The actuary has to determine how significant these situations are relative to the analysis being conducted and treat them accordingly.

The amount of available data may not be sufficiently credible (either according to a statistical measure or by judgment based on the observed volatility by year or age) or suggest a need for adjustment, which may require interpolation or extrapolation in some way, possibly weighted with data from a reference population or other source. This can arise because of a lack of sufficient volume, relevant data or changes in future conditions.

Adjustments are often made to even a statistically reliable set of data, for at least the following purposes:

- *Inadequate size of the underlying population*. Some form of smoothing is often applied, as a smooth pattern of mortality rates may be desirable (otherwise too much random noise would be reflected). Even using national data, all-cause data by individual (or even quinquennial) age may not be large enough for a sufficiently smooth set of mortality or mortality improvement rates. If the all-cause data are split into population segments or causes of death, for example, the size problem can be exacerbated.
- *Edge issues*. These can arise, for example, at old ages, young ages or the most recent year when only preliminary data are available, where data need to be adjusted or smoothed (e.g., through some form of graduation, such as one of the Whittaker-Henderson series, smoothing splines, kernel estimates or local polynomial fitting or a curve such as Gompertz). The method can be selected considering a subjective view as to the expected shape of mortality rates at the oldest-old ages (e.g., greater than 105) or mortality improvement rates. Smoothing can be quite important; if not done, mortality improvement factors will also be quite volatile.
- Age cells with a *very low probability of death or no or few deaths or exposures*.
- In some cases involving public data by detailed category, the number of deaths in certain categories is too small to be available because of *privacy concerns,* although in most cases these situations arise when there is a lack of statistical credibility anyway.
- Treatment of *incomplete data*. For example, data may be available only for a certain part of the age distribution or not split by gender. Decomposition of that data based on trends or other sources may be necessary.
- Experience is *inconsistent with expectations*. For example, conditions may differ between the historical and future periods, such as from historical or shocks-in-process due to an epidemic such as AIDS or drug overdoses, or an outlier such as a natural disaster.

Short-term mortality fluctuations (or disruptions) do not usually affect long-term mortality trends. Rather, their effect tends to dissipate totally or to have a relatively small residual (endemic) effect. Whether or how they are reflected in a mortality projection depends on several factors, some of which the actuary may need to consider depending on the particular situation or application, such as in pricing (premiums or contributions), reserving or financial reporting, risk management or capital assessment. Factors involved in a spike or wave can include the following:

- Its nature, in particular, what is(are) its driver(s) and the effectiveness of the mitigation techniques applied to control its adverse effects, if any;
- When it began, its severity, its trajectory and when it has completed its course, if applicable;
- Its effect by socioeconomic or sociodemographic characteristics, e.g., age, gender, employment status, race/ethnicity or disability status;
- Where in its trajectory the evaluator sits, e.g., in the initial rise or decline stage;
- If during its course, what are its prognosis and the uncertainty regarding its future trajectory, especially regarding its duration, severity and expected effects on the population being evaluated;
- Whether it has a period, cohort or hybrid effect;
- How it (or its mitigation) affects or is related to other causes or drivers of mortality;
- The reliability and quality of the supporting data used or considered, given the difficulties in attributing deaths to particular drivers and causes and their nature; and
- Its residual effects, in particular, whether it will establish an endemic or what a "new normal" would be like.

At the time this paper was written, at least two of these fluctuations may be relevant to actuaries and provide a range of spike and wave considerations:

- *COVID-19*. This pandemic materially affected population mortality in many countries from 2020 through at least 2022. It has affected those of all ages in one or more ways, with a greater impact at older ages, although a shift has been seen in the distribution by age in some countries throughout this pandemic. In the U.S., it heavily affected those at older ages (over age 75) in 2020, shifting to affecting younger ages in 2021, and increasing its concentration in those of older ages again in 2022. These shifts resulted from such factors as trends in vulnerability, percent vaccinated, mitigation actions taken and variants. The number of reported COVID-19 deaths may not be completely accurate or complete because of the difficulty in attributing causes of death or misclassification bias. As a result, a more reliable measure of its impact may be the "excess deaths" measure over an expected benchmark, although setting the benchmark may be problematic in some cases. It is not clear at the time this paper was written the extent of its effect on future mortality rates due to long COVID (post-COVID), delayed care for other medical conditions, and any consequential effects on or from other causes of death.
- *Drug overdoses*. Although this cause of death has existed for centuries, deaths due to opioids and certain synthetic drugs such as fentanyl have adversely affected mortality experience in North America for the last two decades that increased again during the COVID-19 pandemic. Drug overdoses are a significant part of the deaths of despair (due to drug overdose, alcohol abuse and suicide) that have mostly affected those in their 20s through their 50s, partly due to addiction, pain medication or mental health. These deaths have contributed to recent all-cause mortality deterioration at these ages and years, with significant differences by country and geographic area. The future course of this cause of death is currently uncertain, although it definitely will not go away.

Usually, the mortality from a spike or wave, if any, is included in historical all-cause mortality experience data. If deemed appropriate, an adjustment could be made to the baseline mortality experience from which a projection is based to explicitly remove the historical effect of a significant past or current spike or wave, with future expectations regarding any remaining effects included explicitly. This may be especially important if the population being assessed is different from the source of the historical experience used or if the period to which the projection is to be applied is not expected to have the same type of exposure to the spike or wave. For example, the actuary may wish to reduce (or increase) or eliminate its effect in period 2 and/or future periods, depending on the circumstances. The method applied will depend on the specific application and available data.[13] In any case a projection of the future trajectory of the spike or wave may be appropriate, to the extent estimable. If the mortality of the spike or wave has been eliminated from the historical mortality experience, any expected remaining portion of the spike would be added to the "normal" expected future mortality.[14]

In most cases the "final" smoothed or adjusted mortality rates and resulting improvement rates for period 1 may represent only fairly crude estimates, although, as indicated, for certain ages and categories, especially those at or near the boundary edge of the available data.

8.2. PERIOD 2: HISTORICAL PERIOD IN WHICH LIMITED OR NO RELEVANT RATE EXPERIENCE DATA ARE AVAILABLE (USUALLY A SHORT PERIOD, FROM ONE TO THREE YEARS)

For period 2, lags in reporting detailed (including, for example, age-specific or cause-related) death data can range from a few months to several years, depending on the reporting processes involved. Unlike the historical data in period 1, data for this period may either be incomplete, available only in the aggregate (e.g., the total number of reported deaths) or not available at all. Some lag between death and its reporting or recording is to be expected; life insurance companies commonly receive experience death claim reporting between a week and several months.

Corresponding lags in exposure or population (the denominator for mortality rates) can also occur. For general population values, an accurate census in many countries is conducted once a decade, with intercensal extrapolation or interpolation often developed by the applicable government census or vital statistics division. Extrapolations from the prior census are often developed with the benefit of estimates of trends in demographic decrements or increments, including births, deaths and migrants.

Similar techniques can be applied to other types of population segments, such as employment rolls for active and retired employees or participants for pension plans and inforce records for life insurance companies (e.g., reflecting new business, voluntary terminations, policy exchanges and deaths). Note that for most such private-sector programs, the lags involved are periods specifically allowed for (e.g., the grace period for premium submission) and reporting lags by the participant or the processor.

Before discussing the peculiarities of and development of rates for this period, it is important to note that its estimates can serve two purposes, to (1) represent estimates of mortality and mortality improvement rates for this period or (2) form the starting point or baseline from which future mortality improvement factors are applied. Although rates developed to satisfy each objective do not have to be the same, if different, the reason and basis for this difference should be disclosed and documented. Because the estimated annual rates can be volatile and the rates can be developed to satisfy one or the other of the objectives, discontinuities between rates may arise between the estimated or projected rates for the last year of one period and the first year of the next one. In any case the effect of these discontinuities, in the

[13] For example, if a spike is estimated to increase base mortality by 20% in a year and the population studied is expected to experience about half the effect, then for the demographic segments affected the historical mortality would be reduced by [−(1−1/120%)/2] = 8.33%.
[14] For further discussion of significant fluctuations (disruptions), see the chapter of the International Actuarial Association's *Risk Book* entitled *Mortality Disruptions* (forthcoming) by Gutterman.

aggregate, should be minimized if practical. Further smoothing between the rates of the periods may be desirable.

Similar to mortality rates in period 1, the estimates for period 2 may be derived from a combination or weighting of data from several sources, although they may be different from the sources used for period 1. Data for this period may be available only in an aggregate form, or for the most recent period, no reliable data may be available at the time the estimates or projections are made.

A spike or wave in progress at the time of the analysis would usually be presumed to continue in some way during period 2, although, depending upon when the evaluation is conducted, its trajectory, including its rebound period, may not yet be discernable. Given the short period involved, the relative importance of period and cohort effects that have been observed may continue, unless evidence arises to the contrary, although the difference in outcomes for this short period between these effects may not be material.

An extrapolation of the period 1 mortality rates or another source appropriate for this purpose (benchmark rates) based on either (1) recent or longer-term historical mortality improvement, (2) aggregate trended experience for period 2 applied to the most recent smoothed experience or (3) a combination of (1) and (2) might be used. In any case a comparison of any reliable aggregate information that is available, such as the raw number of deaths during the period compared with that of the immediately prior period, is a useful validation metric to either confirm the reasonableness of the extrapolation or estimation or to apply an aggregate adjustment to rates of the immediately prior period. Where available, this comparison should be performed on smaller groups as well, such as a broad age group, where available.

The extrapolation to the rates in period 2 might be performed by a numerical analysis technique, such as regression, P-splines, a Whitaker-Henderson fit or average annual rates of mortality improvement applied to recent rates. The base period from which extrapolation is performed could cover rates of between five and 15 years, possibly unweighted or reflecting a trend, depending on the situation. If weighted, it might be decided that the more recent part of the period should be assigned a greater weight. The base period might exclude one or more outlier years considered not to be representative and thus excluded from the calculations, for example, if it included a spike, such as an epidemic or pandemic.

If the extrapolation can be performed for each attained age separately or independently, the differences in rates between age groups should be reviewed for consistency with other age groups. In some cases the rates could be directed by an expected spread between ages or age groups, with adjustments for significant deviations if there is no known reason for them. It could also be performed through the application of a formula (data curve) representing multiple age groups, which would avoid the necessity of interage group adjustments.

It may be appropriate to compare for each year in period 2 (1) the estimated number of deaths derived from multiplying the mortality rates by the corresponding population, (2) the estimated number of deaths applying the baseline rates if different than (1) used as a jumping-off point for period 2 to the estimated population and (3) the corresponding actual aggregate number of deaths for each year. If the resulting differences are significant, further analysis and adjustment if warranted in the estimation methodology or assumptions may be called for to reduce any discontinuity between the rates from the end of period 1 and those of period 2.

If the objectives of an analysis are directed only to the development of a forecast or projection of future mortality rates, rather than the development of reasonable estimates for period 2, a separate determination of rates for period 2 may not be needed. In this case the transition (period 4) rates would begin immediately after period 1. Separate estimates for period 2 may prove worthwhile in any case to validate the rates for the early part of period 3. In this case forecasts developed would then begin immediately after the end of period 1. Discontinuities may nevertheless still arise between the last period of period 1 and the beginning of the transition period.

9. Projection Periods

Developing a projection presents numerous challenges, because the future is by definition difficult to foretell. Nonetheless, methodologies are available that can assist the process followed. Most forecasting exercises require a set of relevant historical data (i.e., estimates derived for rates in periods 1 and 2), together with insights into whether historical patterns of mortality and population are expected to continue as they have been or to take another trajectory. Although the use of a strictly statistical model may be used, judgment remains a key component of any projection process.

Historical mortality trends show mixed patterns, reflecting changes in the population and conditions being studied that may not be relevant in the future or to particular population segments. Many important developments have led to dramatic changes in mortality over the last century—for example, improved water quality and sanitary systems, antibiotics and immunizations, cardiovascular risk factor mitigation, and the implementation of Medicare and Medicaid in the U.S.—that will likely make smaller contributions to mortality improvement going forward.

The future will see drivers of improvement in the areas of new breakthroughs in medical technologies and treatments. But adverse developments will also be inevitable, such as the effects of the substantial increase in the prevalence of obesity and diabetes, as well as possible pandemics and antimicrobial resistance. Other factors will involve effects that are difficult to predict, such as government action, new technologies and human behaviors that have not even been thought of yet.

Much of the discussion so far relates to situations in which many or at least some relevant data are available. What happens when we have no recent relevant or reliable data? Historically, a safe harbor forecast might have been a 1.0% to 1.5% annual improvement over a national population life table, possibly lower at older ages. However, there are at least two caveats:

1. Overall population experience levels may not provide a sound basis for the initial mortality rate starting point. For that purpose, an understanding of the risk profile of the population for which the projection is being made is needed. In particular, an indication of how this population is similar to or differs from the reference country's (or region's) risk profile would be assessed. To the extent this information is available and understood, the reference population mortality rates (or an adjustment of the reference population, as appropriate) might be used.
2. Similarly, the trend in recent historical mortality improvement for the general population (excluding the effects of COVID-19) may not be appropriate for many population segments. Experience over the last several decades suggests that there has been a significantly different pattern by socioeconomic groups (see section 9.1 for further discussion). Nevertheless, it is unlikely that such differential trends will continue indefinitely.

As the experience of COVID-19 has shown, turning points or the existence, trajectory and severity of current mortality disruptions can be problematic and quite difficult to predict. Although many experts believed that a future pandemic was inevitable, few were so bold as to incorporate its severe effect in their quantitative forecasts. Nevertheless, understanding the nature of such events or trends on an early warning basis can help in addressing such developments (see section 12 for further information).

Section 9 assumes that developing an estimate of the "ultimate" mortality rate improvement represents the first step of the projection process, followed by a projection of how current experience (period 2) can lead or transition to the ultimate mortality rates. However, an alternative approach could reverse these steps: an extrapolation of mortality from current levels could be the first step, followed by an estimate of medium and ultimate levels of mortality improvement. In either approach, current estimates and ultimate rates need to be coordinated or linked, because the foundation of the ultimate rates by attained age is based on the relevant baseline rates and the cumulative mortality improvement between the end of period 1 or 2 and the beginning of the ultimate period 3. The following assumes that the former approach is used. In any case the final step in mortality projection involves determining a set of final mortality improvement rates ("closing off the mortality curve").

Consistent with the CMI (2009) model, long-term (ultimate) rates of mortality improvement are determined (separately by age, possibly gender, cohort or cause components) by user inputs for the convergence period. Effectively, this approach assumes that in the immediate term, an acceptable baseline (starting rates) and likely subsequent rates of change in mortality rates are based on the most recently observed experience and best estimates of the short-term future. In the long term, the forces driving mortality change may be different than those currently influencing patterns of improvement, although what the drivers will be and their significance remain highly uncertain. Therefore, the long-term rates may be better informed or at least supplemented by expert opinion input and analysis of long-term historical patterns and key causes driving them. Over time, the relative weight placed on the recently observed past versus the more subjective longer-term view can change accordingly. This is also consistent with the approach taken by the model currently used by RPEC and the Society of Actuaries Research Institute (2021) MIM model.

9.1. PERIOD 3: ULTIMATE MORTALITY RATE IMPROVEMENT FACTORS

Three key assumptions regarding the ultimate mortality rates for period 3 are (1) the starting base rates, (2) the period at which the ultimate rates are achieved and (3) the ultimate rates of improvement.

The starting base rates are usually derived from the last year of periods 1 or 2, depending on when the forecasting method is first applied. As noted above, the base rates may or may not be the estimates for the period immediately prior to the beginning of the forecast period. They may differ if the mortality rate estimates are not sufficiently smooth for forecasting purposes.

Actuaries assess the latter two assumptions according to the objectives of the application, the population to which they are to be applied, preconceived notions regarding the future and desired refinements. Practice has differed for each of these assumptions.

A key decision needed is the number of years until the ultimate improvement rates are achieved. In practice, a single period is usually used, although that is not necessary. Alternatively, it can differ either by attained age, gender or period measured from the valuation date. Common periods used in North America and the United Kingdom for most ages range from 10 to 25 years. The U.S. Social Security Administration trustees, in their annual report, indicate that a 25-year period for overall population projections is used until ultimate improvement rates are reached. In most cases the period used tends to be set by judgment, depending upon the experience studied and what was used in prior studies.

Several considerations or situations might influence the length of this period, including the following:

- The size of the difference between current mortality levels and the expected level at the beginning of the ultimate period, that is, the expected slope of the transition (discussed in section

9.2). The larger the difference, the longer the period until ultimate rates are reached that may be used.

- If the mortality rate pattern used is a select and ultimate scale (often applied in the pricing of individual life insurance with underwriting selection performed), the ultimate period used is often the same as the period for the select part of the mortality table. The select period represents the period over which the effects of selection (considering the target market, the extent or intensity of underwriting, and antiselection by the insureds, e.g., as a result of voluntary termination of those with favorable risk characteristics, further discussed in section 4) are expected to wear off.
- If applied to a group of disabled lives (e.g., to set the liability for active claims for an insurer), the period can be selected to be consistent with those who are not disabled at those attained ages.
- The period does not have to be the same for all ages and genders. In some cases this period for older ages can be shorter than that for younger ages. The period could represent a certain number of years since a policy was issued in the case of insurance or from the assessment date, partly because the effect of some underwritten mortality drivers may wear off at older ages.
- If mortality by cause is being studied, a separate period might be used for different causes.

Because of the increasing uncertainty regarding mortality rates or improvement rates over the period, mortality rate improvement in the ultimate period is usually less refined (i.e., with greater aggregation) than is studied in an historical period, particularly as compared with the considerations for periods immediately prior to the assessment date.

Approaches that could be taken include the following:

- Use of *expert opinion*. For example, the Society of Actuaries' Retirement Plans Experience Committee has based its ultimate rates of mortality improvement primarily on expert opinion (the consensus of members of this committee). That said, expert opinion can differ, sometimes widely, although it is often based on the long-term experience of those participating in the decision making. If the advice from a group of experts is considered, the rates used can be developed through various methods, including discussion that results in agreement or a Delphi consensus approach involving multistage inputs. These differences in opinion can result from many factors, such as the experts' backgrounds and experience.
- *Disaggregation* of mortality rates into causes of death and/or mortality risk drivers. This disaggregation can be reflected explicitly or implicitly, or simply ignored. This type of analysis may provide insights into what future patterns and rates may look like, although the process followed can be quite complex. Projection of long-term mortality rates on a per-cause method is currently only used (that is, publicly disclosed) by the U.S. Social Security Administration's Office of the Chief Actuary. This office develops ultimate (using a 25-year period) assumptions for seven major causes of death (the seventh being "all others"). Their assumed ultimate rates of improvement are the same for all genders. The values are determined by judgment and agreed upon by the trustees of the programs, based on a combination of observations of historical long-term experience and judgment. Issues that need to be resolved using this approach include the handling of currently unidentified or underestimated causes or drivers of death, the relationships among causes that can be problematic, and if using primary causes of death how to treat the "other" category. In addition, in the long term, the rates for the cause with the lowest rate of improvement can dominate the projections.
- Use of a *target rate*, set either by a scientific or statistical method or by the judgment of experts, the analyst or a combination. Two examples include the following:
    - In their World Population Prospects (2019), the experts of the Population Division of the United Nations select a target for life expectancy at birth for females, estimate the difference

in these expectancies between males and females, develop mortality rates for quinquennial age groups for subsequent five calendar year groups, and interpolate from them to derive mortality rates for each gender by individual age and year.

- The most recent U.S. Census Bureau projections converge to an ultimate Life Table for 2150 based on a survey of experts regarding mortality improvement trends and rapidity of change until 2150, with mortality rates extrapolated from current levels based on deaths from the National Center for Health Statistics and population from the U.S. Census Bureau and past trends, using the method described in Lee and Tuljapurkar (1998).

The key risk-related variables that appear to drive annual mortality improvement rates need to be identified and relativities and trends determined, often at least based on a historical review and the assistance of expert opinion. Expectations as to whether and how changes in these relativities and trends will occur are needed, based on a comparison between historical mortality drivers to those expected. As indicated above, this analysis is performed over the most appropriate period over which historical trends are based.

Although it may be statistically desirable to use a very lengthy period as a base, such as 50 or 100 years, it is difficult to believe that experience 100, or even 50, years ago can be relevant to rates of mortality and its improvement for a period between 20 and 50 years in the future. In contrast, choosing too short a period runs the opposite risk: of being overly responsive to short-term trends or only one part of a mortality disruption (similar to the problems of financial modeling that only reflects part of a business cycle).

Looking at a heat map of a country's mortality experience (possibly using a rolling two-to-five-year average to reduce volatility; see Figure 2 for U.S. males) suggests some of the significant trends, cycles and waves that have occurred in a relevant population by period, age group and cohort, some of which may or may not be suggestive of subsequent changes in mortality. However, that information may be useful for use in developing rates for period 4.

Although the multiplicity of relatively minor, possibly offsetting, short-term and long-term trends and disruptions may increase reliance on long-term all-cause averages or trends, certain patterns may be important to recognize (e.g., the general relationship between the mortality of males and females has been reasonably consistent over long periods). This type of relationship may at best be problematic, for example, because of significant differences in smoking prevalence, as opinion is split as to whether this difference will continue to gradually narrow or remain broadly similar to today's level.

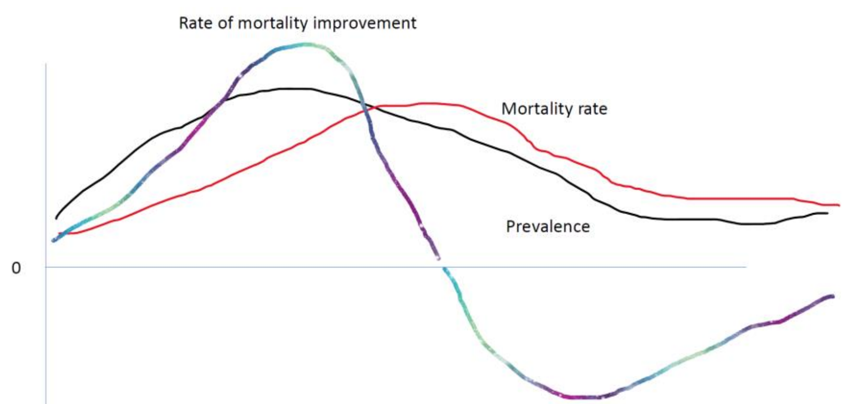
The future of mortality in many high-income countries will not continue to benefit from the favorable mortality improvement of the last several decades that resulted from favorable long-term trends or waves. For example, this includes an increase in effective cardiovascular disease risk factor control and the long trend in the decrease in smoking prevalence (note that over the short term, a large number of older women still experience adverse mortality because of their relatively high smoking prevalence of several decades ago). The increase in obesity in the United States since the 1970s, which at least to some extent may represent an analogous mortality risk driver that will adversely affect future mortality, has begun to have a significant effect.

The effect of previous and current drivers of adverse mortality can reverse in sign as the frequency or prevalence of the driver diminishes. To do this, the actual and expected effects of the driver's lifecycle should be assessed. For example, mortality improvement in the United States has been driven, with a long lag, by a significant amount of smoking cessation (from a peak for males since the 1940s and 1950s and females since the 1970s, although there remain many remaining smokers and a growing number of vapers). After a huge runup of adverse smoking-related premature deaths, for many years overall mortality experience benefited from a large amount of smoking cessation. Basing future expectations on a period in which only favorable trends were experienced may produce biased projections.

Another example is whether the large favorable effects of cardiovascular risk factor diagnoses and treatments will continue. It is now not clear whether the percentage using these mitigation techniques will continue to increase, stabilize or diminish, although it is unlikely that, without significant improvement in diagnosis or treatment, the resulting favorable improvement trends in cardiovascular disease will be able to continue. Further examples are easy to identify.

An illustration of the effect of a mortality wave can be seen in Figure 3, although the three lines included are not drawn to scale. The mortality wave generally begins with the emergence of a significant driver or cause of death that peaks and then dissipates. The effect of the condition, such as smoking, may have a significant lag between prevalence or exposure and consequential death. Mortality improvement is the rate of change in the mortality rate for the condition, turning negative immediately after the peak of the condition's rate of mortality. The contribution to all-cause mortality improvement depends on (1) the percentage of the all-cause mortality rate contributed to by this condition, (2) the mortality level of the condition and (3) the rate of either improving or deteriorating mortality. The effect on all-cause mortality in the period of mortality improvement in the last part of the wave is often forgotten.

Figure 3
MORTALITY IMPROVEMENT (DETERIORATION) OF A TEMPORARY CONDITION



Not drawn to scale. Horizontal line is calendar time.
Prevalence is the percentage of people who have a given condition.
Rate of mortality improvement is the annual percentage change in the *mortality rate* caused by the condition, becoming negative (deterioration) after the mortality rate decreases.

Some mortality drivers have a finite life, diminishing with aging and as time goes on, or just continuing; it may be difficult to distinguish these patterns in some cases. It is unrealistic to believe that the effect of every existing and potential mortality driver will continue forever, even after applicable lags are reflected. It may be important to distinguish mortality drivers and mortality improvement drivers; the former, although continuing to have an effect, will not have a continued effect on mortality improvement.

The examples also suggest that the use of a historical period may generate biased forecasts unless the increasing and decreasing effects of significant temporary and permanent disruptions are appropriately distinguished. This concern might be overcome by incorporating the entire life of temporary drivers or stages that generally corresponds to the projection period for a different driver. These mortality drivers can arise in areas such as public health practice, human behavior, communicable diseases and medical diagnosis or treatment. For at least this reason, an understanding of historical, current and expected future mortality drivers may be important in the process of developing mortality improvement forecasts.

It is difficult to determine, early in 2023, whether (1) the robust mortality improvements seen in the 1990s and 2000s in many higher-income countries, particularly due to cardiovascular diseases, will continue or (2) the flat or at best gradually improving experience of the 2010s will continue. In addition, it is uncertain whether the increasing differentials in improvement over the last 50 years between those with unequal income will continue to widen or remain at their current levels. These are important questions that will resolve only over time.

The following are examples that may warrant consideration in the development of improvement factors, although the use of too many factors may overcomplicate the model used. If this occurs, too many independent assumptions may be required, and incorporating the expected effects of interrelations between the factors may be difficult, which could introduce inappropriate or unintended results, especially over the long term.

- *Age*. Mortality drivers and causes of death can differ dramatically by age group. It is also common to assume an age gradient beginning at older ages, such as 80 to 90.
- *Gender*. Although the ultimate mortality improvement factors in some mortality forecasts include separate factors by gender, not all do. On the one hand, distinct factors require separate schedules of mortality improvement and twice the number of estimates, which will make the model used more complex and may lead to what might become inadvertent divergent or inconsistent forecasts by gender in the long term. However, many drivers will differ in their effect on changes in gender-based mortality, e.g., a later decline in smoking prevalence by women and a lower prevalence of excess alcohol use. It is usually better to make an explicit decision regarding the desired relative importance of internal consistency and model complexity.
- *Geographical* area. Examples of population segments that might be used include country or region and rural or urban areas. Similar to gender, because of their cumulative effect, applying a significantly different set of factors between two or more geographical or population segment groupings can lead to unexpected differentials in mortality, especially over the long term.
- *Birth cohort*. Birth cohort has been shown to have had a significant effect on mortality in some cases and some periods. The scope of a cohort can range from those exposed to a sudden event to up to 30 or more years, depending on the type and severity of the driver(s) involved (see the diagonal patterns in Figure 2). To help determine the significance of a cohort and the period over which its effects are or are expected to be observable, it is important to attempt to identify the driver(s) of cohort differentials, although the attempt can be controversial. In some cases its effects might be shown to be of a short-term nature, or at least be time limited; in others, it can involve more than a generation—an example of the latter is the impact of smoking on a generation. The effect for a population segment (e.g., a cohort that contained an inordinately large percentage of smokers) could continue for quite a while because of the lagged effects involved. But with a high percentage who cease smoking, its effect on that cohort will dissipate over time. Another example is if a birth cohort was subject to a reduction in nutrition for a lengthy period at younger ages (due, for example, to a war or a financial depression), the corresponding effect on that cohort's mortality may be long lived. In any case, if the effect of the mortality disruption driver is temporary, a reversion in trend may be possible for the succeeding cohort (see section 8.1 for a discussion of mortality spikes and waves). Usually these differential effects wear off at older ages.
- *Socioeconomic* or sociodemographic category, for example, by income, wealth and educational attainment. Similar to other factors that can be used to differentiate mortality and mortality

improvement among population segments, the impact on mortality from one or more of these factors can be significant. Although educational attainment is fixed once an individual reaches about age 25, income level and wealth can be dynamic, that is, they can change over a person's lifecycle.

- Changes in *educational attainment* levels in a population, given the lagged effects involved, can have a strong long-term effect on mortality. Although studies have concluded that educational attainment, at least in the United States, is strongly related to both future mortality rates and in some cases rates of mortality improvement, the strength of the cause-and-effect effects may not be easy to demonstrate. Other factors, such as the percentage of the population in poverty and the quality and accessibility of health care services, may prove to be just as relevant to mortality or longevity. Even if the extent of educational attainment does not have a direct causative effect on mortality, changes in educational level are indicative of other simultaneous mortality drivers and may provide an early-warning indicator of long-term trends—it points to the importance of family, neighborhood and peers involved in growing up.
- *Employment status*, particularly if the application relates to the mortality experience of a pension program (e.g., blue-collar or white-collar jobs). Mortality studies of a group of employees often distinguish between groups of employees because the experience of higher-income employees (and their dependents) and those with less physically demanding jobs may experience a higher rate of improvement than those with physically demanding jobs. This differential has existed for at least the last 50 years. Note that as a consequence of stress on the job, differentials between employee segments in the future may not be as significant as those in the past. However, to the extent those differentials also reflect differences in socioeconomic status or condition, they may continue.
- *Behavior* or *other disease risk factors*. The relevance and expected future of smoking and vaping prevalence and cessation has changed considerably over the past several decades, especially relevant because of the long lags between exposure and resulting premature death. The level of obesity, and in particular morbid obesity, has increased in many countries; it has been shown that peak levels, volatility, normal lack of success in long-term weight loss programs and period of organ exposure to white adiposity tissues all may have lagged adverse mortality consequences. In certain cases mitigation techniques are available but may not be utilized (e.g., medicine for high blood pressure or cholesterol, or bariatric surgery).
- *Medical advances,* public health care upgrades and research. Many exciting potential medical advancements are on the horizon, such as CRISPR-Cas9 and mRNA vaccinations. Other favorable developments are also possible, such as improved dementia treatment. However, the likelihood that these advances will be successfully implemented, along with the time it takes to affect mortality experience in the population segment in question, may be longer than hoped for. Also, those with higher income will typically receive roll-out benefits earlier than those with lower income.
- *Other external factors*. These include future epidemics and pandemics and environmental factors such as climate change. The characteristics of their effects include their incidence, duration, severity and residual effects.

As indicated, the historical relationship between the rates of mortality and its rates of improvement for the population segment to which the mortality forecast will apply and the corresponding population from which the historical mortality experience has been drawn needs to be assessed. A further dimension is whether this relationship will be maintained or changed, the rapidity of change.

Significant heterogeneity and inequality within a population may affect future mortality improvement of that population. All population groups consist of individual members with a diverse set of risk characteristics. In assessing whether or how this aggregation of risks will affect long-term mortality improvement, it should be decided whether significant subpopulations are expected to experience significantly more or less mortality improvement than that of the average of the population and whether their share of the population is expected to change significantly. Because of the longer expected survivorship of those with higher average benefits from many pension or life insurance programs, the average benefit of those who survive will tend to increase over time, while the age-adjusted mortality will tend to decrease. These movements may also impact evolving average mortality and improvement rates.

If an ongoing mortality disruption is considered permanent, the residual or continuing part of a spike or wave will be expected to continue, on a cohort and/or period basis, depending on the nature of the disruption. Although it is likely that temporary shocks or waves will not affect the mortality of the ultimate period, the forecaster may wish to utilize one or more scenarios that include additional shock(s) to or endemic parts of mortality, depending on the purpose of the projection. This is especially important in a study of the effect of uncertainties (see section 12), for example, including a range of pandemic scenarios considering both direct and indirect effects (e.g., regarding a currently in-process mortality disruption such as COVID-19), there may be consequential effects (e.g., on drug overdose, excessive alcohol consumption, smoking and obesity).

The approach taken can differ depending on the application, the applicable population or population segment, the extent to which population mix changes and the preferences of the analyst or user. Although refinements in explicit assumptions or implicit considerations may be theoretically appealing and be derived from historical experience, projections need to consider their expected permanence, interrelationships, trajectories and significance.

9.2. PERIOD 4: TRANSITION MORTALITY RATE IMPROVEMENT RATES BETWEEN 2 AND 3

In developing transition improvement rates for period 4, it may be worthwhile to assess the objectives of the projection during this period, which might consist of the following characteristics or choices:

- *Blend smoothly* between historical and future mortality rates, as well as not creating or inadvertently expanding discontinuities between ages.
- Whether the transition period should be viewed *as a whole*, that is, a single curve for an age group, or *split into two or more parts*. Multiple time segments may be desirable to be more responsive to recent trends or circumstances. For example, a more refined initial part of the transition period may be expected to be more responsive to recent mortality spikes, cohort(s), current or anticipated causes of death or risk factors, or simply as a continuation of prior trends using regression. This is in contrast to a less refined numerical or statistical convergence to an ultimate mortality improvement used for the later portion of the transition period.
- *How rapid* a transition is desired, that is, whether the percentage change in rates during the earlier part should be larger, smaller or similar to those of the later part of the transition period. This will determine the formula(s) used for transition purposes. For example, one possible approach would be to have a rapid short-term shift, with the end of the period converging more slowly to the ultimate (step 3) mortality improvement factors. The chosen steepness might depend on whether there is a significant difference between the rate of mortality improvement at the end of period 1 or 2, as applicable, and that of period 3.
- Since every projection involves some degree of *expert opinion*, the desired extent of opinion relied upon and used should be capable of being justifiable, documented and communicated to the user.
- Consider and appropriately communicate the *uncertainties* involved (see section 12).

In developing transition mortality or mortality improvement rates, a starting point (the estimated rate(s) at the end of period 1 or 2, as applicable) for the transition period needs to be set, with its ending point (year) at the beginning of period 4. If the starting point consists of a set of mortality rates applicable to a prior multiyear period, such as five years, a trend might be applied from the midpoint of that period to the beginning of the transition period (or the midpoint of the jumping-off year). However, in most cases the jumping-off point of the transition period would be the annual rates of mortality for the middle of the last year of period 1 or 2, as applicable. There are at least three possible starting rates, that is, relating to a specific age or range of ages:

- The estimates for the last year in period 2;
- The estimates for the last year from period 1, in which case if period 2 is used to estimate the most recent year's mortality rates, the transition rates might be independent of the last year's estimates, thus likely creating some discontinuities between the periods; or
- A trend generated from the last several years of mortality rates including the jumping-off year; this has the advantage of being able to use a line or curve that directly reflects recent trends, without the possible fluctuations from a single year's observation.

If the third option is used, several choices present themselves for short-term transition rate patterns, which may need to be adjusted further, especially if the historical period(s) has one or more significant disruptions. These could be applied to (1) all-cause mortality, (2) major risk factor or cause components or (3) either (1) or (2) adjusted for such factors as current or very recent spikes or edge conditions. Possible trend lines and formulaic techniques include the following:

- Numerical analysis trends applied from recent (possibly 10 or 15 years) experience, such as a linear trend or linear regression line on mortality rates or the log of these rates for a short period from the last year in period 1 or 2, as applicable, exponential smoothing or an autoregressive integrated moving average (ARIMA) model;
- A cubic polynomial;
- Horizontal and diagonal (to simulate cohort effects) cubic splines. This approach has been used by the Society of Actuaries' RPEC.

If the number of years of available experience is limited (e.g., three or five years), too few years may be available to develop an improvement rate curve, and alternative approaches, including significant reliance on judgment, may have to be applied.

If a mortality spike, such as a one- or two-year pandemic, occurred during the historical period, several approaches might be taken. Note that the inclusion of the effects of such a temporary shock can introduce a bias into the projection trend line, even if its effects have dissipated during the transition period. Note that the approach taken may depend on when it occurred and its severity. They include the following:

- *Ignore the period(s) during which the spike occurred*. If this approach is taken, it must be decided whether the immediately prior trend should be assumed to continue during the period ignored or whether the period should be totally ignored for trending transition purposes. The approach taken depends on the understanding of whether the disruption was independent of other mortality drivers or causes and whether the prior trend, if any, should be assumed to continue throughout that period. If, for example, a regression line is normally applied over the historical period, such as 10 years, and the disruption covered two of these years, the mortality rate estimated for the first year of the transition period could be derived from the eight-year period or a lesser amount corresponding to the 10-year period.

- *Ignore or estimate the effects of the disruptive type of death*. In some cases it may be difficult to attribute deaths to the spike. To assess how to treat excess deaths, several items need to be considered, including (1) how to identify them, either from separately identified or reported data or through an "excess deaths" calculation, and (2) whether a residual continuation needs to be considered. One approach would be to assume a continuation of the otherwise-present trend (similar to the immediately prior approach), equivalent to an excess deaths approach, for which the "normal" number of deaths or rates need to be estimated, possibly employing an average or trended set of mortality rates or subjective judgment. An excess death exclusion approach is used when there is some doubt as to data attribution accuracy or bias in the shock's attributed and reported deaths. These can result from indirect consequential deaths (e.g., due to deferred medical diagnosis or treatment or where multiple contributing causes of death are involved) or indirect related deaths (e.g., fewer influenza deaths due to masking taken to avoid an infectious disease). In addition but difficult to estimate, long-term medical research efforts may be interrupted and delayed, thus adversely affecting longer-term mortality. Note that some of the excess deaths may be deaths of those who were frail and who might have otherwise died soon anyway ("harvesting"), which might reduce future mortality rates immediately afterward, possibly lowering mortality rates in a subsequent period.
- *Set a range of alternative scenarios*, especially if there is a large amount of uncertainty involved, with each scenario reflecting plausible trajectories of direct and indirect effects of the shock.

Further adjustments may be needed for certain disruptions, particularly if in the form of a wave. For example, reductions in mortality could reflect the lagged effects of smoking cessation (see Figure 3) or to revert to an after-the-peak drug overdose trajectory. With respect to the treatment of a mortality disruption, historical data and projections need to be reviewed regarding whether an adjustment is needed to reflect its increase, plateau and decrease stages. In most cases these patterns will not affect ultimate rates (discussed in section 9.1).

If the pattern (e.g., age or cohort shift, increase or decrease) of significant recent drivers or causes of mortality are expected to change materially, it may be appropriate to explicitly adjust for them, at least in the early part of the projection period.

The transition, especially during the first part of the transition period, can be based on a period cohort (see Figure 2) or a hybrid approach (discussed in section 3) for, for example, the first five years. See the Society of Actuaries Research Institute (2021) MIM model for an example.

Although significant improvement has been seen in longevity over longer than the past century (with a few noticeable hiccups), as seen in Figure 1, this may not continue at the same rate in the future. Mortality deterioration has occasionally arisen, recently due in part to COVID-19 and drug overdose deaths in North America. Deterioration can also emerge after a significant temporary spell of mortality improvement. When this occurs, other than simply because of random fluctuations, especially in small populations, a known reason is usually involved. Recognition of the nature of the source of the mortality deterioration may suggest whether the deterioration will continue and for how long it might be expected to continue. In certain cases no special action will be warranted, whereas in others it might be.

It could be argued that any lengthy period of mortality deterioration will result in increased research funding or mitigation actions to counter its effects. However, depending on the cause, a reversal may take a significant period for implementation, especially for all communities in the population affected. As a result, all recent periods of mortality deterioration should be rigorously assessed and reflected appropriately.

9.3. PERIOD 5: FINAL ULTIMATE PERIOD OF MORTALITY RATE IMPROVEMENT RATES

The objective of period 5, the final ultimate period, is to close off the projection. It is common to assume that a final, relatively low rate of mortality improvement will be reached or begin at an advanced age or a lengthy number of years from the present. Although this, the final improvement rate, can be equal to the ultimate rates in period 3, it does not have to be.

The final improvement rate is usually expressed in a less refined manner than those in period 3, often selected by judgment or based on a long (or if by necessity, short) historical period. It is sometimes a constant improvement rate for all subsequent years. It can also be reached at a given age or period measured (or a combination) from the assessment date, sometimes on a linear pathway between the last year in period 3 and the final ultimate age or period, although it might be somewhat more refined. The effect of any current mortality disruption is not usually expected to continue through this period, although sometimes a simple adjustment expressed as a percent for all or a range of ages might be applied, especially in response to the uncertainties involved.

The method used during period 5 to transition to the mortality improvement rate for the year immediately prior to the close off of the mortality table at advanced ages is usually quite simple, either a constant or at a linear rate of decline, usually based solely on judgment. It also may be zero or a value such as 0.25% or 0.1% annually after an age as young as age 100, partly since limited data are available to base a scientifically robust estimate regarding mortality rates after that age. Sometimes this low improvement rate is reached at a more advanced age, such as 110, 115 or 120. At these advanced ages, most socioeconomic differences tend to be much smaller or nonexistent than at younger ages.

The age at the end of a mortality table is often referred to by demographers and actuaries as “omega,” the assumed limit of human longevity. The assumed omega is reached when the annual mortality rate is 100%. A significant amount of academic discussion has surrounded whether there is an omega and what it should be. But in most cases there is little practical reason to care whether omega is 115, 119, 121 or 130, because so few people will reach any of these ages. Although theoretically, several people in the world will live to these ages, for use in any financial and in most actuarial modeling, the choice of such an omega will make a negligible financial difference in any practical problem.

10. Modeling Mortality

Underlying a great deal of the prior discussion are two uses of mortality modeling: (1) to characterize and understand mortality rates and improvement factors based on historical experience (descriptive) and (2) to describe or hypothesize the future trajectory of mortality applicable to the population to which the projection is applied (predictive). Two key mortality-related metrics are the focus of actuarial modeling: mortality rates and mortality improvement rates. The latter is the focus of much of the above discussion and the approaches described, while this section primarily addresses the mortality rates themselves, with the ultimate outputs of both being mortality rates.

In the course of developing a mortality forecasting model, it should be recognized that the objective is to build a model(s) to project the future, which is inherently unknowable. In doing so, any aspect that does not appear immediately relevant, tractable or important should be ignored. This can help reduce the problem to its essence. At the same time, this hopefully will not lose significant information or impede understanding, enabling a focus on the major elements of the underlying assumptions and processes. Models by themselves cannot make decisions—however, model outcomes can provide useful and sometimes valuable contributions to developing sound decisions.

Desirable characteristics
Key desirable characteristics for a descriptive or predictive model include the following, the relative relevance and importance of which will depend on the application, as well as whether the focus is more on the short, medium and/or long term:

- *Responsiveness*, especially to recent experience, although not to every twist and turn of nonrepeating volatility, disruption or unexplained outliers;
- *Robustness* of the model and *reasonableness* of the model's assumptions;
- *Historical fit* (back-testing): throughout the analytical process followed, there is always a tension and tradeoff between goodness of fit (accuracy) and smoothness;
- *Forecast accuracy* and reliability;
- *Stability*, that is, not producing volatile rates each year of the projection or between projections;
- *Smoothness*, avoiding unexplainable discontinuities;
- *Practicality* and ease of application;
- *Simplicity*, avoiding undue complexity;
- Identification of *key drivers*;
- *Relevance* of the basis of historical data relied on;
- Maintenance of reasonable *relationships* between subpopulations, e.g., mortality between males and females;
- *Understandability* and explainability, including the basis for and output of the model, especially regarding the implications of the application; and
- *Well documented* and *reproducible*.

There are obvious tensions and trade-offs between some of these characteristics. For example, responsiveness and stability, fit and stability or smoothness, or relevance and reliability cannot be equally satisfied. A weighting or compromise among these characteristics or objectives may have to be accepted. For example, unwarranted volatility in mortality forecasts can lead to more expensive risk capital for an affected insurer.

Many mortality models and projections are deterministic in form, that is, they produce a single set of mortality and mortality improvement rates that vary by age and possibly other risk characteristics. For example, mortality models are used to develop national life tables, for life insurance or pension valuations, or use in MIM-2021 (Society of Actuaries Research Institute 2021). However, other models are stochastic, that is, based on one or more probability distributions, in which confidence intervals can be developed. Although further discussed in section 12, such distributions do not predict any type of significant mortality disruptions, although a separate model might be used for this purpose or specifically address permanent shifts in one or more parameters.

Expert opinions, individually or through a survey of experts, can be drawn on from either inside the forecast team or as external advisors, consultants or peer reviewers. Where practical, they should be selected from a range of professions or backgrounds, such as actuaries, epidemiologists, demographers, medical professionals, sociologists or academics. However, even when the "best" experts' views are relied upon, their findings may be subjective, inconsistent or biased (e.g., following the "herd's" conventional wisdom, and certain professionals focus on only what can be currently observed rather than likely future developments). Nevertheless, their consensus findings may turn out to be more accurate than solely relying on data or an individual's opinion. Aspects of a model being reviewed might include the process or methodology followed, the model applied or the parameters used in the forecasting process.

Visualizations are increasingly used to assess the reasonableness of mortality trends and communicate the results of the models. Heat maps (e.g., Figure 2) are prime examples, especially useful in identifying the effects of mortality disruptions and differentiating between period and cohort trends.

Projection techniques

Numerous mortality projection techniques are currently being used, including those developed by actuaries and demographers, some of which may have had many refinements proposed. Most techniques are based on a model, the choice of which can depend on the availability of relevant and reliable data, the experience of the analyst and the purpose of the projection. Transformations are often used, for example, using logarithms (logs) of mortality rates in the Lee-Carter method, rather than the mortality rate or improvement rate. Common variables used include the following:

- One-factor—usually age only;
- Two-factor—often age and gender;
- Three-factor—age, gender and period and/or cohort; and
- Four-factor—age, gender, period/cohort and another characteristic, such as socioeconomic or geographic.

Some techniques are fairly straightforward numerical analysis methods to interpolate or extrapolate the rates, including the following:

- Linear or exponential;
- Regression or ARIMA models;
- Generalized linear models;
- Age-period-cohort (APC) P-Spline model regression (used by the United Kingdom's CMI); and
- Whittaker-Henderson two-dimensional, order 2 or 3, commonly used in North America.

However, the use of these formulaic approaches can prove problematic. First, they often rely on historical averages or straightforward trends over time, while the complexities of the mortality process may not be considered. For example, in the first decade of the 2000s in the U.S., several observers strongly suggested that projections that did not reflect significantly positive improvement trends greater than 2% annually in mortality that had been observed in the prior decade should not be used on account of their inherent bias.[15] Shortly thereafter, however, rates of mortality improvement for most ages decreased significantly (due in large part to slowdowns in improvement in cardiovascular disease mortality and increases in the number of deaths due to drug overdose).

Some forecasters, reflecting more recent experience, have reduced the size of their mortality improvement projections, but it is not yet known what trajectory mortality will take in the future. Differences in expert opinion illustrate the difficulty in determining whether a recent (favorable or adverse) mortality disruption will be a temporary or permanent phenomenon.

An important question is how refined a model should be, particularly because in light of the difficulty a quantitatively based model may have in dealing with nonrandom fluctuations. Although a more refined approach may produce a more accurate fit for a given set of mortality experience, it may not produce a more accurate model under other conditions or periods. Nevertheless, it has to be remembered that, as

[15] For example, Soneji, Samir and Gary King. 2012. "Statistical Security for Social Security." *Demography*, 49, 3, Pp. 1037-1060. https://tinyurl.com/y66t3spw

the often-quoted George Box observation in 1976 states, "All models are wrong, some are useful."[16] Although a more refined model may produce a better fit in relation to a given set of data, that does not guarantee it will with respect to other risk factors, populations or periods. This is particularly applicable when a spike, wave, permanent disruption or structural shift arises.

Many approaches or formulas used are refinements of earlier ones. Among the earliest ones include one developed by Gompertz (1825), followed by Makeham (1860), the latter incorporating a somewhat more refined model that provided a better fit for mortality at older ages.

Another more recent example is the Lee-Carter model[17] (first described in Lee and Carter 1992). Its original version involved a decomposition of logarithmic mortality rates into an age-specific base level and a time-varying component (period effect), multiplied by an age-modulating parameter (age effect), fitting a bilinear model to past log-mortality rates using least squares. Subsequently, many variations have been proposed or used, usually enhancing the subsequent fit of a population projection, for example, those using a set of age group data, more recent periods (rather than beginning with the 1950s when reliable data became available), cause-of-death data and birth cohorts (e.g., Renshaw-Haberman based on a generalized linear model).

A limitation of all of these models is their assumption that a selected set of average historical mortality patterns and trends, including the age curve, will continue in the future (thus, choosing the period is all important). A key to all models is the underlying assumptions used.

Other models include those produced by Heligman and Pollard (1980), who used a mathematical formula over the entire range of ages—childhood, young adulthood and maturity, with multiple parameters—Cairns-Blake-Dowd (Cairns, Blake and Dowd 2006), who used a stochastic model, and Currie (Currie, Durban and Eilers 2004), who used a generalized linear model with an age-period-cohort approach. Other developments have been introduced by epidemiologists and demographers, including Perks (1932), who used a Gamma-Makeham formula, and Kannisto (1992), who used a logistic curve for older age mortality. In addition, several authors have utilized a range of probability distributions, such as the Weibull, Gamma, log-normal and Gini/Lorenz, the last of which has often been used to study inequality.

Predictive analytical methods, machine-learning techniques or multivariate analysis may be especially useful in the analysis of historical data when teasing out the most important variables in multivariate data or using multiple variables in a predictive model. Big Data have been used to supplement or augment the primary source of data and information. In many cases a single dimension or variable, even though significant by itself, may hide other important factors and drivers that would otherwise not be identified as being important in the estimation or forecasting process. These methods can also be used to back-test the outputs of other models (e.g., Deprez, Shevchenko and Wüthrich 2017). Other statistical tools, such as bootstrap sampling approaches, have also been used.

As mentioned several times in this paper, many factors affect mortality and mortality improvement rates. Multivariate methods (such as using Cox proportional hazard ratios; Cox 1972) can be used to estimate or

16 Box, George E. P. (1976), "Science and statistics" (PDF), *Journal of the American Statistical Association*, 71 (356): 791–799, doi:10.1080/01621459.1976.10480949

[17] $\mu_{ag,yr} = \alpha_{ag} + \beta_{ag}\kappa_{yr} + \varepsilon_{ag,yr}$, where α is the age shape, β is the age-specific pattern of mortality change and κ is the year trend. After fitting α and β by maximum likelihood, a time-series model, typically a random walk with drift, extrapolates the index linearly over the historical period used, resulting in constant rates of improvement.

project differences in the hazard or force of mortality experienced by groups that have different risk characteristics. To develop a single set of mortality best estimates based on a single mortality response curve (e.g., Ludkovski, Risk and Zail 2018), these methods may not rely on a single distributional assumption or a predictive analytical method, although typically they do not reflect many of the factors discussed in this paper.

Relational approaches have also been used. They reflect relationships between the subject population and a reference population for which more reliable data are available or where understanding is greater, for instance, by gender or socioeconomic groups. Maintaining or expanding upon the knowledge of the relationship between key variables may also enhance the overall resulting estimates and projections.

No consensus has yet been reached (and may never be) as to what is the best approach or model, as there is no single model that performs best for all ages, years or populations. Even ones accepted by many, such as time series (Lee-Carter), have had numerous modifications and refinements proposed over time.

Risk factors or causes and drivers

An alternative to the use of mathematical models incorporates an understanding of the effects of mortality drivers and/or causes of death, which can be crucial in developing sound and supportable mortality models (Gutterman and Vanderhoof 1998) or a combination of drivers and causes and mathematical models. The following describes key aspects of these risk factors and causes of death that might be reflected in such models that are most relevant and material to the estimation and forecast of mortality.

Intentionally, this section does not constitute a thorough discussion of these drivers and causes; that is outside the scope of this paper. Note that this type of knowledge may prove incomplete and even disproven as time goes on, whether because of a lack of reliable data, inadequate understanding or changing conditions.

Nevertheless, these drivers and causes can provide valuable inputs or considerations in understanding the past and present and forecasting the future. They can be reflected in either an implicit or explicit manner. Implicitly, any recent or expected changes are often considered in deciding the methods used or adjustments to historical estimates made, no matter what model or approach is used for the short and medium term. They can also help in deciding what type of model should be used.

Many actuaries consider historical trends in causes of death or risk factors in developing their projections, especially for the short-term or intermediate future, eschewing their use in the long term because of the uncertainties involved. To better understand mortality, it may be important to look behind the proximate causes of death to the underlying drivers. However, because of the lack of direct pathways between and multiple contributions of drivers to causes of death, direct associations between causes of death and risk factors can be difficult to establish directly.

The only explicit model that directly and explicitly incorporates causes of death available on a public basis is that used in the calculation of long-term cash flows of the U.S. Social Security (Old Age, Survivors and Disability Insurance) system, based on seven causes of death (cardiovascular, cancer, violence and accidents, respiratory, dementia and other). Its 2019 Technical Panel on Methodology and Assumptions recommended that this method be restricted to the short- and medium-term future, as they believed that too much uncertainty and subjective decisions are required for the long term. In applying such a method over a lengthy period, great care is needed, because if the mortality improvement rate applied to one cause has a much lower improvement rate than the others, the improvement rate for that cause will tend to dominate the rates for the other causes.

Table 3 shows historical rates of mortality improvement for this projection for several U.S. age groups, sex and cause for 2009–2019 and for sex and cause for 1979–2019 for comparison purposes. This shows wide differences in mortality improvement by cause. It shows significant deterioration for violent and accidental deaths, as well as deterioration for dementia (the latter of which, in part, is due to an increasing tendency to report deaths due to that cause).

Table 3
ANNUAL RATE OF IMPROVEMENT BY CAUSE, AGE AND GENDER—UNITED STATES

	2009–2019						1979–2019
Males	0–14	15–49	50–64	65–84	85+	Total	Total
Cardiovascular	2.22%	0.75%	0.19%	1.31%	1.01%	0.99%	2.27%
Cancers	1.69	2.26	2.04	2.18	0.83	1.87	1.04
Violence and accidents	0.25	−2.44	−3.34	−1.78	−1.80	−2.40	0.12
Respiratory	2.04	1.92	−0.33	1.50	1.59	1.33	0.20
Dementia	3.93	0.50	−2.80	−1.89	−2.15	−2.06	−7.71
Other	1.75	0.32	−0.66	−0.63	0.06	−0.26	−0.20
Total	1.49%	−0.74%	0.03%	0.92%	0.38%	0.45%	1.02%
Females	0–14	15–49	50–64	65–84	85+	Total	Total
Cardiovascular	1.62%	0.42%	0.06%	1.69%	1.24%	1.26%	2.18%
Cancers	1.51	1.77	1.79	1.84	0.13	1.44	0.62
Violence and accident	−0.09	−2.35	−2.73	−1.58	−2.16	−2.16	−0.27
Respiratory	2.63	2.08	−0.75	1.07	0.51	0.69	1.54
Dementia	−1.78	1.84	−3.53	−2.39	−2.36	−2.38	−8.77
Other	1.67	−0.06	−0.84	0.15	0.75	0.19	−0.25
Total	1.42%	−0.21%	−0.02%	−0.93%	0.16%	0.45%	0.62%

Source: Social Security Administration, Office of the Chief Actuary, March 31, 2023. "The Long-Range Demographic Assumptions for the 2023 Trustees Report," Mortality, p. 20.

In assessing trends, care is also needed in interpreting the data used in cause-based methods because of attribution or survey response-related issues. For example, discontinuities can arise when ICD coding is updated, if not appropriately adjusted, and can adversely affect the analysis of long-term trends by cause.[18] Nevertheless, information gained from looking at prospects for causes and drivers of death is often used to influence the development of all-cause mortality improvement modeling.

Observed differences in mortality or mortality improvement rates for causes of death and risk factors can be used to help define risk groups, e.g., by age, gender, geography, marital status, racial, ethnic or national origin group, income or socioeconomic status, pre- and post-lapse status for certain life insurance products, and disability status. These differences can be significant for the development of refined models of relative mortality rates, for instance, when the mix of the population changes over time or an unrelated population is used as a historical baseline, especially for such purposes as analysis and determination of public policy, life insurance underwriting guidelines and actuarial decisions.

An example is marital status. It has been found that rates of mortality have differed significantly depending on an individual's marital status. There are many reasons for this, including self-selection. However, since in

[18] Inconsistent application or timing of introduction in many countries of versions of the International Classification of Diseases (ICD), which in 2022 uses version 10, can affect the results of time series analyses by cause. Since there are so many individual codes, they are almost always grouped.

some countries the family structure has been changing significantly over time, both the relativities and future trends may affect future mortality improvement rates. It is not yet clear the extent to which historical differentials resulting from changes in the family structure will apply in the future.

Knowledge of risk drivers is best applied if the determinants are well understood (Gutterman and Vanderhoof 1998), such as smoking and COVID-19 disruptions. Most problems involved are practical ones, relating to data and whether observed trends will continue, cycle, be exacerbated or fade away.

Prior to using cause and driver information, several additional issues should be addressed, including to what extent is all-cause mortality affected by a change in an individual driver or cause of death. A related question is to what extent would the elimination or dramatic change in the cause of death or risk driver reduce all-cause mortality. Naturally, the answer depends upon the driver or cause of death. An example might be the effects, safety and take-up rate of new antiobesity medications.

The author believes that it is highly likely that one or more other causes of death or risk drivers will in turn be affected, the direction of which will depend on the causes and drivers, their interactions and the circumstances involved. This is partly because some underlying conditions may affect multiple causes or drivers of death, the correlations among certain causes and risk factors, and the existence of multimorbidities, especially at older ages. Thus, for example, if a medical treatment or external risk factor changes, the effect on all-cause mortality may be different than just the direct change in the single identified factor.

Although reflection of correlation between causes and drivers in a model is possible and possibly desirable if proper support has been identified, it also may introduce additional complexity and require hard-to-quantify assumptions. Nevertheless, it may also enhance the goodness of fit to historical experience and more accurately reflect future expectations. However, there is no guarantee that the refinement(s) will enhance future accuracy and may not reduce uncertainty.

Causation is a desirable risk characteristic. However, it can be difficult to demonstrate. A high correlation between variables can be used where an association can be shown, although with caution. An enhanced understanding of the processes involved may provide an early warning signal regarding the beginning or continuation of a more significant relationship or a discontinuity in mortality.

The lag between an underlying driver and consequential mortality outcome, often quite long, can contribute to the difficulty of recognizing underlying trends. These lags differ significantly by the nature of the underlying condition or behavior, comorbidities involved and their consequences.

Once identified as significant, adverse mortality caused by a specific disease or condition might be expected to be addressed by large amounts of health care research and development financing and actions by governments, academic researchers and large pharmaceutical companies. However, the return period for this type of investment between financing and effective implementation in the population being assessed may be long and may prove difficult to achieve. Especially if the lags involved are long between exposure and recognition, between recognition and mitigation technique development, or between development and rollout to the entire population, the time until mortality is “turned around” may be quite lengthy.

Smoking is an example of an adverse health driver that took a really long time to obtain a consensus as to the seriousness of its adverse implications. Although obesity and dementia are other examples of adverse conditions that have been recognized for quite a while, nonsurgical “solutions” to them remain elusive, although breakthroughs may emerge.

Usually only the most significant risk drivers or causes are explicitly considered in the estimation and forecasting process. Although many of the following represent mortality risks, there are also corresponding favorable factors, including improved prevention programs, technologies and diagnostics and treatments. The following is a selection of demographic, behavioral, external and diseases that might be appropriate to consider, either explicitly or implicitly.

Demographic. As indicated earlier, the most common variables to recognize in this process have been age and gender. All useful estimates and projections consider age (measured by individual or quinquennial age). Most mortality models and mortality rate estimates also differentiate by gender, in large part because of the significant advantage females have in mortality in almost all populations, population segments and age groups. Other demographic factors that could be studied separately include genetic factors, race/ethnicity/national origin, marital status and occupation. However, some of these are problematic, as pointed out elsewhere in this paper.

For example, approaches that have been taken to reflect gender include the following:

- Study and forecast mortality rates for each gender independently;
- Assume the current relationship between their mortality continues, such as in terms of the number of years of life expectancy or mortality rates for large age bands;
- To reduce complexity, mortality improvement factors can be unisex, applied to gender-specific mortality rates, assuming that the mortality rate differentials will continue; and
- For national forecasts, female life expectancies are estimated and then life expectancies for males are developed based on the expected relation between the two sets of life expectancies, the approach used by the Population Division of the United Nations (2019).

Behavior and lifestyles. The prevalence and importance of behavioral and lifestyle factors in the historical or expected future populations differ by country, age, gender, socioeconomic and other groupings. Mortality rates often differ on a J-curve basis (e.g., regarding obesity, both being underweight and obese can represent adverse mortality factors relative to being of standard weight, with morbid obesity more adverse than underweight). A few of the most commonly used adverse behaviors are the following:

- *Smoking* combustible cigarettes (or other tobacco-related products) has been a key driver of all-cause premature mortality in many countries. In the U.S. it may have been the most significant factor driving adverse mortality experience over the last 75 years. As a result, in the U.S. a person's smoking habit or addiction has been at the heart of underwriting and pricing risk classification for individual life insurance since the 1970s. This behavior provides a prime example of the considerable lag between a behavior and consequential premature mortality, resulting in a long-wave mortality pattern. This time lag has resulted in a low incentive to cease smoking, as well as veiling other cause-and-effect relationships.
- The reduction in the prevalence of smoking was a key contributor to the rapid improvement of mortality in the 1990s and early 2000s. As this effect waned, it contributed to the slowdown in mortality improvement in the 2010s. Since the smaller U.S. female smoking prevalence peaked after that of males, its corresponding effects on female mortality occurred later and were not as pronounced. These patterns also contributed to the reduction in the difference between the mortality rates between the two genders over the last few decades.
- The increment to all-cause mortality is often modeled by assuming all smoking deaths are a multiple of lung, bronchia and trachea cancers (in some countries, 20% to 25% of all-cause mortality was due to smoking). The smoking mortality wave could be estimated by age and gender if several additional assumptions are made. Some mortality projections have ignored the effect of

the reduction in this source of premature mortality in developing projections of mortality improvements.

- *Obesity, diabetes and nutrition*. Obesity, especially long-term exposure of internal organs to white adipose tissues, has been increasing in the U.S. since at least the 1970s. It too has a significant lag between its prevalence and consequential adverse health conditions, averaging about 30 years. Consequential diabetes has become a major disease in both the U.S. and Mexico but is also becoming increasingly important in other parts of the world. Some actuaries believe that, with a significant lag, obesity, especially extreme and morbid levels of obesity, will become a significant driver of death in the years to come. Large food servings and unhealthy fast and processed food can lead to ill health. At the same time, a massive amount of food insecurity is found around the world, in some areas leading to undernutrition and malnutrition.
- *Physical inactivity, sedentary lifestyle*. As more jobs in many societies have moved from agriculture and industrial work to service and information economies, the amount of work-related physical activity has also diminished. Increased screen time worldwide, as well as other sedentary habits, exacerbated by COVID-19 in some areas has also led to increased ill health and adverse mortality outcomes.
- *Alcohol*. If managed ineffectively, alcohol use can be addictive, with significant adverse effects on the number of accidents and diseases, especially from binge and heavy drinking.
- *Drug overdose* and other external causes. It is difficult to assign drug overdoses solely to a lifestyle or behavioral issue, because factors such as addiction, disease, poverty, stress and other external influences can also drive these causes of death.

Although with the possible exception of smoking, mortality due to several behavioral risk drivers can be difficult to estimate and project, partly because of inadequate data reporting for some of these risk drivers. Nonetheless, their effects can be important to consider in both estimates and forecasts of mortality. Some of these lifestyle factors can affect mortality at most adult ages, with some having a greater effect on males because of their higher prevalence or exposure.

External causes. A wide range of external factors can have a material impact on mortality and mortality trajectories. Some are directly observable when they result in mortality spikes, while others represent continuing mortality risk factors. They include the following:

- *Infectious pathogens* can lead to epidemic or pandemic diseases, including viruses (e.g., underlying AIDS, COVID-19, influenza), bacteria (e.g., tuberculosis) and protozoa (e.g., malaria). Their trajectory can have a significant seasonal pattern, depending on the method of their human-to-human transmission. Particularly where reasonably well defined, these can be observed and thus modeled separately. Although future epidemics and pandemics are impossible to predict, we know that they will continue to emerge. Expanding antibiotic resistance may pose a threat to mortality in the future, particularly because of what is currently a limited antibiotic development pipeline.
- If a population is unfortunate to become involved in a *conflict* or *war*, there is a wide spectrum of effects, including direct deaths from the conflict and indirect deaths due to such factors as inadequate accessible health care services to diagnose and treat other ailments, other physical and mental effects, food insecurity and emigration-related issues.
- *Poverty, inequality and food insecurity* can exist in any country. Driven by certain demographic, socioeconomic or geographic characteristics, certain vulnerable population segments can be particularly exposed to material mortality risks. As a result, over the last several decades, mortality has been lower and mortality improvement in many countries has been greater for those with higher income and wealth or better access to health care resources. In fact, in some countries

such as the U.S., the mortality of those with lower socioeconomic status has been moribund, if not deteriorating. This is a major reason that the mortality improvement of life insurance and annuity policyholders has been better than the corresponding improvement for the overall population improvement.

- *Environmental natural disasters, climate change, climate risk and pollution*. All of these risk factors can have both immediate and long-term effects resulting from exposures to a wide range of these external drivers. They may take the form of a spike or wave, or represent a permanent effect, depending on their nature and prognosis. Climate change has been a growing concern over the past several decades. Depending on the effectiveness of applicable mitigation and adaptation measures that have been and will be taken, this factor will represent a significant mortality risk factor, especially among those most vulnerable, although its significance will differ significantly by location and exposure to climate-related risks. For example, excessive wildfires constitute both a risk to those in the vicinity and an air quality risk to those farther away, such as from the wildfires in Canada in the summer of 2023.
- *Migration*. Immigration or emigration can result from one or several other external causes, such as conflicts, poverty and environmental crises, or simply economic hardship. Immigrants may contribute to a spike, wave or permanent effect, depending on their risk characteristic profile. In many cases the mortality for immigrants may be better than that for the native born (partly because of self-selection). This has been referred to as the "healthy immigrant effect" (Fernandez, García-Peréz and Orozco-Aleman 2023).
- *Firearms*. Although the frequency of these deaths differs dramatically by country, geographic area and age group, firearms can contribute to a large number of premature deaths due to accidents, murders and homicides or suicides.

Noncommunicable and infectious diseases. Medical developments can offset the effect of some of these, such as the following:

- *Cardiovascular*. Overall, cardiovascular diseases represent the major cause of death in many countries at middle or older adult ages. In higher-income countries, a key source of mortality improvement has been a result of improved diagnosis and treatment of these diseases or their risk factors including cholesterol or hypertension. However, it is possible that the significant favorable effects in the past may diminish over time, as new developments, diagnostics and treatments may not be as significant as those over the last several decades. In addition, it is not clear that the use of these key risk factor mitigation actions will continue to increase, as cost-related nonadherence has loomed large, estimated by Van Alsten and Harris (2020) to result in about 125,000 deaths per year in the U.S. Thus both prevention and treatment are important.
- *Cancers.* Although progress against some types of malignant neoplasms (cancers) has been made, this disease category is made more difficult because it consists of many different diseases, with a range of causes. Early diagnosis will continue to be important for significant future progress in this area.
- *New diseases or outbreaks*. The unknown unknowns of mortality drivers.
- *Mitigation*.
    - Public health actions and their communication focus on fighting the expansion of and eradication of diseases and accidents, as well as their other consequences. These include enhanced sanitation, safer water and food, more nutritious food and drink, new medical technologies and medications, as well as changing lifestyle habits. However, these mitigation actions may not sufficiently offset the adverse drivers and causes involved and may result in diminishing returns (e.g., after a peak in the percentage taking statins was reached in the U.S., their use has recently diminished).

- Special attention has recently been given to new developments, such as mRNA-based vaccines for infectious diseases, antiobesity medications, CRISPR-Cas9, cell therapies and the use of increased knowledge and manipulation of the human genome.

As mortality has improved, more people are living to older ages. As a result, increased attention is being given to the aging process and its management. However, the increase in the number of multimorbidities at older ages may increase exposure to other drivers of premature death, especially mental and neurological illnesses such as dementia and Parkinson's, especially at older ages.

Some of the above drivers and causes may be sufficiently identifiable and attributable to consider and possibly adjust for if sufficient evidence exists to reflect changes in their prevalence or severity.

An aggregation of risk factors, expressed as an index of deprivation or fragility, can be used to aggregate or summarize a set of related risk characteristics, such as socioeconomic or health conditions. Such an index could be used to assess historical drivers for a population and to help forecast future trends, especially if one or a few of its components have experienced or are expected to experience significant change. Another approach uses or develops a composite factor, such as self-assessed health, which, although often highly related to future mortality, may also provide a useful early warning indication for use in developing short-term mortality improvement forecasts. However, although possibly useful as a descriptive measure, without rigorous testing, such indices may not be as useful for predictive purposes.

11. Uncertainty

Uncertainty is involved, not only in every step in the process of projecting mortality rates, but also in developing estimates of the past. Even a "good" projection may inherently have a large amount of uncertainty associated with it. Although this paper and thousands of other papers deal with the topic of projecting mortality, many authors and researchers unsurprisingly have focused on their best estimates based on a deterministic or single future, rather than emphasizing their stochastic variability. Nevertheless, it is important to recognize the relevant and significant complexity and multiplicity of uncertainties involved in human health and mortality processes. Analysts and students of these processes have to mix knowledge with humility, recognizing the uncertainties involved.

Many sources of uncertainty relate to the complex and ever-changing nature of individual and group aspects of human health and forces that can affect health. They include inadequate and incomplete data and sampling, mortality disruptions, model misspecification, informed and uninformed judgment, misestimation risk (that is, the less that the starting point is understood, the consequential projection may be inaccurate to a greater extent), misunderstood systematic bias, and unrecognized correlations and competing risks. This is in addition to difficulties in projecting human behavior, political actions and future changes in various technologies.

In actuarial literature, the categorization of uncertainty has often been addressed as being composed of three types: process risk (normal fluctuations of random elements and expected changes), parameter risk (with respect to the ultimate relative importance of explicitly identified factors) and model risk (misunderstanding or lack of recognition of the effect of drivers of change in overall mortality). Drivers of uncertainty can consist of currently known or unknown favorable factors (e.g., a cure for Alzheimer's or effective and affordability antiobesity medications) and adverse ones (e.g., widespread antibacterial resistance and future pandemics). In some cases risk refers to quantifiable deviations from that what is expected, while uncertainty refers to nonquantifiable deviations. To the extent practical, a range of possible outcomes should be considered.

Although a set of confidence intervals are sometimes developed for a set of expected mortality rates (typically a symmetric range around those expected values), these typically reflect only statistical (process) risk and not parameter or model risk, which is a higher likelihood of an outlier outcome. To reflect these latter types of risk, a wider range may better reflect the total risk and uncertainty involved. One approach taken to reflect a wider range is to use a greater level of statistical confidence (e.g., a 99.5% confidence rather than a 95% confidence level), a regime-switching approach or a more extreme individual or multiple scenarios.

A new spike or wave may unexpectedly develop (e.g., a new global pandemic, a nuclear accident or increased air pollution). The frequency, severity and timing of new spikes or waves are difficult to anticipate. However, although a specific event or condition may be unpredictable, it can be anticipated that such an event or condition may arise and that appropriate planning should be conducted, e.g., through the accumulation of capital to provide for the sustainability of the entity.

Certainly, the size and makeup of the population studied affects the extent of its mortality uncertainty—in some cases it can be as difficult to determine the mortality profile of a small population as it can be to interpret mortality outcomes because of the multitude of factors involved. This is one reason national statistics are followed so closely, as their size can reduce the amount of process risk, as well as overall expected volatility.

Parameter risk can increase substantially as a result of a pervasive or multifaceted mortality disruption. That is, in the middle of a multiyear pandemic, it is difficult to determine the underlying trajectory of “normal” mortality and mortality risk factor processes, which can increase the level of uncertainty regarding where current and future mortality will be headed. In some cases such disruptions may affect only the short-term and to a lesser extent the intermediate-term future rates, and not at all those of the distant future.

Of course, the range of probable values developed assumes that the best estimate (or statistical mean) is accurate, which naturally puts a great deal of pressure on the development of an accurate estimate of the mean. The width of the confidence interval for the rate of mortality improvement can mathematically be expressed as being inversely proportional to the number of years in the period analyzed and inversely proportional to the square root of the cohort size.

It is important to distinguish between short-term and long-term mortality uncertainty. Although short-term uncertainty can reflect the effect of a specific type of disruption and volatility (process risk), long-term uncertainty is more often expressed in terms of a relatively stable difference from the baseline projection. Its true uncertainty (due to process or model risks that can accumulate over time) may be far larger.

Regardless of the source and extent of uncertainty, the method of expressing uncertainty can depend on the experience of the forecaster and the forecaster’s profession, the approach used to develop an estimate of the expected value, and the users of the forecast. It can be especially large at the oldest ages where data regarding the cause of death or even population exposure may be problematic.

The best estimate of the rates of improvement often falls in the middle of the range of uncertainty, that is, in a symmetric range. However, opinion can be divided as to whether the effects of the favorable or adverse factors will dominate. If an asymmetric range is selected, the analyst needs to confirm whether the best estimate has been chosen appropriately. Since tail outcomes can occur in either direction, whether a probabilistic or scenario-based method is used, process, parameter and model risks can be favorable or adverse compared to the usually expected gradual favorable trend.

Uncertainty is often expressed utilizing a mathematical model, judgment or a combination. Its quantification can be developed through the choice and characteristics of the data set used, the probability distributions or risk assessment and tolerance applied, and the selected confidence interval (that is, the range of probable values, possibly expressed by a given number of standard deviations) chosen. Ultimately, the approach taken and the type of quantification may be expressed in a qualitative form depending on the analyst, user and use.

Although overall long-term trends in mortality and longevity have historically been overwhelmingly favorable for quite some time in most areas, this may not always be the case for a given population segment or a wider population. Experts and others (e.g., using the Lee-Carter based on historical trends, extrapolation and even scenario analysis based on expert opinions) often base their long-term forecasts on relatively recent experience ("current bias") under the human assumption that if there has been considerable improvement in the relatively recent past or over the long term, this level of improvement will continue. Thus, if anything, mortality estimates tend to be optimistic or rates of mortality improvement may be overstated. If recent mortality drivers have been unfavorable, unfavorable mortality improvement may be expected to continue.

Uncertainty can be expressed in terms of the level of mortality or of its rate of improvement, although the former is more common. Estimating a rate of mortality improvement across time can be more challenging than estimating the mortality rate for a particular year or period. To estimate an improvement rate, a comparison of the annual rate of change in mortality rates across time can be the starting point. Each annual mortality rate estimate should be rigorously reviewed to produce robust improvement rate estimates in relation both to history and to one's expectations (e.g., commonly between 0.75% and 1.5% annually, possibly lower at older ages) that are not overly influenced by unexplained outliers.

Stochastic approach
Mortality as a stochastic process, corresponding to the risk that aggregate mortality might differ from that expected, is a fundamental aspect of the mortality trajectory. Uncertainty is quantified according to the mathematical model applied to develop its mean value; alternatively, the model can be developed given the mean value, to obtain insight into the type, form and extent of the uncertainty involved.

To develop a stochastic model, one (or more) probability distributions are needed. The distributions can represent the likely values and trajectory of rates of mortality or the rates of mortality improvement. Although the basis of these distributions needs to be established, overreliance on the statistical volatility observed from, say, the last half a century or more or solely on process risk may be problematic in a future world full of possible drivers and disruptions that may be quite different from those from the past and does not consist of a single set of random elements. Indeed, output from such a model will be only as reliable as the probability distributions used.

Because of both parameter and model risk, there are also possible tail events, that is, disruptions that may not behave solely in a solely probabilistic manner. These can take the form of a shock or wave, or represent a regime change or discontinuity in rates or rate improvement. This process might be able to be captured by a wide probabilistic tail, with probable (or at least possible) mortality disruptions of unknown frequency, severity and duration that could dominate what may be purely statistical volatility. In examining or forecasting a tail event, it has to be remembered that in a tail situation, the correlation among mortality drivers and causes can become highly correlated and might cascade either in an adverse (or favorable) direction that can exacerbate the effect of the driver or cause.

The objective of the forecast will guide the acceptability of limits of uncertainty quantified by the stochastic model, expressed in terms consistent with the objective. Even if a probabilistic model is used (e.g., a

Weibull or Poisson distribution), the confidence level needs to be set. For example, if the projection is used for an insurance regulatory purpose, a preset confidence level would usually be applied, set based on its soundness or sustainability objectives. If used for another purpose, the level may be subjectively set based on the risk tolerance of the user.

In any event judgment is by necessity involved throughout. In real life the confidence interval used can be quite difficult to derive in a scientific manner, because it is somewhat arbitrary and is ultimately based on subjective judgment developed for other purposes.

The probability distributions used are usually fitted to a selected set of historical mortality data. Care is needed to ensure that the data are reliable and the period used is reasonably relevant to the period for which the forecasts are to be applied. The question of whether to include outlier or tail data, like that associated with COVID-19 (and the two world wars of the twentieth century and the influenza epidemic of 1918–1919), needs to be determined. If practical, both directions of significant shocks and waves should be considered. The extent that parameter risk, model risk and the desired degree of asymmetry are not fully represented in the historical experience used needs to be recognized.

Even if a seemingly refined stochastic model has been developed, there is still a significant role for expert opinion. This opinion might address the choice of the confidence level of acceptable risk, and how much to incorporate or weight parameter and model risk.

Examples of stochastic models are Lee-Carter (Lee and Carter 1992 and its variants) and Cairns-Blake-Dowd (Cairns et al. 2006).

Representative scenario approach

A scenario represents a description of how the future may develop, based on a coherent and internally consistent set of assumptions about key driving forces (e.g., reflecting a given rate of technological advancement and disease trajectories) and their relationships, as well as their ramifications. Scenarios are neither predictions nor forecasts but are used to provide narratives and trajectories that result in one or more likely multiple alternate outcomes. They are usually comprehensive, that is, all assumptions used in a scenario are consistent with a similar set of conditions, often based on a story.

A scenario analysis includes both the selection of the story from which the scenario is developed and the assumptions used in each scenario. This can take the form of three scenarios, including a low-end, midrange and high-end projection. A scenario can also include a temporal component (short-term, medium-term and long-term). The scenarios selected are reasonable ones, not being best or worst cases, although that will depend on their use. Their description needs to describe the story and how it should be interpreted.

Once the best estimate (or best-estimate scenario) has been developed, the study of deviations can be meaningfully illustrated with alternative scenarios and sensitivity analyses. These can be expressed either quantitatively or qualitatively to illustrate the effect of specified deviations from the best estimate. Developing useful scenarios can be a challenging but worthwhile task. First, just thinking through alternative assumptions can force the analyst to consider all sources of uncertainty. One approach is to use a common theme, such as the continuation of the level of mortality trends of the 2000s or the lower trends of the 2010s.

A static scenario is one for which the analyst selects a set of conditions that are expected to apply, with a smooth trajectory. A dynamic scenario employs either a stochastic model with a stochastic process assumed or a set of representative alternate but realistic scenarios; both approaches embed a recognition of the uncertainties involved. The following discusses how these approaches have been used to recognize, quantify and illustrate the relationships between the mortality process over time and space and its corresponding uncertainties.

Scenarios can focus on rates of improvement or mortality, possibly measured in terms of life expectancy at certain ages in periods 3 and 5, or the speed and trajectories of transition in period 4.

It can represent a numerical sensitivity test or be based on a holistic scenario (e.g., a high or low level of medical or behavioral advancement, probably differing by broad age group). In some cases relatively simple alternative sets of assumptions (through deviations from expected) can provide useful information. Sensitivity testing, with a range of alternative sets of improvement factors (e.g., 0.5% or half of the rates of annual improvement in each direction), can demonstrate how sensitive the application's outcome is to these alternative assumptions.

Differences between components of a holistic alternative and baseline scenario represent the uncertainty associated with level (lack of understanding of the population or future structural change), improvement (trend) or shock (temporary).

Sensitivity tests can address the likelihood of deviations in all related assumptions applicable, for example, to a pension plan or an insurance company, not just the mortality assumption. When a significant spike occurs (e.g., a global pandemic), it is likely that the general economy and other demographic assumptions will also be significantly affected.

A risk margin?
In some applications a margin for uncertainty is needed. To determine an appropriate size of a mortality margin needed for safety or sustainability in the face of uncertainty, the basis and extent of comfort needed, that is, the extent of risk aversion inherent in the application or by the users of the outputs of the analysis, must be assessed. Usually this margin is based on the risk preference of the user, such as how many standard deviations or the most reasonable adverse scenario (e.g., a deviation of 10% of mortality levels or half the expected rate of improvement) the user wants protection against. For life insurance a safety margin would increase mortality, whereas for annuities a safety margin would decrease mortality. Because of their sensitivity to these factors and the many risk factors involved, tail situations are often best illustrated by representative scenarios.

Preferably any such margin should be expressed explicitly, rather than implicitly. If implicit, the actuary should identify how it was incorporated and its level of deviation. In any case, where practical, the key assumptions underlying the uncertainty calculations need to be communicated in a manner that is clear to the users and bearers of risk, in terms of explicit assumptions or implicit considerations. This is important for both future updating and proper communication.

The basis and extent of "comfort" needed and the risk aversion desired by the user or bearer of risk, if applicable, will drive the extent of such a margin. Usually, this margin is expressed in terms of how many standard deviations or the level of confidence is used as a threshold for acceptance or action, even if the distribution does not include a full provision for parameter, model or tail risk.

12. Validation and Governance

Using a dynamic view of the processes involved in projecting mortality improvement represents best practice. One necessary component of this process is a periodic (re)assessment of how accurate the most recent mortality forecast has and is anticipated to perform, reflecting all up-to-date and relevant information. The evaluation of a model or a set of mortality projections is dependent on the objectives of its application and the degree of accuracy sought.

The periodicity of reviews can differ by application and users, or when a sufficient deviation from the expected arises. However, there is also a danger of being overly responsive to short-term fluctuations, as some volatility is natural and should be expected. That said, when to act or to begin further analysis can be subjective: it can be useful in advance to set a threshold or timeframe beyond which further steps may be warranted. Sometimes an annual review is performed, but assumption changes will not be made more frequently than every three years.

Actuaries can develop overly complex models, possibly including being overfitted, using variables that are not significant and difficult to understand and thus to justify, communicate and update in the future. They also can lead to overconfidence in their projection accuracy. It has to be remembered that the purpose of developing estimates of mortality improvement is to develop sufficiently reliable and unbiased projections relative to the application, extent of perceived uncertainty involved and available resources.

Since a forecast can cover a lengthy period, monitoring its accuracy can be somewhat problematic. Although measuring short-term actual-to-expected comparisons can be of interest (and in, e.g., the case of a life insurance or annuity company, it can identify or help explain a source of profit or loss), the results may not be able to provide an accurate indication of an overall projection. Periodic experience analysis can, however, provide insight into trends and comparisons with other entities or countries, depending on the application. It also might be able, through negative assurance regarding deviations from that expected, to provide an early indication that the underlying assumptions may have to be revisited.

To identify at what point a forecast should be reviewed, it may be useful to represent expected performance by a probability distribution. Alternatively, a reasonable range of plausible trajectories could be applied on a heuristic or judgmental basis. Although simply extrapolating recent trend lines might be used, because of many of the factors discussed earlier, this may not represent the best approach to follow, unless concerned only with the short term. Consistent with past practice, preconceived notions of improvement consistent with past trends (e.g., with regard to the experience of past decades of national experience) and their average rate of improvement might serve as useful bases for alternative mortality scenarios.

Nevertheless, it remains important to validate the mortality estimates and projections in some fashion. Commonly used metrics used for this purpose include actual-to-expected ratios and mean-squared errors (focusing on deviations in both directions) in the aggregate or for major drivers or causes of death components. Although commonly applied to expected mortality rates, the examination of improvement rates may also be useful.

Validation of data accuracy and overall reasonableness testing against alternative methods or trends are all important steps to consider taking in some form. Careful consideration of significant changes in circumstances or population mix, as well as the effect of any mortality disruptions, can be part of the validation process. As easy as it is to be skeptical of the interpretation of life expectancies, changes and comparisons of period life expectancies (or age-standardized rate indices) over time can serve as useful harbingers of needed actions and areas for further study.

Although expert opinion is to be sought, experts need to be properly vetted. In addition, since this is not an exact science, expert opinions can be and usually are diverse. Bias can emerge due to a variety of reasons. In many cases the logic or reasoning used can be as useful as numerical forecasts. A reasonableness review (sometimes referred to as a “smell test”) is often performed.

Several projection models assume that forecast noise (i.e., the error or residual term) follows a normal distribution, with a zero bias. This assumes that the effects of future disruptions offset each other, that is, there is an equal impact of positive and negative deviations from that projected. The purpose of the validation process is to identify any non-offset bias. Depending on the application and use of the mortality projections, multiple models might be used to help the validation process.

Regarding what to do about validating long-term assumptions, one approach might be to compare the forecast against recent mortality experience of other countries or similar population segments that might serve as an early indicator of future mortality of the population being assessed that are farther along on the mortality improvement curve. However, the analyst needs to be aware of any significant differences in the risk profile and cultural or behavioral differences between the populations being compared. An example is the prevalence of immigrant groups, which in many cases represent a more favorable mortality profile than those who are native born. This can be because of a higher likelihood of self-selection, that is, those who immigrate to a country (or new job entrants) tend to self-select (e.g., in Canada, immigrants need to pass a physical examination). International (or intergroup) comparisons may suggest mortality improvement capacity.

At each step in the improvement analysis process, a review for reasonableness avoids having to start from scratch once a round of the process has been completed. This includes a search for the identification of unexpected or unintended discontinuities in mortality or mortality improvement rates. Visual inspection (e.g., through heat maps) or automatic means (checking whether a rate of improvement is less than or greater than a prespecified range in rates) can be used.

Caps and floors may be appropriate; for instance, in the extreme, mortality rates cannot be negative or greater than 100%. Another is that, in the long term, mortality rates cannot go below a reasonable rate for accidents, which is possible for youths if an overly optimistic continuation of past favorable trends is relied upon. Although rates may not necessarily be revised as a result, at least reasons for the results will have been identified beforehand.

Governance or oversight of the process followed and decisions made regarding methodology and assumptions are integral components of an analytical or forecasting process. Depending upon the application, the methodology and assumptions may have to be formally approved or reviewed and be acceptable by an independent party, such as through a peer review process. Good governance requires reporting, explaining and justifying the findings and, if applicable, any recommendations. Some form of oversight or peer review is always desirable, because the process and factors involved are numerous, although the depth and intensity can differ depending on the situation. In any case the methodology, data sources and basis for assumptions should be documented in a form that a reviewer(s), a user(s) and a future analyst can understand and, where appropriate, reproduce.

References

Barbieri, Magali. 2022. Mortality by Socioeconomic Category in the United States. February. Schaumburg, IL: Society of Actuaries.

Booth, Heather, and Leoni Tickle. 2008. Mortality Modeling and Forecasting: A Review of Methods. *Annals of Actuarial Science* 3(1–2): 3–43.

Cairns, Andrew J. G., David Blake, and Kevin Dowd. 2006. Pricing Death: Frameworks for the Valuation and Securitization of Mortality Risk. *ASTIN Bulletin* 36(1): 79–120.

Continuous Mortality Investigation (CMI). 2009. CMI Mortality Projections Model Feedback on Consultation and Issue of 'CMI_2009.' Working Paper 41. London: Institute and Faculty of Actuaries.

Cox, David. 1972. Regression Models and Life-Tables (with Discussion). *Journal of the Royal Statistical Society Series B* 34: 187–220.

Currie, Iain D., Maria Durban, and Paul H. C. Eilers. 2004. Smoothing and Forecasting Mortality Rates. *Statistical Modeling* 4(4): 279–298. doi:10.1191/1471082X04st080oa.

Deprez, Philippe, Pavel V. Shevchenko, and Mario V. Wüthrich. 2017. Machine Learning Techniques for Mortality Modeling. *European Actuarial Journal* 7: 337–352.

Fernandez, José, Ménica García-Peréz, and Sandra Orozco-Aleman. 2023. Unraveling the Hispanic Health Paradox. *Journal of Economic Perspectives* 37(1): 145–168.

Gavrilov, Leonid A. and Natalia S. Gavrilova. 2002. Evolutionary Theories of Aging and Longevity. *Scientific World Journal* 2: 339–356. doi:10.1100/tsw.2002.96.

Gompertz, Benjamin. 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. doi.org/10.1098%2Frst1.1825.0026.

Gutterman, Sam, and Irwin T. Vanderhoof. 1998. Forecasting Changes in Mortality: A Search for a Law of Causes and Effects. *North American Actuarial Journal* 2(4): 135–138.

Heligman, Larry, and John Pollard. 1980. The Age Pattern of Mortality. *Journal of the Institute of Actuaries* 107(1): 49–80.

Jenkins, Willard A., and Edward A. Lew. 1949. A New Mortality Basis for Annuities. *Transactions of the Society* of Actuaries 1: 369–466.

Kannisto, Väinö. 1992. Presentation at a workshop on "Old-Age Mortality", Odense University, Odense, Denmark. June 1992. Described in A. Roger Thatcher, Väinö Kannisto and James W. Vaupel 1998. The Force of Mortality at Ages 80 to 120. Monographs on Population Aging, 5. Odense University Press.

Lee, Ronald, and Shripad Tuljapurkar. 1998. Population Forecasting for Fiscal Planning: Issues and Innovations. Unpublished manuscript.

Lee, Ronald D., and Lawrence R. Carter. 1992. Modeling and Forecasting U.S. Mortality. *Journal of the American Statistical Association* 87(419): 659–671.

Ludkovski, Mike, Jimmy Risk, and Howard Zail. 2018. Gaussian Process Models for Mortality Rates and Improvement Factors. *ASTIN Bulletin* 48(3): 1307–1347.

Makeham, William M. 1860. On the Law of Mortality and the Construction of Annuity Tables. *Assurance Magazine and Journal of the Institute of Actuaries* 8(6): 301–310. doi:10.1017/S204616580000126X.

Perks, W. 1932. On Some Experiments on the Graduation of Mortality Statistics. *Journal of the Institute of Actuaries* 63: 12–40.

Population Division of the United Nations. 2019. *World Population 2019: Methodology of the United Nations Population Estimates and Projections.* New York: United Nations Department of Economic and Social Affairs Population Division.

Society of Actuaries Research Institute. 2021. Developing a Consistent Framework for Mortality Improvement, MIM-2021-v2. November. Schaumburg, IL: Society of Actuaries Research Institute.

Van Alsten, Sarah C. and Jenine K. Harris 2020. "Cost-Related Nonadherence and Mortality in Patients With Chronic Disease: A Multiyear Investigation, National Health Interview Survey, 2000–2014". *Preventing Chronic Disease* 2020;17:200244. doi: http://dx.doi.org/10.5888/pcd17.200244.

Yang, Yang. 2008. Trends in U.S. Adult Chronic Disease Mortality, 1960–1999: Age, Period, and Cohort Variations. *Demography* 45(2): 387–416.

About The Society of Actuaries Research Institute

Serving as the research arm of the Society of Actuaries (SOA), the SOA Research Institute provides objective, data-driven research bringing together tried and true practices and future-focused approaches to address societal challenges and your business needs. The Institute provides trusted knowledge, extensive experience and new technologies to help effectively identify, predict and manage risks.

Representing the thousands of actuaries who help conduct critical research, the SOA Research Institute provides clarity and solutions on risks and societal challenges. The Institute connects actuaries, academics, employers, the insurance industry, regulators, research partners, foundations and research institutions, sponsors and non-governmental organizations, building an effective network which provides support, knowledge and expertise regarding the management of risk to benefit the industry and the public.

Managed by experienced actuaries and research experts from a broad range of industries, the SOA Research Institute creates, funds, develops and distributes research to elevate actuaries as leaders in measuring and managing risk. These efforts include studies, essay collections, webcasts, research papers, survey reports, and original research on topics impacting society.

Harnessing its peer-reviewed research, leading-edge technologies, new data tools and innovative practices, the Institute seeks to understand the underlying causes of risk and the possible outcomes. The Institute develops objective research spanning a variety of topics with its strategic research programs: aging and retirement; actuarial innovation and technology; mortality and longevity; diversity, equity and inclusion; health care cost trends; and catastrophe and climate risk. The Institute has a large volume of topical research available, including an expanding collection of international and market-specific research, experience studies, models and timely research.

Society of Actuaries Research Institute
475 N. Martingale Road, Suite 600
Schaumburg, Illinois 60173
www.SOA.org