Scanning the horizon: integrating expert knowledge into the calibration of stochastic mortality models

Richard G. A. Faragher*, Arne Freimann† [corresponding author], Jochen Ruß‡

This version: 17th of December, 2024

Abstract

Expert knowledge from many different disciplines has the potential to inform on developments that could significantly increase or decrease human life expectancy. However, such knowledge is typically not considered in longevity risk management, since stochastic mortality models are generally only calibrated to historical mortality patterns, i.e., fully data-driven.

Following an interdisciplinary approach, we develop a methodology how expert knowledge on the (uncertainty of the) future of human life expectancy can be integrated into the calibration of stochastic mortality models. We argue that this approach is particularly relevant if there are "low probability / high impact" scenarios on the horizon, that are considered plausible by experts in their respective field but are "virtually impossible" in models calibrated to historical data. Based on current research on treatments that might be effective in slowing down ageing, we motivate and propose an exemplary plausible scenario for the future development of human life expectancy. We assign a potential impact on life expectancy as well as a plausible probability of occurrence to the scenario and present a method for calibrating stochastic mortality models so that the resulting projections are in line with these parameters. In a case study, we analyse and compare the longevity risk in an exemplary annuity portfolio and show that this so-called "driver-driven" calibration can lead to a structurally different assessment of longevity risk than the traditional "data-driven" approach, especially with regard to tail risks.

Keywords: Longevity risk, mortality trends, stochastic mortality models
JEL-codes: G17, G22
Preliminary version has been presented at the Longevity 19 conference in Amsterdam.

* School of Pharmacy & Biomolecular Sciences, University of Brighton, Brighton, East Sussex, BN2 4GJ, UK (R.G.A.Faragher@brighton.ac.uk).

† Institute for Finance and Actuarial Sciences (ifa), Lise-Meitner-Str. 14, 89081 Ulm, Germany (a.freimann@ifa-ulm.de); corresponding author.

‡ Institute for Finance and Actuarial Sciences (ifa), Lise-Meitner-Str. 14, 89081 Ulm, Germany & Institute of Insurance Science, University of Ulm, Helmholtzstr. 20, 89069 Ulm, Germany (j.russ@ifa-ulm.de).

1 Introduction

Although human life expectancy has shown relatively steady past increases, particularly in the 20th century (cf. Oeppen and Vaupel, 2002), uncertainty about future developments currently appears rather high which is, e.g., illustrated by a strongly shifting view on future mortality by the British House of Lords Science and Technology Committee which, back in 2005, was optimistic regarding trends in population life expectancy stating that *"for each hour spent in reading this report, life expectancy will have increased by 12 minutes.*[1] However, by 2019, the same committee reported in a bleaker tone that *"the rate of improvements [of life expectancy] decreased […]. Reasons for this slowdown are under investigation…*[2] In line with this shifting view, in 2015, BBC News posed the question *"Life expectancy: Is the party over?"*.[3] They argued that *"increasing obesity or antibiotic resistance might slow or even halt the trend to improving life expectancy"* and extrinsic factors as varied as air pollution, micro-plastics in food and climate change have been postulated as significantly decreasing life expectancy with, it should be noted, equally varied evidence bases (cf. Patkee and Strange, 2023). On the other hand, population level improvements in lifestyle (cf. Khaw et al., 2008) or medical breakthroughs (particularly in the field of slowing down the human ageing process – see below) hold the potential to significantly decrease mortality and hence increase life expectancy.

Knowledge about potential scenarios that might impact human mortality exists in many different disciplines along with the potential for subject specialists to provide educated estimates for the probability of both occurrence and potential impact. This well-founded knowledge would be particularly useful whenever mortality risk or longevity risk is being quantified, e.g., in risk management of life insurers or pension funds, or in the calculation of solvency capital in risk-based frameworks like Solvency II in the European Union. However, it appears that such knowledge is hardly considered in mortality/longevity risk management. Common stochastic mortality models such as the widely used models of Lee and Carter (1992) or Cairns et al. (2006) are typically calibrated in a purely "data-driven" way. This means that model parameters are calibrated to historical mortality data so that the model captures observed past mortality patterns and dynamics as closely as possible, cf. e.g., Currie

[1] House of Lords Science and Technology Committee (2005) *Ageing: Scientific Aspects.* 1st Report of Session 2005–06.

[2] House of Lords Science and Technology Select Committee (2020) *Ageing: Science, Technology and Healthy Living*. 1st Report of Session 2019–21.

[3] https://www.bbc.com/news/business-35070410

(2016), and extrapolates them stochastically into the future. This is a suitable approach whenever there is no indication that the structure of possible (albeit uncertain) future fluctuations of mortality rates might be different than a structure one would estimate from looking at historic data. However, at some points in time, there might be a situation where experts in their respective field (e.g., medical experts considering a potential future drug on the horizon or biogerontologists considering a potential breakthrough with respect to slowing down human ageing) might come to the conclusion that there is a significant probability for a rather high future increase in life expectancy, whilst at the same time such an increase would be considered "virtually impossible" by a model that is solely calibrated to historical data.

To illustrate this point, imagine the year is 1929, and Alexander Fleming has recently published a paper suggesting that "*penicillin in regard to infections appears to have some advantages over chemical antiseptics. It is a more powerful inhibitory agent…and it can be applied to an infected surface undiluted as it is non-irritant and non-toxic*" (Fleming, 1929). Contemporary expert judgement would probably have concluded (and indeed did conclude) that there was a significant probability that penicillin can be purified and produced at sufficient scale to have a high impact on human life expectancy. Although nobody in 1929 could have exactly predicted the full impact or correctly specified the probability and timing of the events, any "educated guess" by experts in the field would have provided a better understanding of the uncertainty of (then) future development of human life expectancy than a purely data-driven approach.

In view of current developments in the field of anti-ageing research, we argue that we are currently at a similar point in time where uncertainties regarding the future development of human life expectancy might be larger than a purely data-driven approach suggests. Hence actuaries / risk managers should consider all available information when calibrating stochastic mortality models, with a particular emphasis on the inclusion of expert knowledge from other areas.

Although uncertainty is high "in both directions", i.e., with respect to a potential decrease as well as increase of life expectancy, we focus on a potential increase in life expectancy since the topic of this paper is on modelling, measuring, and managing longevity (as opposed to mortality) risk. But our proposed methodology can of course also be applied to incorporate expert scenarios that lead to an increase in mortality, i.e., a decrease of life expectancy.

Also, since the focus of our paper is on risk (i.e., the potential deviation from a best estimate), we concentrate on modelling "low probability / high impact" scenarios. We would like to mention, however, that expert opinions could of course also be relevant for creating a best estimate scenario. This would particularly be relevant if a certain scenario comes with a

rather high probability or if, for whatever reason, past data is not considered representative of the future best estimate trend. For instance, the Institute and Faculty of Actuaries (2023) suggests using a driver-driven calibration approach based on forward-looking expert judgement for the task of deriving best estimate mortality assumptions in a post-pandemic era.

The use of expert opinions for the calibration of stochastic risk models is not as uncommon as it might seem at first glance. Often, this is done implicitly. For instance, if – for whatever purpose – an estimate for future interest rates is required, this is typically derived from forward rates based on market prices of bonds or swaps. So, estimates for future interest rates are typically not derived by an extrapolation of historical interest rate data but rather by prices that depend on expectations of market participants which constitutes some kind of expert opinion. The same holds true whenever an implied volatility of stocks is used as a risk measure. This is not derived from looking at historical fluctuations but also rather from market participants' opinion on future uncertainty. Since a deep and liquid financial market for longevity-linked securities that would allow to derive implied volatilities from market prices does not (yet) exist, we need a different way to consider expert opinions in risk analyses. One possible methodology to achieve this will be presented in this paper.

The remainder of this paper is structured as follows: In Section 2, we derive an exemplary plausible scenario for the future development of human life expectancy by an interdisciplinary approach. We would like to stress that (although we make sure that our considered scenario is plausible and considered realistic by experts in the respective field), our intention is not to claim that this is "the correct" or "the most relevant" scenario that stochastic mortality models should be calibrated to. Rather, the aim of our paper is to propose a methodology that can be applied for calibration once a modeller has decided that some concrete scenario should be used. Such a calibration methodology is introduced in Section 3. We explain how a purely data-driven calibration of a stochastic mortality model can be modified by adjusting the model parameters such that the resulting model is consistent with a given expert scenario. In Section 4, we apply this methodology using the scenario from Section 2 and analyse the impact of the modified calibration on an insurer's risk assessment in a case study. Finally, Section 5 concludes with a brief discussion of potential future extensions of our methodology.

2 A plausible "low probability / high impact"-scenario

In this section, we present a concrete expert scenario about how human life expectancy could develop in the future. A scenario consists of three components: **What** might happen? **When** could it happen? **How likely** is it to occur?

2.1 Why consider the field of "anti-ageing medicine"?

Many potential scenarios could impact human mortality but events with a low impact are probably not relevant for our purpose since their effect does not significantly exceed "normal" volatility (that is also covered in-data driven approaches). On the other hand, high probability events should be considered in the Best Estimate (rather than in a model for potential fluctuations around this Best Estimate). Hence, we have chosen to focus on "low probability / high impact" events. Those are relevant in any practical applications of life insurance and pension funds, where the focus is on the tail of the distribution (i.e., events with high impact that are possible but only with rather low probability), e.g., when determining risk capital for longevity risk under Solvency II. Therefore, the focus of our paper will be the calibration of **volatility parameters** that control the degree of (long-term) uncertainty in the stochastic projection, i.e., making sure that any **plausible** intensity of potential future change of life expectancy is covered by a model with **appropriate probability**.

As mentioned in Section 1, we concentrate on longevity risk and hence consider a scenario in the direction of an increase of life expectancy. We argue that although the adoption of a healthier lifestyle may have a huge impact on an individual's lifespan, it is very unlikely that a large portion of a population changes their lifestyle. Hence this is unlikely to significantly change life expectancy in a whole population (or in an insurer's book of business). We also argue that typical medical progress that aims at finding better cures for individual medical conditions cannot significantly reduce mortality in the elderly since typical seniors have a rather large number of conditions at the same time (cf. Collerton et al., 2009). Curing any single disorder is unlikely to significantly reduce overall mortality rates. But without a significant reduction of mortality rates of the elderly, life expectancy cannot rise significantly (cf., e.g., Olshansky et al., 2024). Therefore, we base our scenarios on research into the fundamental biology of ageing because slowing this down will delay all, or at least a majority, of age-related diseases and hence has significant potential to reduce even older-age mortality rates.

Although ageing is complex, there is now ample evidence from multiple species to support the possibility that ageing could be 'druggable', resulting in either compressed morbidity or extended healthy lifespan (cf., e.g., Partridge et al., 2020; Faragher and Hartley, 2024). Considerable progress has been made recently in uncovering major "hallmark" mechanisms which significantly improve mammalian health and thus longevity (cf., Lopez-Otin et al., 2013 and Lopez-Otin et al., 2023). These "hallmarks" include deregulation of nutrient sensing, cellular senescence, mitochondrial dysfunction, and loss of proteostasis, and the higher-level outputs from multiple biochemical pathways, such as the mammalian target of

rapamycin (mTOR) axis, each of which in turn is a series of discrete molecular events. Rather than being separate mechanisms the hallmarks identified to date appear to be co-dependent, with increasing evidence that positively or negatively modulating one hallmark can influence others in the same direction.

By way of example of some areas of biogerontology that are moving out of the laboratory and in the clinical pipeline, Guarente et al. (2024) review the evidence base and clinical trials status for eight drugs or drug classes (metformin, NAD+ precursors, glucagon-like peptide-1 receptor agonists, TORC1 inhibitors, spermidine, senolytics, probiotics, and anti-inflammatories) most of which have the potential to modulate hallmark mechanisms. Whilst initial data are promising in many instances (cf., e.g., Mannick and Lamming, 2023 and Justice et al., 2018) it needs to be kept in mind that the total number of licensed molecules (i.e., drugs) available globally is still well below 8000[4] (a number which includes compounds such as aspirin) and the overall rate of failure of clinical trials to phase 3 is of the order of 90%. However, subsets of clinical trials (particularly those in which drugs are 'repurposed' for uses other than the one for which they were originally licensed) have much higher success rates, sometimes reaching 60%. Although doubts remain in some quarters concerning whether a theoretical framework currently exists with sufficient explanatory power for some ageing phenomena (cf. Gems et al., 2024) successful outcomes for trials based on hallmarks are an established fact.

Of course, at this stage, nobody can know which interventions will prove most effective, at what point in time they will come to the market, and what the effect on human life expectancy will be. Hence any expert opinion will naturally be far from a perfect prediction. It seems, however, clear that with a probability that is larger than zero, an unprecedented change can happen that occurs well within the term of a deferred annuity or pension contract taken out by a young person today. Therefore, we argue that, although a model that considers a plausible expert opinion in the calibration of its "uncertainty parameters" will not be perfect, it will be more meaningful than a model that is calibrated to historic data only – at least at points in time, when there are potential medical breakthroughs at the horizon.

When it comes to estimating the three components of a scenario mentioned above (what, when, and how likely), we would like to stress one aspect that might seem counterintuitive at first glance: If a concrete scenario is discussed between mortality modellers and medical experts, one might develop a feeling what a plausible intensity of changes in life expectancy could be (e.g., from early clinical trials or results in model organisms). One might also rather easily agree on an estimate for the timing (e.g., from experience with the different stages of

[4] cf. De la Torre and Albericio (2024) as well as go.drugbank.com/stats.

bringing dugs to market). However, assigning a probability is extraordinary tricky: For calibrating a stochastic mortality model, one is typically not interested in the probability of the very specific scenario but rather in the probability (x%) of anything (maybe a different scenario) happening that leads to a life expectancy increase (relative to the current best estimate projection) that is at least as high as in the specified scenario. Before we discuss in Section 3, how a model can be calibrated such that x% of all random scenarios generated by the model lead to a life expectancy increase that is at least as high as the specified scenario, we develop such a scenario in the remainder of Section 2.

2.2 The scenario: Senolytics

We have chosen to base our scenario on the clinical use of senolytics. These are drugs that selectively eliminate senescent cells. Such cells are derived from conditional renewal populations in tissues (such as the dermal layer of the skin or the liver) where cells are regularly lost and must be replaced throughout life. As an anti-cancer mechanism division of any one cell is limited eventually producing populations of cells that cannot replicate but do not die. These cells become agents of sterile inflammation and accumulate in the body over time. In rodents, the elimination of such senescent cells considerably extends healthy lifespan, and in rare human diseases (especially Werner's syndrome) their premature accumulation causes accelerated ageing. Hence, their elimination in humans (more than 30 human trials are currently in progress) clearly has the potential to prevent clusters of diseases, slow down or even reverse ageing, and increase life expectancy.[5] The questions are to what extent mortality and hence life expectancy might be affected, within what timeframe this might happen, and how likely it is to occur. Recall again that we are interested in "tail events", i.e., "low probability / high impact" scenarios.

As a starting point, we consider the life expectancy extension caused by senescent cell removal in transgenic mice (cf. Baker et al., 2016). This is a maximum of 37% depending on genetic background and sex. Direct extrapolation from rodent data into humans is a gross oversimplification for a variety of reasons, some of which form the basis of our later "out-of-the-box"-scenario in Section 2.3, and one key difference is cause of death. Cancer prevention mechanisms are far less efficient in rodents than humans and tumours are thus the primary cause of death in most laboratory rodent strains (cf. Snyder et al., 2016). However although cancer is also a leading cause of death in humans, it kills a lower percentage of the population (estimates vary by country but a figure of 16% globally is often quoted with higher percentages of 25-30% for first world countries) and thus if, as seems

[5] Cf., e.g., Di Micco et al. (2021) or Baker et al. (2011).

probable, much of the lifespan extension effect of senolytics seen in rodents is via tumour suppression (perhaps paradoxically whilst senescence is a cell intrinsic tumour suppression mechanism senescent cells have a pro-carcinogenic effect on their neighbours if they are not removed) then pro-longevity effects could be smaller in humans. One might argue that a conservative estimate for the impact on humans would be only a third of the effect size seen in mice.

Conversely, it could be argued that the effect on humans might actually be larger than that for several reasons: First, other causes of death in humans (such as cardiovascular disease) have a clear senescent cell component and thus a senolytic could achieve equally large-scale improvements in healthy human lifespan by impacting these pathologies. Second, the potential for positive modulation of other hallmarks through senolytics exists and should be considered alongside. Finally, the extent to which simultaneous modulation of multiple Hallmarks (e.g., senolytics combined with modulation of nutrient sensing pathways using compounds such as Rapamycin or Everolimus) would allow "stacking" of interventions to achieve larger effects has not yet been tested and is thus unknown. This last aspect in particular is a critical knowledge gap within the field considered either from a biological or a risk forecasting perspective.

Accordingly, we have selected an impact on human life expectancy of 25% (approximately midway between the 37% observed in mice and the more conservative lower estimate of one third of this value).

When it comes to timing, we use a value of 10 years in our scenario. This consists of two years for phase II, another two years for phase III, one extra year for an indication license. Hence, off license prescription might be possible in five years. It might then take another five years until it reaches sufficient portions of the population to have a permanent and significant effect on life expectancy. This is consistent to the 16-year timescale often used as a rule of thumb for a drug to pass from phase zero to licensing.

Finally, we use an exemplary value of 1% for the probability of occurrence and end up with our illustrative expert scenario:

25% increase in life expectancy over a time horizon of 10 years with 1% probability.

We will use this expert scenario to perform our driver-driven model calibration in Section 4. Since the focus of our application is on pensioner mortality, we will consider an 25% increase in remaining life expectancy at age 65 (rather than at birth). Of course, a 25% increase in life expectancy at birth would result in an even larger absolute gain in years compared to a 25% increase in remaining life expectancy at age 65. Note that the effect of

senolytics has also been tested in mice that have already lived through a significant portion of their lifespan, see e.g. Yousefzadeh et al. (2018).[6]

The potential of Senolytics was also recently taken up by Club Vita (2024) who outline a possible scenario for the future development of human mortality. In their scenario, they assume that the discovery and distribution of a Senolytics drug will lead to a 10-year increase in remaining life expectancy at the age of 65 over the next 20 years. Unlike us, however, they refrain from specifying a probability of occurrence for their scenario so that their scenario cannot readily be used for calibrating the volatility parameters of a stochastic mortality model. Also, their analysis was not specifically focussed on low probability/high impact scenarios. Hence although their scenario is somewhat less extreme than ours, this is not necessarily inconsistent if theirs would be assigned a higher probability than 1%.

2.3 Out of the box: the killifish scenario

Due to the long-term nature of pension risks it might make sense to also consider scenarios over longer time horizons than 10 years. Also, since annuity and pension providers need to be able to meet their obligations even in very extreme scenarios, it is prudent to consider rather extreme scenarios, i.e., very low probability events.

We therefore look at an "out-of-the-box"-scenario which is highly improbable but not biologically impossible. The line of argument is based on the evolutionary biology of ageing, the life history of humans as a species, and analogies drawn between this and the modulation of ageing in the only other species currently known to have followed a similar evolutionary trajectory to mankind.

In evolutionary terms, ageing evolved early in the history of life on Earth in populations of organisms exposed to high extrinsic mortality (a feature of the biosphere to this day). This results in a combination of selection for genetic variations that enhance reproductive success early in life regardless of their deleterious effects later on in time (antagonistic pleiotropy or AP) coupled with the inability of natural selection to remove fitness neutral deleterious mutations (mutation accumulation or MA). However, unlike most species, humans have undergone at least one and possibly more extended evolutionary bottlenecks (essentially everyone alive on Earth today descends from a population of less than 1,000 individuals, cf. Zhivotovsky et al. (2003) and Hu et al. (2023)).

[6] They administered the senolytic fisetin to mice which were already 20 months of age (approximately 80% lifespan completed).

Although fitness neutral mutation frequencies are insensitive to natural selection (by definition) they are acutely sensitive to genetic drift. Drift rates in turn are an inverse function of the effective population size (N_e) and thus the exceptionally high drift rates resulting from the human genetic bottleneck may have dramatically altered the balance between AP and MA. This may be one factor underling the difficulties the biomedical field currently has in producing rodent models that are representative of human ageing pathologies.

However, *Nothobranchius furzeri* (the African turquoise killifish) is an extensively genetically bottlenecked vertebrate with an average lifespan of six months, and a physiology that recapitulates multiple signs of mammalian ageing at the molecular, cellular, organ and behavioural levels. It is also a species which shows a remarkable increase in life expectancy of 60% in response to resveratrol, the highest known impact of a single intervention in any species' life expectancy so far. Hence, to produce our "out-of-the-box" scenario we consider the possibility that a similarly high increase in life expectancy could occur in humans through a simple, but as yet undiscovered, intervention. Whilst an effect on human lifespan as large as resveratrol in killifish is highly unlikely, the idea that a simple (yet unknown) intervention might have a huge impact on human life expectancy is not implausible. For example, the macrolide antibiotic Rapamycin was first purified in the early 1970s and licensed for clinical use in the late 1990s; however, it was not until 2009 that it was discovered through the NIA Intervention Testing Program that it could increase rodent lifespan by up to 14% when given to old mice (cf. Harrison et al. 2009).

Of course, a 60% increase in life expectancy at birth would mean a significantly higher increase in remaining life expectancy at age 65, which is our reference figure in our application in Section 4. Hence, we use an illustrative scenario of an increase in remaining life expectancy of 100%. We also use a rather long-time horizon arguing that if something exists that works on humans like Resveratrol works on the Killifish, then it has not yet been discovered – so it is not an existing drug and would need a long term to get to the market. We therefore use a time horizon of 30 years. Since the estimation of a probability of occurrence is quite difficult, as already mentioned in Section 2.1, we refrain from specifying the probability of occurrence at this point. Instead, we will use this "out-of-the-box" scenario in Section 4 to test the plausibility of the resulting model calibration with regard to its projected long-term uncertainty. Arguably, a plausible model should assign a probability of occurrence to the killifish scenario that is well below the 1% of our senolytics-scenario to reflect its extremity, but also above 0% so that this scenario is not considered "virtually impossible".

3 Calibration methodology

In this section, we propose a methodology on how expert knowledge on the future of human life expectancy expressed in the form of an expert scenario can be integrated into the calibration of stochastic mortality models.

3.1 Specification of an expert scenario

Having presented an expert scenario for the future development of mortality in the previous section, we now discuss how such an expert scenario can be translated into a mathematical framework. An expert scenario consists of the following four components describing **what** might happen, **when** might it happen, and **how likely** it is to occur:

- The reference figure ($R_{x,t}$) is the underlying quantity the expert scenario refers to. It is an F_t-measurable random variable, i.e., a stochastic quantity as of points in time before t but known at time t. Such a figure might for instance be the mortality rate at a certain age x in calender year t, or the remaining life expectancy in calendar year t for a certain reference age x or at birth.
- The impact (I) describes the effect the scenario might have on the reference figure in the form of a relative change compared to the current best estimate.
- The time period (τ) indicates the time period over which the expert scenario might unfold.
- The quantile (Q) describes the probability that this scenario – or anything else that might have an impact of at least I on $R_{x,t}$ over the time period τ – will occur.

A stochastic mortality model is considered consistent with an expert scenario, if

$$\mathbb{P}\left(R_{x,\tau} \geq (1+I) \cdot \mathbb{E}\left[R_{x,\tau}\right]\right) = Q, \qquad (1)$$

i.e., if the probability that the reference figure $R_{x,\tau}$ will exceed its current (time zero) best estimate by at least the factor of $(1+I)$ is equal to Q.

3.2 Choice of reference figure

The specific choice of a suitable reference figure is of course closely related to the nature of the underlying expert scenario. If, for example, the expert scenario makes a statement about changes in mortality for a certain age range, a natural choice would be the corresponding one-year probability of death $q_{x,\tau}$, possibly averaged over a given age range. If the scenario – like our scenario discussed in the previous section – refers to the future development of

(remaining) human life expectancy, a suitable reference figure would be the expected remaining lifetime of an individual aged x at time τ, that is,

$$R_{x,\tau} = \mathbb{E}\big(T_{x,\tau}|F_\tau\big) = \sum_{t\geq 1} \mathbb{P}(\, T_{x,\tau} \geq t|F_\tau) + \, 0.5.$$

It corresponds to the remaining life expectancy of an individual aged x in year τ taking into account the current (as of time τ) best estimate survival rates[7] including the current best estimate mortality trend[8] assuming that people die – on average – in the middle of the year.

The remaining cohort life expectancy represents a suitable reference figure for our application for several reasons. First of all, it is an intuitive quantity that is widely known and easily understood by professionals from other disciplines without the need for in-depth actuarial knowledge. It is therefore well suited for an interdisciplinary approach. In contrast to the period life expectancy, which is solely based on mortality rates in a particular calendar year without taking into account future mortality improvements, cohort life expectancy represents the current best estimate for the actual remaining lifetime of an individual. Furthermore, it provides a natural aggregation of mortality rates for all ages above the reference age, which allows to focus on the most relevant age range for the longevity risk of pension funds and insurers.

3.3 Driver-driven calibration procedure

In this section, we provide guidance on how to incorporate expert knowledge in form of a single expert scenario into the calibration a stochastic mortality model. Of course, in principle, also multiple scenarios – for instance over different reference time periods – could be considered simultaneously. We will discuss this and other possible extensions in Section 5 and only consider the case of a single expert scenario in this section.

[7] In our numerical applications, we derive best estimate survival rates at future points in time by following the deterministic central path of mortality given by the prevailing mortality trend. This pragmatic approach is common in actuarial practice, see e.g., Cairns and El Boukfaoui (2021) or Freimann (2021), to avoid computationally burdensome nested simulations.

[8] In practice, this requires the estimation of the current mortality trend based on the most recently observed mortality pattern, which involves the risk of misspecifying the trend since it can be blurred by "normal" annual fluctuations around the trend, see Börger et al. (2021b). For simplicity, we assume that the mortality trend is observable for the calculation of remaining life expectancies at any future point in time.

Given an expert scenario $\left(R_{x,\tau}, I, \tau, Q\right)$, the aim of the driver-driven calibration is to find a reasonable model parametrization that is consistent with the expert scenario in terms of Equation (1) and produces plausible forecasts. Typically, the solution to this optimization problem is not unique as mortality models generally have several projection parameters. For instance, when relying on a (multivariate) random walk with drift for mortality projections in a parametric mortality model with two period effects, like the standard CBD model, a total of five parameters is required: two parameters for the drift – primarily determining the central projection – and three parameters for the two-dimensional covariance matrix – primarily determining the volatility. Obviously, further specifications are needed in addition to the given expert scenario to obtain a complete model parametrization.

To this end, all available information should be used and sensibly combined. In addition to the given expert scenario, this naturally also includes empirical values from historical mortality patterns. In general, we recommend the following four-stage procedure:

1. Starting point: data-driven calibration

As long as no reliable expert scenarios are available, values derived from historical data generally represent the best and most objective basis. Even if a certain expert scenario is available, it generally only relates to a certain time horizon, age range, and quantile of the probability distribution. For other time horizons, age ranges, or parts of the probability distribution, however, empirical values from historical data might still represent the only source of reliable information. Therefore, an objective data-driven calibration generally constitutes a reasonable starting point and provides an initial estimate for all model parameters.

2. Separation of "location parameters" and "volatility parameters"

The projection parameters of a stochastic mortality model can broadly be classified into two categories: "location parameters" that primarily determine the location of the central projection and "volatility parameters" that primarily control the degree of (long-term) uncertainty in the stochastic projection. Depending on the type of expert scenario, it may be appropriate to modify either "location parameters" or "volatility parameters" and to adopt the remaining parameters from the data-driven calibration. For example, if the considered expert scenario makes a statement about the median ($Q = 50\%$), the "location parameters" need to be adjusted accordingly, while the "volatility parameters" can remain unchanged. If, as in our example, the scenario is concerned with the evolution of future mortality in the tail of the distribution, it seems appropriate to only adjust the "volatility parameters". In order to identify the appropriate model parameters for the expert scenario at hand, a thorough analysis of the underlying model structure and the role of each model parameter is recommended. When dealing with more complex model structures, a sensitivity analysis of

the reference figure with respect to changes in selected model parameters might for instance provide valuable insights into the role of individual parameters.

3. Specification of a scaling method with a single degree of freedom

Next, a relative scaling factor S is introduced that acts on all model parameters that have been selected in the previous step. In many cases, it may be sufficient to select a single model parameter to avoid complexity. If multiple parameters are selected, a further adjustment might be required to ensure consistency between the scaling of different parameters, in particular when they are not of the same type. This can certainly be achieved in several ways, where the optimal choice depends on the specific mortality model at hand.[9] Finally, the optimal scaling factor S is derived numerically so that the resulting model calibration fulfills Equation (1).

4. Validation

Ultimately, the resulting model calibration should be thoroughly validated, which should of course also be the standard procedure for a purely data-driven calibration. For a driver-driven calibration, particularly key figures, time periods, and quantiles other than those of the given expert scenarios should be checked for plausibility. This is important in order to rule out the possibility that the resulting model parametrization – even if it matches the given expert scenario – leads to implausible results elsewhere.

4 Application

In this section, we apply the presented driver-driven calibration approach to a stochastic mortality model and compare the results with a purely data-driven calibration. We rely on our senolytics-scenario from Section 2.2 to perform the driver-driven calibration and use the "out-of-the-box"-scenario from Section 2.3 to test the plausibility of the longer-term uncertainty predicted by the resulting model. Further, we use historical mortality data from the Human Mortality Database for the male civilian population of England and Wales as an example. The details of this data set can be found in Appendix A.1. For the numerical applications, we perform Monte-Carlo simulations with 50,000 sample paths.

[9] For instance, one possible method might be multivariate normalized exponential tilting, see Freimann (2021) for a related application.

4.1 Considered stochastic mortality model

We consider a mortality model with stochastic trend changes as proposed by Börger and Schupp (2018), which builds on the structure of the established Cairns-Blake-Dowd (CBD) model of Cairns et al. (2006).

In the CBD model structure, the logit of the one-year probabilities of death is modelled as

$$logit(q_{x,t}) = log\left(\frac{q_{x,t}}{1-q_{x,t}}\right) = \kappa_t^{(1)} + \kappa_t^{(2)}(x-\bar{x}),$$

where $\bar{x}$ denotes the middle of the considered calibration age range. The first period effect $\kappa_t^{(1)}$ models the development of the general (age-independent) mortality level over time, whereas the second period effect $\kappa_t^{(2)}$ models the steepness of the mortality curve. For stochastic mortality projections, these two time-dependent processes need to be projected into the future.

To this end, Cairns et al. (2006) rely on a two-dimensional random walk with constant drift, which represents the prevailing (estimated) trend in mortality improvements. In the following section, we will use this model as a reference point as this is still the most popular choice in practice.

However, as argued by several authors[10], the assumption of a constant mortality trend over the whole projection horizon does not seem appropriate, especially over longer time horizons. In particular, it ignores the risk that the mortality trend might change again at future points in time, for instance as a result of a medical breakthrough as discussed in Section 2. Since the risk of a future acceleration of the mortality trend represents the main driver of long-term longevity risk, it should be adequately modelled.

Therefore, Börger and Schupp (2018) suggest a trend-change model, which we will also consider in what follows. They model the period effects as trend-stationary stochastic processes, where the underlying mortality trend can experience random changes in both directions at random future points in time. Specifically, the period effects are modelled as random fluctuations around an underlying stochastic process $\hat{\kappa}_t^{(i)}$, i.e.,

$$\kappa_t^{(i)} = \hat{\kappa}_t^{(i)} + \epsilon_t^{(i)}, \qquad i = 1, 2,$$

with $\epsilon_t = \left(\epsilon_t^{(1)}, \epsilon_t^{(2)}\right)$ being a a two-dimensional normal vector with mean zero and covariance matrix Σ. It represents transitory "normal" annual fluctuations in mortality

[10] See, among others, Börger and Schupp (2018) and Liu and Li (2017).

around the prevailing long-term trend, e.g., due to a flue wave. The underlying processes $\hat{\kappa}_t^{(i)}, i = 1,2$ are assumed to be continuous and piecewise linear, where their slopes $\hat{d}_t^{(i)}, i = 1,2$ are interpreted as the prevailing mortality trend at time t. The two processes are projected independently from each other by following the prevailing mortality trend, i.e.,

$$\hat{\kappa}_t^{(i)} = \hat{\kappa}_{t-1}^{(i)} + \hat{d}_t^{(i)}, \qquad i = 1,2.$$

In each projection year t, the mortality trend may experience a permanent change with probability of occurrence $p^{(i)}, i = 1,2$, where $S_t^{(i)} \in \{-1,1\}$ denotes the direction (i.e., the sign) and $M_t^{(i)} > 0, i = 1,2$ the magnitude of the trend change. Specifically,

$$\hat{d}_t^{(i)} = \hat{d}_{t-1}^{(i)} + O_t^{(i)} S_t^{(i)} M_t^{(i)}, \qquad i = 1,2,$$

where

- $O_t^{(i)}$ is Bernoulli-distributed with parameter $p^{(i)}$,
- $S_t^{(i)}$ is a discrete and uniformly distributed random variable on the set $\{-1,1\}$,
- and $M_t^{(i)}$ is modeled as a lognormal distribution with parameters $\mu_M^{(i)}$ and $\sigma_M^{(i)^2}$.

Börger and Schupp (2018) recommend using the same parameters for upward and downward changes, which leads to a symmetric distribution. This is a reasonable assumption in many cases, in particular in a purely data-driven calibration, as it assures that the prevailing mortality trend at any point in time represents the best estimate of the future trend. However, in cases where one-sided information is available, it may make sense to relax this assumption and to use different parameters for upward and downward trend changes, i.e., to model the conditional distribution $M_t^{(i)}|S_t^{(i)}$. Such a situation is, in particular, present in our case, in which only expert scenarios are considered that make a statemant about one side of the probability distribution, i.e., an increase in life expectancy, and no scenario that provides information on adverse developments. We will further discuss this in Section 4.3.1.

4.2 Starting point: data-driven calibration

As a first step, we carry out a purely data-driven calibration for both considered models (i.e., for the random walk with drift, as well as for the trend change model) for the male population of England and Wales. All rather technical details including the resulting model parameters can be found in Appendix A. For both models, we consider two variants: one that accounts for parameter uncertainty, i.e., the risk of not being able to reliably estimate the "true" model parameters from historical data of limited length, and one that does not. We compare the

Model	**Senolytics** $\mathbb{P}(\%)$	**Killifish** $\mathbb{P}(\%)$
RWD w. param. uncertainty (data-driven)	0.25%	0.07%
RWD w/o param. uncertainty (data-driven)	0%	0%
Trend w. param. uncertainty (data-driven)	0.31%	0.20%
Trend w/o param. uncertainty (data-driven)	0.13%	0.11%

Table 1 Resulting quantiles for the expert scenarios under a data-driven calibration.

resulting model parametrizations by looking at quantile charts for the remaining cohort life expectancy at age 65, which are shown in Figure 1.

We start by comparing the variants that do not incorporate parameter uncertainty (top figures). It is immediately apparent that both models provide a structurally different form of uncertainty. The random walk with drift produces rather narrow confidence intervals with light tails that widen only slowly over time. It has been claimed by several authors including Börger and Schupp (2018) and Liu and Li (2017) that this tends to underestimate long-term longevity risk. The trend change model, on the other side, generates confidence intervals with much more pronounced tails that widen at considerably faster rates in line with the growing uncertainty in long-term mortality trends.

When parameter uncertainty is taken into account (bottom figures), the uncertainty for both models increases. However, the incorporation of parameter uncertainty appears to have a stronger impact on the random walk with drift than on the trend change model for our exemplary calibration date. The reason for this is that the uncertainty in the starting trend for the trend change model is rather low in our example calibration (cf. figures and tables in Appendix A.3), whereas the uncertainty in the estimator for the drift is typically quite pronounced when using a calibration window of 20 years. The results of a data-driven calibration clearly depend heavily on the specific calibration date and the assumptions made in the calibration process, particularly with regard to the consideration of parameter uncertainty and the length of the calibration window when relying on the random walk with drift.

For these reasons, one should not expect a clear picture of the prevailing uncertainty in future mortality from a purely data-driven calibration – without any further expert judgement. Therefore, data-driven calibrations should always be thoroughly validated. For this, expert scenarios can provide valuable points of reference.

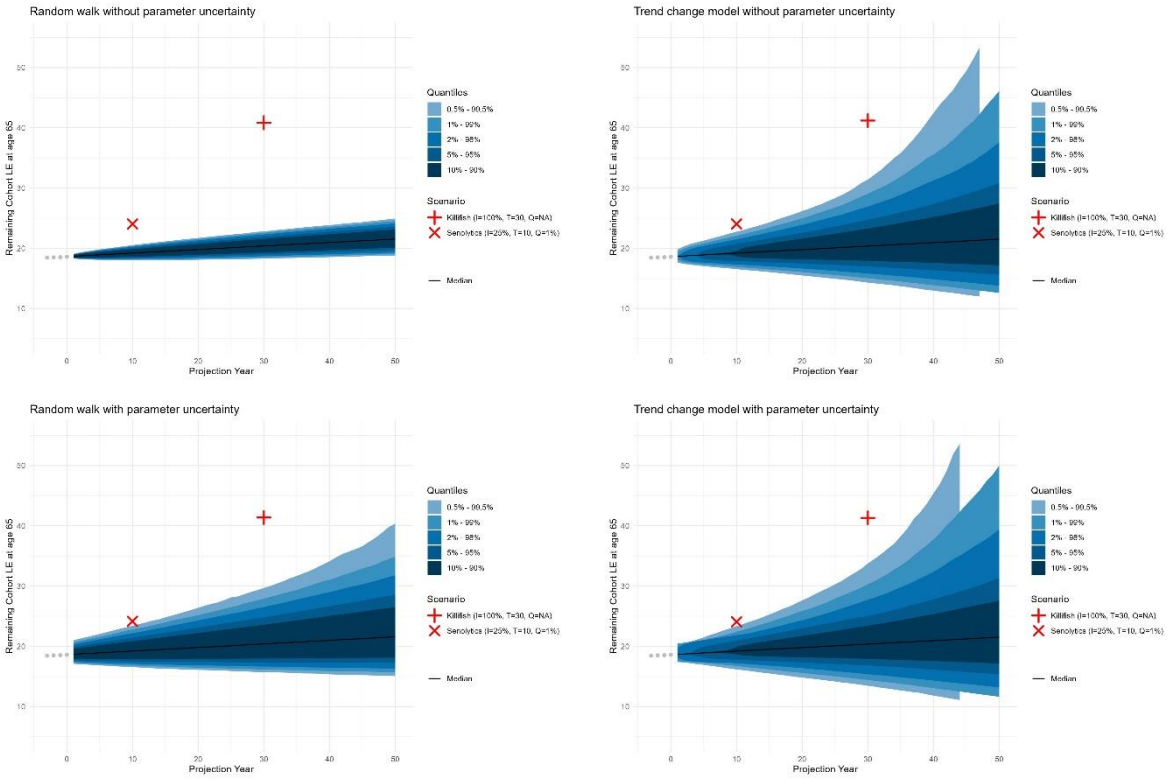

Figure 1 Data-driven calibration: resulting remaining cohort life expectancy for age 65.
Left: random walk with drift; Right: trend change model.
Top: without parameter uncertainty; Bottom: with parameter uncertainty.

The red crosses in Figure 1 show the impact of our expert scenarios from Section 2. Obviously, for all four considered models, the expert scenarios lie outside the considered confidence intervals. Table 1 shows the resulting quantiles, i.e., the exceedance probabilities, for our two expert scenarios resulting from the four considered data-calibrated models. For the random walk with drift without parameter uncertainty, increases in life expectancy as considered in our scenarios are "virtually impossible". When parameter uncertainty is taken into account, such increases are within the realm of possibility, but are significantly less likely than the 1% predicted for the senolytics-scenario. This is not surprising as a model with constant drift does not structurally fit a scenario in which there is a future fundamental and permanent change in mortality dynamics. The stochastic trend model, on the other side, is structurally better suited for this purpose. This is particularly evident in the shape of the confidence intervals, which are much more consistent with the assumption that uncertainty increases strongly over time which is also consistent to our two expert scenarios. Nevertheless, the resulting quantiles in Table 1 also remain significantly below the expert

assessment of 1% for the senolytics-scenario. In view of these results, the uncertainty implied by the four data-calibrated models appears too low.

In the following section, we therefore apply our proposed driver-driven calibration procedure to adjust the trend change model (with allowance for parameter uncertainty) so that it exactly matches the 1% probability of occurrence from our senolytics-scenario. For the sake of brevity, we will focus on the trend change model in our subsequent discussions and no longer consider the random walk with drift, since the structure of its projected long-term uncertainty does not match that of our expert scenarios.

4.3 Driver-driven calibration

Building on the data-driven calibration from the previous section, we now apply our driver-driven calibration procedure from Section 3.3 to the trend change model. As expert scenario, we consider the senolytics-scenario introduced in Section 2.2, i.e., an increase in remaining life expectancy at age 65 of at least 25% with a probability of occurrence of 1%.

4.3.1 Discussion of model parameters

As our scenario is concerned with the future of human life expectancy in a rather extreme scenario, i.e., in the tail of the distribution, the driver-driven calibration should target the model's "volatility parameters". Since the trend change model has several such parameters, it first needs to be determined which one should be targeted. In our model, the uncertainty in future life expectancy is jointly driven by the two period effects, whose future volatility is controlled by the following parameters (cf. Section 4.1):

- The noise term ϵ_t reflects annual fluctuations around the prevailing mortality trend resulting from transitory effects, such as a stronger or weaker seasonal flue wave. It therefore models short-term volatility rather than long-term volatility.
- The trend change probability $p^{(i)}$ indicates the probability that the mortality trend $\hat{\kappa}_t^{(i)}$ will undergo a permanent change from one year to the next. Obviously, a higher frequency of trend changes increases long-term volatility.
- Finally, the parameters $\mu_M^{(i)}$ and $\sigma_M^{(i)}$ of the lognormally distributed trend change intensities jointly determine the magnitude and volatility of future mortality trend changes and are therefore a main driver of long-term uncertainty.

In cases like this, where the model structure contains several "volatility parameters", it seems reasonable to choose the model parameter for scaling that is structurally best suited to the character of the expert scenario at hand. Since our scenario makes a statement about general

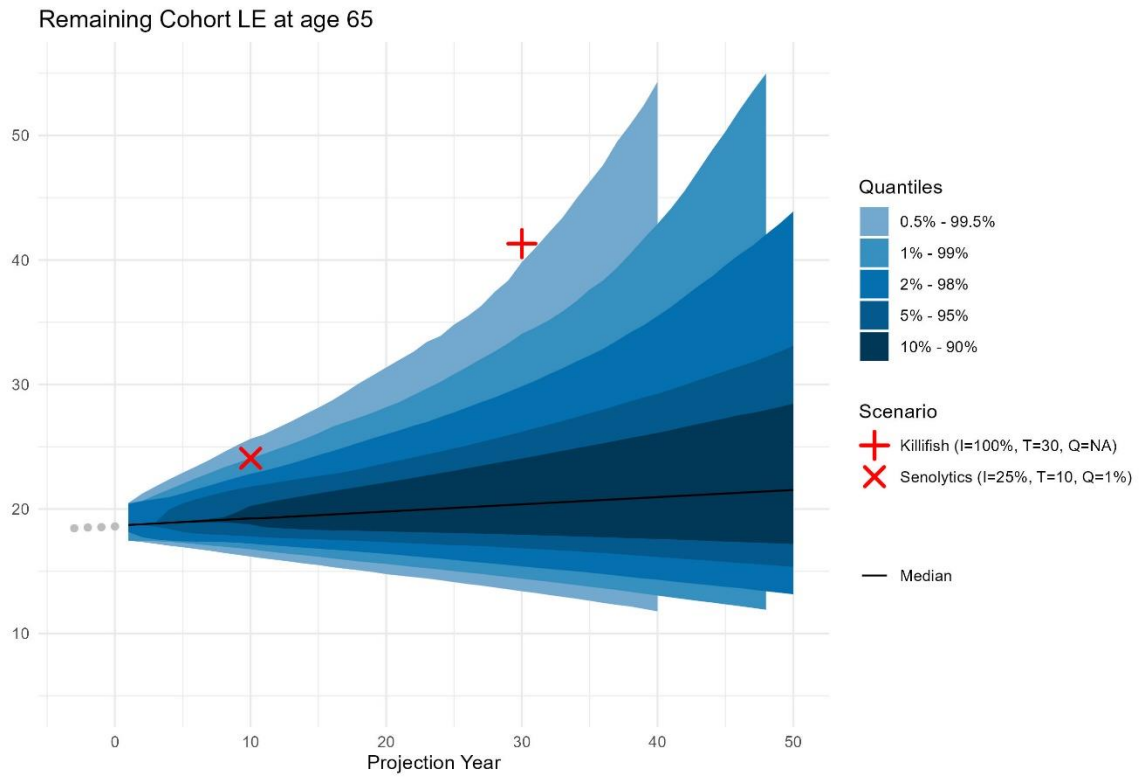

Figure 2 Driver-driven calibration of the trend change model: resulting remaining cohort life expectancy for age 65.

(rather than age-specific) population mortality, it seems appropriate to scale a parameter belonging to the first period effect $\kappa_t^{(1)}$, that drives the overall age-independent mortality development, rather than a parameter for the second period effect. Further, as our scenario makes a statement about long-term uncertainty, a parameter controlling the future dynamics of mortality trends (rather than transitory fluctuations in mortality) should be chosen. From the remaining set of eligible parameters $(p^{(1)}, \mu_M^{(1)}, \sigma_M^{(1)})$, we have considered two variants:

- a driver-driven variant with respect to the trend change probability $p^{(1)}$ and
- a driver-driven variant with respect to the trend change intensity $\mu_M^{(1)}$, where the parameter is only adjusted for trend changes that point in the direction of an increase in life expectancy. For trend changes in the other direction, we leave the data-driven parameter unchanged. As already noted in Section 4.1, this is in line with the character of our expert scenario, that makes a statement only about one side of the probability distribution, i.e., about a potential increase in life expectancy.

When validating the resulting models with regard to various key figures, projection horizons, and quantiles, the one-sided variant based on $\mu_M^{(1)}$ turned out to yield the most plausible results.[11] We show and discuss these results for this variant in the next section.

4.3.2 Resulting calibration and validation

The driver-driven calibration yields an optimal value for the parameter $\hat{\mu}_M^{(1)}$ of -4.24644 compared to its data-driven estimate of -4.61589, which increases the intensity of future mortality trend changes. When applying the model, it should be addressed whether the driver-driven estimate is used for the entire projection horizon or just for a limited projection period. Since our senolytics-scenario only makes a statement about the mortality evolution over the course of the first ten projection years and no statement about mortality dynamics thereafter, it seems reasonable to return to the data-driven estimate after the tenth projection year. Consequently, our calibrated model contains two regimes:

- a "stressed" driver-driven regime for the first ten projection years using $\hat{\mu}_M^{(1)} = -4.24644$ and
- a "normal" data-driven regime using $\hat{\mu}_M^{(1)} = -4.61589$ thereafter.

The latter is consistent with the principle that a data-driven estimate is always the most objective and sensible starting point as long as no further information is available.[12]

The resulting quantile chart for the remaining cohort life expectancy at age 65 is shown in Figure 2. As intended, the model's predictions after ten years correspond exactly to the prediction of the senolytics expert scenario with an exceedance probability of 1%. Compared to the data-driven results in Figure 1 in Section 4.2, the confidence intervals are now wider reflecting an increase in short- and thus also in long-term uncertainty.

To check the resulting long-term uncertainty for plausibility, a comparison with our "out-of-the-box" expert scenarios from Section 2.3 can provide a valuable point of reference. The resulting exceedance probability is 0.43% which seems plausible, given that it is significantly less likely than the 1% from our senolytics-scenario reflecting its extremeness,

[11] In particular, it turned out that in the first variant the annual trend change probability $p^{(1)}$ would have to be scaled up quite drastically to an implausibly high value of more than 10% in order to reach the expert scenario's prediction after ten projection years.

[12] The driver-calibrated parameter could of course also be used for the entire projection horizon, which would lead to a model with an even higher long-term uncertainty.

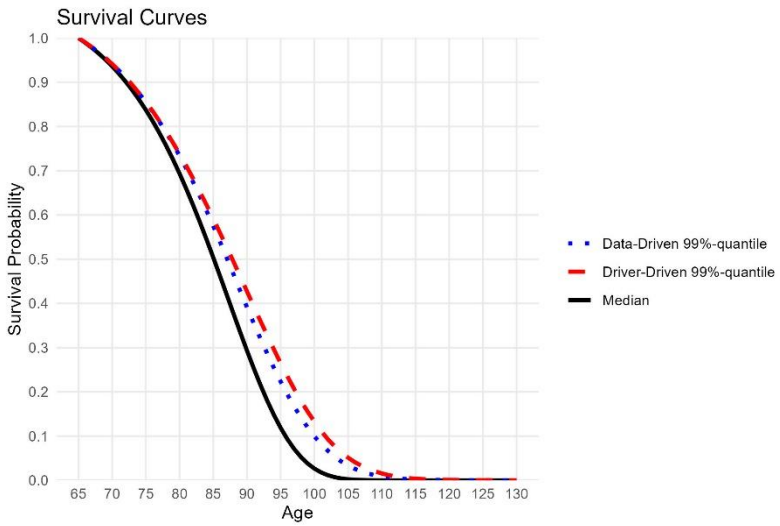

Figure 3 99%-quantile of the survival curve for age 65 for the driver-driven calibration in comparison to the median and the data-driven calibration.

but is still above zero, meaning that the scenario is not considered "virtually impossible" by the model.

For validation purposes, Figure 3 shows the 99%-quantile of the survival curve for age 65 for the driver-driven calibration in comparison to the median and the data-driven calibration. As expected, the 99%-quantile of the survival probability under the driver-driven calibration lies significantly above its purely data-driven counterparty. Nevertheless, the results appear plausible overall, given that the shape of the survival curve is not structurally distorted, and the results are not too extreme.

Overall, we conclude that the model provides a reasonable depiction of the uncertainty in future mortality, particularly in the long run, and constitutes a valuable alternative to purely data-calibrated models.

Risk Measure	$\boldsymbol{\alpha = 90\%}$	$\boldsymbol{\alpha = 99.5\%}$
Value-at-Risk at level α	13%	33%
Tail-Value-at-Risk at level α	31%	46%

Table 2 Relative increase in risk for the insurer when using the driver-driven calibration compared to the purely data-driven calibration under different risk measures.

4.4 Risk analysis

Finally, we take a look at how an insurer's assessment of longevity risk differs between our driver-calibrated and the purely data-calibrated trend change model in a simplified case study.

We consider an insurer holding a simplified portfolio of life annuity contracts that is closed to new business. These contracts pay a constant benefit of one currency unit at the beginning of each year for as long as the beneficiary lives. Regarding the initial structure of the portfolio, we assume that this portfolio has been built up over the past 50 years by selling the exact same type of contract to the same number of customers with an initial age of 65 at the beginning of each year. The age structure of the portfolio at the start of the simulation thus results naturally from the realized survival rates in the past. For simplicity, we assume that deaths occur according to realized mortality and disregard idiosyncratic small sample risk that arises in portfolios of limited size since it is typically much less relevant compared to mortality trend risk.

Following the standard approach for quantifying longevity risk in annuity portfolios, we look at the probability distribution of the centred random present value of future liabilities $L - \mathbb{E}(L)$[13]. To keep the focus on longevity risk, we use a constant annual interest rate of 2% for discounting. We quantify longevity risk by means of the risk measures Value-at-Risk and Tail-Value-at-Risk at threshold probabilities of 90% and 99.5%. The resulting increase in the insurer's risk when using the driver-calibrated model compared to the purely data-driven model is shown in Table 2.

The increase in $VaR_{99.5\%}$ of 33% and in $TVaR_{99.5\%}$ of 46%, respectively, show that the driver-calibrated model leads to a significantly higher assessment of longevity risk than the purely data-driven reference model. The fact that the increase in risk is more pronounced at the 99.5% threshold than at the 90% threshold shows that the differences are primarily due to a different assessment of tail risks. Since the adequate assessment of tail risks is of particular relevance for pension funds and insurers, especially in the context of risk-based solvency regimes like Solvency II, the driver-driven method we propose offers a valuable addition to established purely data-driven approaches.

[13] See for instance Cairns et al. (2014), Coughlan et al. (2011), or Börger et al. (2021a). The corresponding formulas can be found in Börger et al. (2021a).

5 Conclusion

The future of human life expectancy is currently marked by significant uncertainty in both directions, in particular due to environmental effects, multi-resistant germs, and lifestyle factors in one direction and due to recent advancements in anti-ageing research in the other direction. When modelling and managing longevity risks, it is important that stochastic mortality models provide an adequate picture of the prevailing uncertainty. Expert insights from other disciplines can help to get a more meaningful impression of the prevailing uncertainty than a look in the "data rearview mirror", yet such knowledge is typically not considered in longevity risk management as stochastic mortality models are typically calibrated in a purely data-driven way.

To close this gap, we have proposed a novel "driver-driven" approach for the calibration of stochastic mortality models. The core idea is to calibrate the volatility parameters of a stochastic mortality model such that its projection matches a given expert scenario with a given probability of occurrence. The methodology presented is, in principle, applicable to any mortality model and to expert scenarios with respect to increasing or decreasing mortality. Our work offers guidelines which calibration steps should be performed in order to achieve an adequate and plausible model calibration. Using exemplary (but plausible) scenarios from the field of anti-ageing research, we have demonstrated that such a "driver-driven" calibration can lead to a plausible and structurally different assessment of longevity risk than traditional "data-driven" approaches, especially with regard to tail risks.

Such a driver-driven calibration is of course not without limitations and should be viewed as a complement to, rather than a substitute for a data-driven calibration. In particular, it offers a valuable addition to traditional purely data-driven approaches in times when there are "low probability / high impact" scenarios on the horizon, that are considered plausible by experts in their respective field but are "virtually impossible" in models calibrated to historical data only.

Finally, we would like to mention several possible extensions of our approach that we leave for future research. First, it would be interesting to jointly consider multiple expert scenarios in the calibration. For instance, one could consider different scenarios that relate to different time periods, different quantiles, different directions (increase and decrease of mortality) and combinations thereof. Since a model can generally not exactly match multiple expert scenarios at the same time, this would require some kind of optimization criteria which naturally comes at the cost of higher complexity. Second, it would be worthwhile to take a closer look at expert scenarios that are of structurally different type than the scenarios we considered. Particularly relevant could be the exploration of scenarios that anticipate the possibility of sudden "jump-like" shifts in future mortality levels. This is especially relevant

when considering scenarios where conditions remain stable for an extended period, followed by an event that leads to a sudden drastic shift in mortality. In such cases, it should be critically examined whether models that include a jump component would be more suitable than typical stochastic mortality models.

Appendix

A. Model calibration

A.1. Data

For our data-driven calibrations, we use data of the civilian male population of England and Wales for the years 1841 – 2021 over the age range of 60 – 109 years from the Human Mortality Database (data downloaded on 01 July 2024 from: http://www.mortality.org). We calibrate the CBD model structure via a standard maximum likelihood estimation approach based on the assumption of binomially distributed deaths, see Villegas et al. (2018).

A.2. Data-driven calibration for the random walk with drift

For the calibration of the RWD to this data, we follow Cairns et al. (2006) and calibrate the drift and covariance matrix to the most recent 20 years of data using a standard maximum likelihood estimation approach. The two coronavirus years 2020 and 2021 are treated as outliers and omitted in order to avoid an overestimation of long-term volatility. To ensure consistency of the central projection between the RWD and the trend change model, the drift estimate is set equal to the expected starting trend of the trend process, which is specified in the following section. This yields the following estimates:

$$\hat{\mu} = (-0.007386, 0.000142), \quad \hat{\Sigma} = \begin{pmatrix} 6.02923 \times 10^{-4} & 1.97495 \times 10^{-5} \\ 1.97495 \times 10^{-5} & 1.12944 \times 10^{-6} \end{pmatrix}.$$

Parameter uncertainty is accounted for by sampling the drift vector and the covariance matrix from a suitable normal-inverse-Wishart distribution, see Cairns et al. (2006) for details.

A.3. Data-driven calibration for the trend change model

For the calibration of the trend change model to historical data, we follow Börger et al. (2019) and apply an iterative pseudo maximum likelihood estimation approach. We refer to this paper for technical details. Since data on mortality trend changes are sparse, we account for parameter uncertainty in the starting values as well as in the trend change parameters.

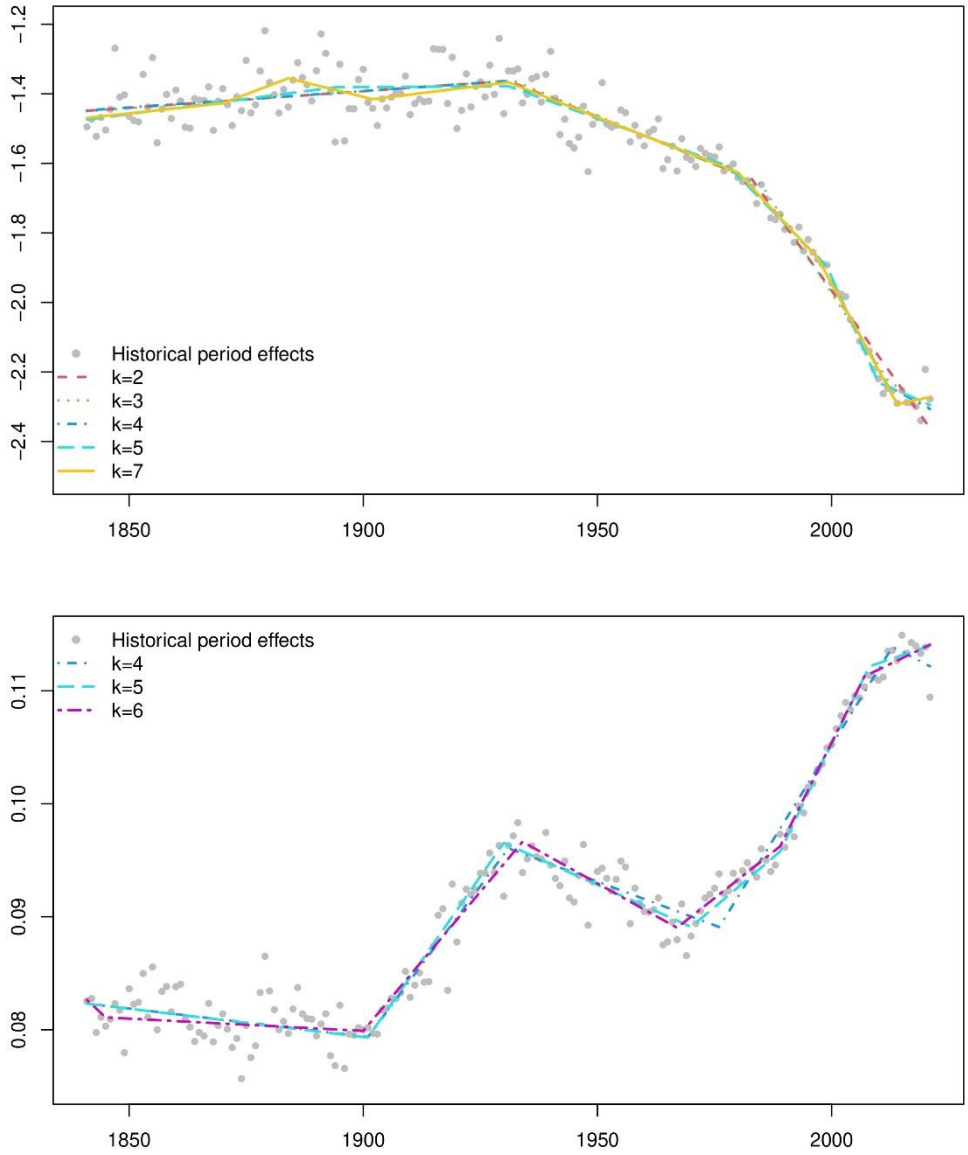

Figure 4 Period effects for English and Welsh males (dotted), optimal realizations for the actual trend processes given different numbers of trend changes k (colored dashed lines) for $\kappa_t^{(1)}$ (upper panel) and for $\kappa_t^{(2)}$ (lower panel).

For a given number of trend changes in the time series (k), the calibration algorithm aims to find the historical mortality trend with the highest likelihood, where the trend is assumed to be continuous and piecewise linear with k trend changes over time. Each value of k is then assigned a relative probability corresponding to the resulting goodness of fit. Parameter

uncertainty in the starting values for the trend processes is accounted for by sampling the initial values from this discrete probability distribution.

k	$\widehat{\boldsymbol{\kappa}}_{t=0}^{(1)}$	$\widehat{\boldsymbol{d}}_{t=0}^{(1)}$	$\mathbb{P}(\%)$	**Trend change years**
2	-2.36149	-0.01883	1.59856	1930, 1982
3	-2.30558	-0.00907	0.20882	1931, 1984, 2011
4	-2.30645	-0.00733	95.62285	1929, 1978 1996, 2011
5	-2.29494	-0.00586	1.52582	1893, 1930, 1977, 1998, 2009
6	N/A	N/A	0	N/A
7	-2.27184	0.00303	1.04395	1869, 1883, 1901, 1930, 1979, 1996, 2013

Table 3 Optimal historical trend changes and empirical distributions for the starting trend for $\kappa_t^{(1)}$.

k	$\widehat{\boldsymbol{\kappa}}_{t=0}^{(2)}$	$\widehat{\boldsymbol{d}}_{t=0}^{(2)}$	$\mathbb{P}(\%)$	**Trend change years**
4	0.11216	-0.00021	0.14402	1900, 1930, 1975, 2012
5	0.11399	0.00014	99.52092	1900, 1929, 1969, 1988, 2007
6	0.11408	0.00020	0.33506	1844, 1899, 1933, 1966, 1988, 2006

Table 4 Optimal historical trend changes and empirical distributions for the starting trend for $\kappa_t^{(2)}$.

Figure 4 shows the historical period effects $\kappa_t^{(i)}, i = 1,2$ and the optimal realizations of the underlying trend processes found by the calibration algorithm for different numbers (k) of historical trend changes. The corresponding parameter estimates and their likelihood are given in Table 3 for $\kappa_t^{(1)}$ and in Table 4 for $\kappa_t^{(2)}$, respectively, where only values of k with a relative likelihood of at least 0.1% are considered.

Furthermore, the algorithm provides the following estimates for the trend change parameters:

$$\left(\hat{p}^{(1)}, \hat{\mu}_M^{(1)}, \hat{\sigma}_M^{(1)}\right) = (0.02242, -4.61589, 0.381)$$

$$\left(\hat{p}^{(2)}, \hat{\mu}_M^{(2)}, \hat{\sigma}_M^{(2)}\right) = (0.02795, -7.37, 0.16348)$$

with corresponding covariance matrices of standard errors of

$$SE^{(1)} = \begin{pmatrix} 1.53153 \times 10^{-4} & -5.9467 \times 10^{-6} & 1.80779 \times 10^{-6} \\ -5.9467 \times 10^{-6} & 5.48394 \times 10^{-3} & -2.54425 \times 10^{-5} \\ 1.80779 \times 10^{-6} & -2.54425 \times 10^{-5} & 2.68535 \times 10^{-3} \end{pmatrix},$$

$$SE^{(2)} = \begin{pmatrix} 1.28927 \times 10^{-4} & -3.82315 \times 10^{-5} & 8.81788 \times 10^{-5} \\ -3.82315 \times 10^{-5} & 3.83876 \times 10^{-2} & -2.4198 \times 10^{-3} \\ 8.81788 \times 10^{-5} & -2.4198 \times 10^{-3} & 2.40788 \times 10^{-2} \end{pmatrix}.$$

We account for parameter uncertainty in the trend change parameters as suggested by Börger et al. (2019): At the start of each simulation path, a multivariate normal random vector is generated with mean equal to the estimated trend change parameters and covariance matrix $SE^{(i)}$. Then, the the first component is transformed to a beta distribution with same mean and variance to obtain a reasonable range for the trend change probabilities between zero and one. Analogously, the third component is transformed into a corresponding gamma distribution to ensure positivity for the volatility parameters.

Finally, the covariance matrix for annual fluctuations around the actual underlying mortality trend is estimated as

$$\hat{\Sigma} = \begin{pmatrix} 9.09578 \times 10^{-4} & 2.85244 \times 10^{-5} \\ 2.85244 \times 10^{-5} & 2.20712 \times 10^{-6} \end{pmatrix}.$$

References

Baker, D. J., Childs, B. G., Durik, M., et al. (2016). Naturally occurring $p16^{Ink4a}$-positive cells shorten healthy lifespan. *Nature*, 530(7589), 184–189.

Baker, D. J., Wijshake, T., Tchkonia, T., et al. (2011). Clearance of p16Ink4a-positive senescent cells delays ageing-associated disorders. *Nature*, 479(7372), 232–236.

Börger, M. and Schupp, J. (2018). Modeling trend processes in parametric mortality models. *Insurance: Mathematics and Economics*, 78, 369–380.

Börger, M., Schönfeld, J., and Schupp, J. (2019). Calibrating mortality processes with trend changes to multi-population data. *Living to 100 Monograph 2020*, Society of Actuaries, 1–25.

Börger, M., Ruß, J., and Freimann, A. (2021a). A combined analysis of hedge effectiveness and capital efficiency in longevity hedging. *Insurance: Mathematics and Economics*, 99, 309–326.

Börger, M., Schupp, J., and Ruß, J. (2021b). It takes two: why mortality trend modeling is more than modeling one mortality trend. *Insurance: Mathematics and Economics*, 99, 222–232.

Cairns, A. J. G., Blake, D., and Dowd, K. (2006). A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration. *The Journal of Risk and Insurance*, 73(4), 687–718.

Cairns, A. J. G., Dowd, K., Blake, D., and Coughlan, G. D. (2014). Longevity hedge effectiveness: A decomposition. *Quantitative Finance*, 14(2), 217–235.

Cairns, A. J. G. and El Boukfaoui, G. (2021). Basis risk in index-based longevity hedges: A guide for longevity hedgers. *North American Actuarial Journal*, 25(Supp1), S97–S118.

Club Vita (2024). *The Risk of Living Longer. What if new senolytic drugs fulfil their potential?* https://www.clubvita.net/uk/news-and-insights/the-risk-of-living-longer-what-if-new-senolytic-drugs-fulfil-their-potential (accessed 10/17/2024).

Collerton, J., Davies, K., Jagger, C., et al. (2009). Health and disease in 85 year olds: baseline findings from the Newcastle 85+ cohort study. *British Medical Journal*, 339:b4904, 1–11.

Coughlan, G. D., Khalaf-Allah, M., Ye, Y., et al. (2011). Longevity hedging 101: A framework for longevity basis risk analysis and hedge effectiveness. *North American Actuarial Journal*, 15(2), 150–176.

Currie, I. D. (2016). On fitting generalized linear and non-linear models of mortality. *Scandinavian Actuarial Journal*, 2016(4), 356–383.

De la Torre, B. G. and Albericio, F. (2024). The Pharmaceutical Industry in 2023: An Analysis of FDA Drug Approvals from the Perspective of Molecules. *Molecules*, 29(585), 1–15.

Di Micco, R., Krizhanovsky, V., Baker, D., et al. (2021). Cellular senescence in ageing: from mechanisms to therapeutic opportunities. *Nature reviews Molecular cell biology*, 22(2), 75–95.

Faragher R. G. A. and Hartley R. C. (2024). Drug discovery for ageing: SIMPs, NEDs and screening challenges. *Nature Reviews Drug Discovery*, 23, 725–726.

Fleming A. (1929). On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of B. influenzae. *British journal of experimental pathology*, 10(3), 226–236.

Freimann, A. (2021). Pricing longevity-linked securities in the presence of mortality trend changes. *ASTIN Bulletin: The Journal of the IAA*, 51(2), 411–447.

Gems, D., Okholm, S., and Lemoine, M. (2024). Inflated expectations: the strange craze for translational research on aging: Given existing confusion about the basic science of aging, why the high optimism in the private sector about the prospects of developing anti-aging treatments? *EMBO reports*, 25(9), 3748–3752.

Guarente, L., Sinclair, D. A., and Kroemer, G. (2024). Human trials exploring anti-aging medicines. *Cell Metabolism*, 36(2), 354–376.

Harrison, D. E., Strong, R., Sharp, Z. D., et al. (2009). Rapamycin fed late in life extends lifespan in genetically heterogeneous mice. *Nature*, 460(7253), 392–395.

Hu, W., Hao, Z., Du, P., et al. (2023). Genomic inference of a severe human bottleneck during the Early to Middle Pleistocene transition. *Science*, 381(6661), 979–984.

Institute and Faculty of Actuaries (2023). *Report of the post-COVID biometric assumption setting working party*. 30th November 2023.

Justice, J. N., Ferrucci, L., Newman, A. B., et al. (2018). A framework for selection of blood-based biomarkers for geroscience-guided clinical trials: report from the TAME Biomarkers Workgroup. *Geroscience*, 40, 419-436.

Khaw K. T., Wareham N, Bingham S., et al. (2008). Combined impact of health behaviours and mortality in men and women: the EPIC-Norfolk prospective population study. *PLoS medicine*, 5(1):e12, 39–47.

Liu, Y., and Li, J. S. H. (2017). The locally linear Cairns–Blake–Dowd model: a note on delta–nuga hedging of longevity risk. *ASTIN Bulletin: The Journal of the IAA*, 47(1), 79 – 151.

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

López-Otín C., Blasco M. A., Partridge L., et al. (2013). The hallmarks of aging. *Cell*, 153(6): 1194–1217.

López-Otín, C., Blasco, M. A., Partridge, L., et al. (2023). Hallmarks of aging: an expanding universe. *Cell*, 186(2), 243–278.

Mannick, J. B., and Lamming, D. W. (2023). Targeting the biology of aging with mTOR inhibitors. *Nature Aging*, 3(6), 642–660.

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

Olshansky, S. J., Willcox, B. J., Demetrius, L., et al. (2024). Implausibility of radical life extension in humans in the twenty-first century. *Nature Aging*, 4, 1635–1642.

Partridge, L., Fuentealba, M., and Kennedy, B. K. (2020). The quest to slow ageing through drug discovery. *Nature Reviews Drug Discovery*, 19(8), 513–532.

Patkee, S. and Strange, A. (2023). *The future of life expectancy: Forecasting long-term mortality improvement trends for insurance*, Swiss Re Institute.

Snyder, J. M., Ward, J. M. and Treuting, P. M. (2015). Cause-of-Death Analysis in Rodent Aging Studies. *Veterinary Pathology*, 53(2), 233–243.

Villegas, A. M., Millossovich, P., and Kaishev V. K. (2018). StMoMo: An R package for stochastic mortality modeling. *Journal of Statistical Software*, 84(3), 1–38.

Yousefzadeh, M. J., Zhu, Y., McGowan, S. J., et al. (2018). Fisetin is a senotherapeutic that extends health and lifespan. *EbioMedicine*, 36(2018), 18–28.

Zhivotovsky, L. A., Rosenberg, N. A., and Feldman, M. W. (2003). Features of evolution and expansion of modern humans, inferred from genomewide microsatellite markers. *The American Journal of Human Genetics*, 72(5), 1171–1186.