Asmik Nalmpatian

Advanced Approaches in Mortality Modeling with Application on Complex Data Situations

Dissertation an der Fakultät für Mathematik, Informatik und Statistik
der Ludwig-Maximilians-Universität München

Eingereicht am 12.11.2025

Asmik Nalmpatian

Advanced Approaches in Mortality Modeling with Application on Complex Data Situations

Dissertation an der Fakultät für Mathematik, Informatik und Statistik
der Ludwig-Maximilians-Universität München

Eingereicht am 12.11.2025

Erster Berichterstatter: Prof. Dr. Christian Heumann
Zweiter Berichterstatter: Prof. Dr. Helmut Küchenhoff
Dritter Berichterstatter: Prof. Dr. Ralph Brinks

Tag der Disputation: 29.01.2026

We ought then to regard the present state of the universe as the effect of its anterior state and as the cause of the one which is to follow. Given for one instant an intelligence which could comprehend all the forces by which nature is animated and the respective situation of the beings who compose it - an intelligence sufficiently vast to submit these data to analysis - it would embrace in the same formula the movements of the greatest bodies of the universe and those of the lightest atom; for it, nothing would be uncertain and the future, as the past, would be present to its eyes.

Laplace (1825) on his vision of a universe where the present is shaped by the past and dictates the future. Yet, despite this deterministic view, the complexities of human life and death transcend mere mathematical formulations, suggesting a higher order that extends empirical analysis. This order appears to balance scientific determinism with the enigmatic nature of human existence, while concurrently echoing the enduring impact of data. The data we leave behind continues to fuel scientific advancements for future generations, even beyond death.

Acknowledgments

First and foremost, I thank God for His grace, mercy and orchestration that made it possible to take on this dissertation and to see it through to completion, despite the demands of a challenging professional life in parallel. Throughout this journey, I have been blessed with many wonderful people who have accompanied and supported me. My deepest gratitude goes to:

- *Prof. Dr. Christian Heumann, for his tireless dedication and patient supervision throughout the years, for his guidance, encouragement and countless insightful discussions that have been instrumental to this work.*
- *Bill Jackson, Dr. Stefan Pilz and Levent Alkaya, my co-authors, for their continuous support and sharing their deep expertise in insurance and statistics, as well as their generous feedback throughout.*
- *Prof. Dr. Helmut Küchenhoff, Matthias, Sevag, André, Alex, Andreas, Yvonne, Max and the entire team at the Statistical Consulting Unit, for their inspiration, invaluable support and for laying the foundation of my understanding, values and expertise in complex applied statistics and data science.*
- *All participating entities, institutions and their related contacts, for the engaging collaborations and for sharing their expertise that formed the backbone of the empirical analyses in this dissertation.*
- *My family and loved ones, for their unlimited understanding and encouragement throughout every step of this path.*

Summary

Understanding human mortality is critical in a time of increasing longevity, aging populations, and unanticipated mortality phenomena like the COVID-19 pandemic. While it is impossible to predict individual mortality precisely, it is essential to develop reliable models that capture mortality trends to support decision-making in insurance, public health, and policy planning. Global increases in life expectancy have led to new challenges in pension sustainability, healthcare costs, and longevity risk assessment. Meanwhile, events such as pandemics demonstrate the need for models that can dynamically adjust to extremes on short notice. This dissertation presents a multifaceted approach to mortality modeling, integrating classical actuarial models with modern statistical and machine learning techniques. The focus lies not just on predictive accuracy, but also on addressing complex scenarios, such as data scarcity, whether due to limited experience or insufficient specific details, extreme events and structural differences across populations.

This thesis comprises four papers, each addressing a unique modeling challenge. The first study introduces a Generalized Additive Model (GAM) within an Age-Period-Cohort (APC) framework to incorporate pandemic effects into multi-populational and cross-country mortality forecasting. By modeling COVID-19 as a temporal disruption, the method extrapolates mortality under varying pandemic scenarios. GAM-APC achieved the most accurate post-pandemic forecasts among the observed benchmark models, with the flattening pandemic effect showing best alignment to observed data for middle-aged and elderly.

The second study proposes a hierarchical Gradient Boosting Machine model to support mortality estimation in data-scarce countries by learning from data-rich regions. A global-local modeling structure ensures transferability and interpretability. The two-step model strongly outperformed single-country models, especially in smaller countries, while preserving local specificity.

The third study addresses regions with no portfolio data at all by developing a transfer learning framework. By leveraging overall population data categorized by age and gender, along with comprehensive insured data from eight other countries, the study bridges the gap using correlations and public data from these countries. It reuses the pretrained global model from the previous study on similar countries and projects to the United Kingdom (UK) by synthetic data generation and a drift adjustment model. This approach yields predictions that closely align with official insured tables for UK. While the official tables are used here for evaluation purposes, in many countries, such official references are either unavailable or lack the granularity needed for portfolio-specific applications. Therefore, the framework demonstrates its value in reconstructing detailed mortality structures beyond age and gender, especially for markets or products where no insured data exist.

The fourth study builds a simulation framework using Iterative Proportional Fitting to generate synthetic segmented mortality datasets. These simulations allow researchers to model mortality for populations where detailed demographic or insured data is unavailable. The synthetic datasets preserve realistic distributions and enable modeling precision across different demographic segments.

Across all papers, reproducibility and practical implementation are ensured through open-source `R` and `Python` software. This work contributes to the evolving field of mortality modeling by integrating methodological innovation with real-world constraints, laying the groundwork for further demographic and actuarial advancements.

Zusammenfassung

Das Verständnis der Sterblichkeit ist angesichts steigender Lebenserwartung, alternder Bevölkerungen und Ereignissen wie der COVID-19-Pandemie zentral. Verlässliche Modelle sind nötig, um fundierte Entscheidungen in Versicherung, Gesundheitswesen und Politik zu treffen. Der weltweite Anstieg der Lebenserwartung hat zu neuen Herausforderungen in der Nachhaltigkeit von Rentensystemen, den Gesundheitskosten und der Bewertung des Langlebigkeitsrisikos geführt. Zudem verdeutlichen unvorhergesehene Ereignisse wie Pandemien den Bedarf an Modellen, die sich schnell an extreme Bedingungen anpassen können. Diese Dissertation stellt einen vielseitigen Ansatz zur Sterblichkeitsmodellierung vor, der traditionelle versicherungsmathematische Modelle mit modernen statistischen und Machine Learning Techniken verbindet. Der Fokus liegt nicht nur auf Vorhersagegenauigkeit, sondern auch auf komplexen Szenarien wie Datenknappheit, Extremereignissen und strukturellen Unterschieden zwischen Bevölkerungen.

Die Dissertation umfasst vier Arbeiten, die jeweils eine Modellierungsherausforderung adressieren. Die erste Studie führt ein Generalisiertes Additives Modell (GAM) in einer Age-Period-Cohort (APC) Struktur ein, um Pandemieeffekte in länderübergreifende Sterblichkeitsprognosen zu integrieren. COVID-19 wird als zeitlicher Bruch modelliert, um Sterblichkeit unter verschiedenen Pandemieszenarien zu extrapolieren. GAM-APC erzielte die genauesten Prognosen für die Zeit nach der Pandemie unter den untersuchten alternativen Modellen, wobei der abflachende Pandemieeffekt die beste Übereinstimmung mit den beobachteten Daten für Menschen mittleren und höheren Alters zeigte.

Die zweite Studie schlägt ein hierarchisches Gradient Boosting Machine Verfahren vor, um die Sterblichkeitsschätzung in datenarmen Ländern zu unterstützen, indem von datenreichen Regionen gelernt wird. Eine global-lokale Modellstruktur gewährleistet Übertragbarkeit und Interpretierbarkeit. Das zweistufige Modell übertraf die isolierten Modelle für einzelne Länder deutlich, insbesondere in kleineren Ländern, und bewahrte die lokalen Eigenschaften.

Die dritte Studie entwickelt einen Transfer Learning Ansatz für Regionen ohne Portfoliodaten. Es nutzt Sterblichkeitsdaten der Gesamtbevölkerung, kategorisiert nach Alter und Geschlecht, sowie detailliertere Sterblichkeitsdaten der versicherten Bevölkerung aus acht Ländern, um die fehlenden Informationen durch Korrelationen zu gewinnen. Das vortrainierte globale Modell aus der vorherigen Studie wird auf das Vereinigte Königreich übertragen, liefert Vorhersagen nahe den offiziellen Sterbetafeln und ermöglicht die Rekonstruktion von Sterblichkeitsstrukturen. Da in vielen Ländern solche offiziellen Sterbetafeln entweder nicht verfügbar oder nur stark aggregiert sind, sind sie für portfoliospezifische Anwendungen kaum nutzbar, insbesondere in Märkten oder bei Produkten ohne versicherungsspezifische Daten.

Die vierte Studie entwickelt ein Simulationsmodell, das Iterative Proportional Fitting verwendet, um synthetische Sterblichkeitsdatensätze zu erzeugen. Diese Simulationen ermöglichen die Modellierung für Bevölkerungsgruppen ohne detaillierte demografische oder versicherungsspezifische Daten. Die synthetischen Datensätze spiegeln realistische mehrdimensionale Verteilungen wider und ermöglichen präzise Modellierung über demografische Segmente hinweg.

Reproduzierbarkeit und praktische Umsetzung werden durch Open-Source-Software `R` und `Python` sichergestellt. Diese Arbeit trägt zur Sterblichkeitsmodellierung bei, indem sie methodische Innovationen mit realen Einschränkungen verbindet und die Grundlage für weitere demografische und versicherungsmathematische Fortschritte schafft.

Contents

Part I.

Introduction and background

1. Introduction

1.1. Outline

The focus of this doctoral thesis is on advanced methods of mortality modeling, a topic of significant importance due to its complexity and the multitude of factors influencing human mortality. Gaining a comprehensive understanding of mortality risk, both present and in the future, is crucial, especially in light of recent changes such as increased life expectancy due to medical advancements and the impact of extreme events like pandemics. These changes necessitate reliable planning and decision-making tools for pensioners, life insurers, and public health authorities. All methods are motivated by real-world applications in either mortality research or insurance science.
The remainder of this introductory chapter is organized as follows. This section gives a brief overview of the research questions covered by this dissertation and motivates their statistical relevance in the historical context. Sections 2 to 5 introduce the individual statistical problems and our contributions that are the core of this dissertation, summarize the utilized methodological approaches and potential alternative techniques, comment on the current state of statistical research in the respective field and shortly discuss potential directions for future research. Section 6 concludes the chapter by summarizing the key findings and offering an outlook on future research possibilities. An in-depth exploration of the individual core contributions made throughout this work is provided in Chapters II to IV. In the spirit of open science and to ensure full reproducibility, all contributions are accompanied by either direct implementations or a detailed, step-by-step algorithmic description of the methodological approaches in the statistical open-source software `R` (R Core Team, 2025) and `Python` (Python Software Foundation, 2025).

1.2. Motivation and scope

The risk associated with human mortality is driven by uncertainties and variables that cannot be fully captured by a single mathematical formula or explained by a single statistical model. Rather than modeling individual mortality as in limited clinical studies, the focus is on population-level mortality data. A multifaceted approach is crucial to estimate and predict mortality, thus this thesis introduces four innovative methods that utilize observed trends in several complex demographic and data availability settings. Longevity risk, defined as the risk that people live longer than expected, presents significant challenges for societies today. While increased longevity enhances productivity and welfare, it also imposes financial burdens on pension systems and public health expenditures, threatening the solvency of financial institutions due to unanticipated liabilities. This is particularly relevant in the context of pay-as-you-go (PAYG) systems, which face increased pressure as the ratio of working individuals to retirees shifts unfavorably. Events like the global pandemic further complicate mortality modeling. SARS-CoV-2 is not the first virus to cause a pandemic, nor will it be the last. Studies indicate that globalization and population growth

increase the likelihood of epidemics and pandemics, which can lead to sudden mortality shocks (Engel and Ziegler, 2020). They discuss how globalization has exacerbated the spread of infectious diseases, creating new challenges for mortality modeling. Woolhouse and Gaunt (2007) find that over 6% of all known human pathogens were first reported in humans after 1980, many of which are globally spreading viruses. Mortality modelers must consider whether their models should account for such shocks, which are crucial for pricing mortality catastrophe bonds and making risk management decisions in life insurance. Cairns et al. (2011b) highlight the need to incorporate these shocks into actuarial practices to ensure the financial stability of life insurance companies. To address these rapid changes and mitigate negative consequences, accurately assessing longevity risk is essential. Mortality estimation models must capture the effects of various factors driving changes in life expectancy. While achieving Laplace's vision of perfect predictability is unrealistic, approaching mortality from multiple perspectives enhances the validity of predictions (Laplace, 1825). One promising approach is to integrate traditional actuarial models with modern machine learning (ML) techniques to improve accuracy while maintaining interpretability. This work builds upon existing literature, which has evolved from simple mathematical calculations of life tables in the 18th century to stochastic models like the Lee-Carter (LC) model in the late 20th century. This model introduced a three-dimensional framework for mortality, which has since been expanded with further effects, such as cohort, and ML enhancements. Lee and Carter (1992) propose a log-bilinear model for mortality rates incorporating age and year effects, a framework widely used for long-run forecasts by institutions like the U.S. Bureau of the Census (Hollmann et al., 2000). Subsequent models, such as the Augmented Common Factor Model by Li and Lee (2005), impose coherence in mortality developments across populations, ensuring long-term stability in mortality ratios (Cairns et al., 2011b; Hyndman et al., 2013). Research has also explored the application of ML techniques. Perla et al. (2021) suggest using Convolutional Neural Networks (CNN) with one-dimensional convolutions, while Meier and Wuthrich (2020) employ CNNs with two-dimensional convolutions to detect anomalies in mortality data. Richman and Wüthrich (2021) advocate for the use of feed-forward neural networks to learn mortality intensities from historical data, highlighting the potential of ML to outperform traditional models. To maintain also interpretability, Richman and Wüthrich (2023) propose the LocalGLMnet, a more explainable deep learning model that attempts to maintain the interpretable structure of GLMs while enabling variable selection and interaction identification. Building on this, Perla et al. (2024) extend the LocalGLMnet to mortality grids, enhancing point forecast accuracy but without addressing forecast uncertainties needed by actuaries.

Open-source mortality data, such as that from the Human Mortality Database (HMD) (HMD, 2024), typically exist at aggregated levels like year, age, and gender. While these provide valuable insights into demographic trends, integrating socioeconomic, behavioral, and health variables collected at the individual level in insurance portfolios can enhance mortality analysis with more accurate risk assessments and predictive modeling, while also considering data privacy and protection. Therefore, the proposed models are inherently multipopulational, allowing for the inclusion of diverse variables. A significant body of research assumes that death counts follow a Poisson distribution (Brouhns et al., 2002; Booth et al., 2002; Delwarde et al., 2006). This approach offers the benefit of eliminating the unrealistic assumption of error homoskedasticity present in models like the LC model by assigning greater importance to age groups with higher mortality counts. However, the Poisson model has the drawback of assuming that the expected number of deaths is equal to its variance. Empirical observations often show that the variance is larger, a phenomenon known as overdispersion, which is common in insurance portfolios. This can be addressed by intro-

ducing additional model parameters (Renshaw and Haberman, 2003) or by adopting alternative distribution assumptions, such as the negative binomial distribution (Renshaw and Haberman, 2008). Since determining the optimal distribution for death counts is not the primary focus of this work, we will consistently use the Poisson assumption, which has been extensively validated for mortality modeling both theoretically and empirically by Brillinger (1986). Consequently, all models assume that the observed number of insured deaths D_i follows a Poisson distribution: $D_i \sim \text{Poisson}(\mu_i \cdot E_i)$. Here, μ_i represents the mortality rate for demographic index i, where i denotes a unique combination of demographic characteristics specific to the dataset used for each study and problem. E_i is the exposure for demographic index i, reflecting the population at risk or person-years at risk, which scales the mortality rate.

The data accumulates slowly, particularly in life insurance where death events are infrequent. This varies by company, country, and portfolio, resulting in data-rich and data-poor subgroups. Therefore with this thesis we aim to provide innovative solutions for demographers and actuaries, addressing the complexities associated with mortality modeling and forecasting. It seeks to enhance the estimation of mortality rates and actuarial life tables proposing novel advanced techniques, while evaluating the applicability of these approaches across diverse settings. Thus the focus is on the multifaceted approach, the integration of traditional and modern techniques, the importance of open-source data, and the specific contributions of the thesis in addressing practical challenges.

In contrast to other studies that presuppose data availability, this research prioritizes special data situations, such as extreme events, regional dynamics and cross-country perspectives to establish a solid data foundation for modeling. By utilizing insights from data-rich regions, we aid areas with limited or no data, effectively tackling computational challenges while improving model accuracy and maintaining interpretability. Thus we build upon various historical advancements in mortality modeling, including both single and multipopulation models, with the goal to not only achieve high predictive accuracy but also to tackle practical challenges beyond mere prediction quality. The following four methodological contributions illustrate how this thesis addresses these challenges:

1. Modeling extreme events with Generalized Additive Models (GAMs) in an Age-Period-Cohort (APC) framework. COVID-19 effects are integrated into future mortality projections to capture long-term impacts and quantify mortality shocks reliably. The proposed models dynamically adapt to new data and account for abrupt changes in mortality patterns, even at the boundaries of the observed time series.

2. Regional mortality estimation through hierarchical Gradient Boosting Machines (GBMs). A hierarchical modeling strategy enables data-rich countries to support those with limited mortality data or computational resources. The approach combines data-sharing and collaborative modeling mechanisms to integrate multi-source information while preserving local population characteristics.

3. Transfer learning for data-scarce regions. In settings where direct mortality data are unavailable, with the example of the UK, transfer learning techniques adapt pretrained models from other populations to estimate mortality risks. This facilitates accurate, context-sensitive predictions even under severe data limitations.

4. Segmented mortality simulation using Iterative Proportional Fitting (IPF). To differentiate between population segments, such as insured individuals and the general public, IPF-based simulations generate realistic, granular mortality datasets for flexible feature extensions and tailored analyses across demographic groups, providing a foundational data infrastructure for research.

1.3. Historical context

Mortality modeling has deep roots in actuarial and demographic practice. The formal study of human mortality began with the construction of life tables in the 17th century. Graunt (1662) was the first to aggregate demographic records and demonstrate that, although individual lifespan is random, mortality in large populations follows systematic age patterns. Halley (1693) built on this insight by constructing an empirical life table from parish records of births and deaths, and used it to compute annuity values. Over the next two centuries, mathematicians devised theoretical mortality laws to simplify life insurance calculations. For example, de Moivre (1725) assumed a uniform distribution of deaths between ages 0 and a limiting age, allowing simple annuity formulas. Gompertz (1825) proposed that adult mortality increases exponentially with age, and Makeham (1867) added a constant term (the "Makeham term") to Gompertz's law to capture age-independent background mortality. The resulting Gompertz-Makeham model thus became a classical two-parameter model of adult human mortality, though it is static and cannot model calendar-time improvements on its own. By the late 19th century, actuaries began fitting more flexible formulas to life tables yielding in even more generic parametric curves. Early contributors included Danish mathematicians such as Thiele (1871) and Oppermann (1870) who developed actuarial models, though these ideas circulated slowly due to language barriers and lack of interdisciplinary collaboration (Hoem, 1983).

Classical parametric laws and complex models

Through the late 19th and 20th centuries, actuaries and demographers proposed numerous parametric "laws" to capture the age pattern of mortality. Well-known parametric curves include Beard's logistic formulation (Beard, 1959). In addition, model life tables, e.g., Coale-Demeny, UN model tables, were used for population forecasts. To handle complexity, multi-parameter hazard models were developed. For example, Siler (1983) introduced a five-parameter competing-risk model that explicitly describes three mortality phases (immaturity, accident hump and senescence) by separate location and dispersion parameters. Heligman and Pollard (1980) later proposed an eight-parameter model that fits the entire lifespan by summing three components (childhood, accident, adult). Empirical comparisons found that a logistic form often best describes adult mortality at extreme ages: Beard (1971) and Thatcher et al. (1998) showed that a logistic hazard model outperformed the Gompertz-Makeham model for data above age 80.

Table 1.1 gives an overview of selected parametric models. Each of them have advantages, such as closed-form and interpretability, but also rigidity: they assume fixed shapes and cannot easily adapt to changing mortality trends or structural breaks. These limitations motivated the move towards stochastic and nonparametric methods in the late 20th century.

Table 1.1.: Overview of selected landmark mortality laws

Law	Mathematical formulation
Gompertz-Makeham (1825/1867)	$\mu(x) = Ae^{Bx} + C$, $\mu(x)$: force of mortality at age x, A: baseline mortality, B: exponential age effect, C: constant background risk.
Siler's Hazard (1983)	$\mu(x) = a_1 e^{-b_1 x} + a_2 + a_3 e^{b_3 x}$, models infant (a_1, b_1), midlife (a_2), and senescent (a_3, b_3) mortality components.
Heligman-Pollard (1980)	$\frac{q_x}{1-q_x} = A^{(x+B)^C} + De^{-E(\log x - \log F)^2} + GH^x$, q_x: probability of death at age x, 8 parameters A-H capture child, accident, and old-age mortality.
Beard-Thatcher (1959/1998)	$\mu(x) = \frac{KAe^{Bx}}{1 + Ae^{Bx}}$, logistic, K is the asymptotic mortality limit, models mortality deceleration due to population frailty.

Stochastic mortality models

With advances in data availability and computing power, the focus shifted to stochastic models that incorporate calendar-time dynamics. A landmark contribution was the LC model (Lee and Carter, 1992). They represented the log of the central death rate $\mu(x,t)$ at age x and year t as

$$\log \mu(x,t) = a(x) + b(x)\, k(t) + \varepsilon(x,t) \tag{1.1}$$

where $a(x)$ is the average log-mortality by age, $b(x)$ is the age-specific sensitivity to changes in the mortality index $k(t)$, and $\varepsilon(x,t)$ is a mean-zero error. In practice, $a(x)$ and $b(x)$ are estimated from historical rates, for example by singular value decomposition, and the time index $k(t)$ is fitted as a univariate time series, typically a random walk with drift. This one-factor model often explains 80-95% of past mortality variation and became a standard forecasting benchmark. For example, agencies such as the U.S. Social Security Administration have used LC projections of mortality to forecast life expectancy (Lee and Miller, 2000). Because $k(t)$ is stochastic, LC yields probabilistic forecasts for future death rates. Its simplicity also spurred many extensions: Renshaw and Haberman (2006) added a cohort (birth-year) effect, multi-factor versions include additional principal components, and state-space or Bayesian formulations have been proposed to improve estimation and capture uncertainty. Another influential model is the Cairns-Blake-Dowd (CBD) Cairns et al. (2006). The CBD model targets older ages (typically age 60 and above) and uses a logistic (logit) link for death probabilities. In its simplest form it posits

$$\operatorname{logit} q_x(t) = \kappa_1(t) + \kappa_2(t)\,(x - \bar{x}) \tag{1.2}$$

where $q_x(t)$ is the one-year death probability at age x, and $\kappa_1(t), \kappa_2(t)$ are time-varying intercept and slope. This linear-in-age form captures the roughly linear increase of logit-mortality with age

in late life, while allowing the level and slope to evolve. Because of its focus on retirement ages, the CBD model is popular in pension risk modeling and has fewer parameters than LC. Other models, e.g. Plat's model, combine features of LC and CBD to cover wider age ranges (Plat, 2009). These stochastic models have become the basis for forecasting mortality with quantified uncertainty in demography and actuarial science.

Multi-population models and coherence

Over the past two decades, multi-population models have become a key advance to ensure coherent behavior and share common components across groups, such as males and females, or different countries. Naive independent forecasts can imply implausible divergence of trends, for example, predicting that male mortality will fall so fast that male life expectancy surpasses female. Li and Lee (2005) proposed the Augmented Common-Factor model. In this approach, multiple populations share a single mortality index estimated from pooled data, plus population-specific offsets. This enforces a common long-term trend while allowing group differences. Hyndman et al. (2013) proposed the product-ratio method, which forecasts the average mortality curve of all populations and then models each population's deviation (ratio) from this mean, ensuring that forecasts remain coherent. Raftery et al. (2014) developed a Bayesian coherent model for male/female mortality: first projecting female longevity and then using a hierarchical model for the historical gap to produce joint probabilistic forecasts. Other multi-population approaches include multivariate extensions of single-population models. Jarner and Kryger (2011) and Cairns et al. (2011a) formulated multivariate LC models allowing two populations to co-evolve. These models jointly fit all populations, borrowing strength across groups and capturing correlations in mortality improvements. In general, coherent multi-population model tend to improve accuracy, especially when some populations have sparse data, and ensure that forecasted mortality curves move together in a realistic way.

While coherence prevents unrealistic crossings of trajectories, it comes with trade-offs. Technically, coherence forces age-specific mortality ratios between populations to converge to fixed constants. While this ensures aggregate measures, like life expectancy differences, stay bounded, it can sometimes conflict with data. For example, imposing a converging gender ratio trend may not be empirically supported and can yield implausible future gender-gap dynamics. This points to a need for models that adapt to changing cross-population dynamics without arbitrary constraints.

Machine learning approaches in mortality modeling

In the last decade, ML and Artificial Intelligence (AI) techniques have begun to influence mortality modeling by introducing new approaches. ML methods can ingest large or complex datasets and capture nonlinear interactions that classical models may miss.

Ensemble tree methods, including algorithms such as random forests and gradient boosting have been applied to mortality tables. These models can flexibly fit irregular age-period patterns. Deprez et al. (2017) and Levantesi and Pizzorusso (2019) used random forests to model historical mortality data, finding that ML models captured nonlinear age effects and interactions not captured by LC, yielding lower in-sample error. Gradient boosting machines have similarly been employed. These methods are powerful for complex patterns but require careful tuning to avoid overfitting in limited datasets.

Deep learning models, particularly recurrent neural networks (RNNs), have been explored for time-series mortality data. Empirical comparisons show that these networks can forecast mor-

tality effectively and suggest that RNNs can learn cohort and period patterns without explicit modeling, though they typically require long training histories. For example, Chen and Khaliq (2022) compared Long Short-Term Memory (LSTM), bidirectional LSTM (BiLSTM), and Gated Recurrent Unit (GRU) models to the LC model using U.S. mortality rates (1966-2015), and found that all deep learning models achieved comparable or better forecast accuracy. Similarly, Nigri et al. (2019) applied RNNs to several national mortality series with encouraging results.

Transformer and hybrid models, such as attention-based neural architectures, e.g. the Temporal Fusion Transformer, can incorporate additional covariate time series. These models can learn relationships between mortality and external factors, such as economic, environmental, etc. Postema and van Es (2022) trained a Temporal Fusion Transformer on multi-country, age-specific mortality data augmented with socio-economic covariates, and reported improved forecast accuracy over standard models, such as the LC model and its extensions. More generally, hybrid models and ensembles, combining ML and traditional models, have been studied, often yielding more precise forecasts by blending strengths of different methods, while compromising aspects of interpretability. A key advantage of ML is the ability to incorporate new data sources. For example, features from electronic health records, insurance claims, wearable devices, or even social media and search trends analyzed via Natural Language Processing (NLP) can be used to enrich mortality models. Wang (2024) discusses how AI can analyze large-scale health data streams to identify emerging risk factors, e.g. novel diseases or lifestyle changes, that may affect mortality. Including such high-dimensional inputs, ML models may adapt forecasts based on real-time health indicators and emerging threats. These ML-based approaches have demonstrated promising performance in retrospective studies, also suggesting that ML can be a useful complement to traditional actuarial models when rich data are available.

Despite these advances, integrating ML also raises new challenges. Many traditional models cannot accommodate sudden shocks, such as COVID-19, as they assume constant trends. Data sparsity in subpopulations limits the reliability of direct model fitting. Coherent models sometimes overconstrain gender or regional trends, conflicting with observed divergences. Furthermore, many ML models struggle with regulatory requirements in actuarial practice and lack interpretability, as we discuss next.

The challenge of model interpretability and transferability

As ML techniques become more prevalent in mortality modeling, the challenge of interpretability and transferability becomes increasingly important. While ML models offer powerful tools for capturing complex patterns, their opaque nature can hinder understanding and acceptance in actuarial practice. A primary concern is interpretability. Actuaries and regulators typically require transparency about how models make predictions. Traditional models have parameters with clear demographic meaning, e.g. age curvature or time trend, whereas ML models and deep nets are often considered as "black boxes". As Rajendran (2024) notes, AI-driven mortality forecasts can be opaque, making it hard to explain why a certain projection was made. This lack of transparency can undermine stakeholder confidence. To mitigate this, practitioners may apply explainable AI techniques, such as feature importance measures or partial dependence plots, to interpret complex models and ensure they align with expert judgment. Transferability and data bias are another concern. A model trained on one population or period may not generalize to another. For example, a neural net trained on pre-2020 data may fail to anticipate mortality shifts during or after the COVID-19 pandemic. Different countries or demographic groups can have vastly different mortality determinants, health systems, behaviors, so an ML model blindly calibrated on

one population may mis-predict another. Furthermore, ML models can inadvertently perpetuate data biases: if certain groups, e.g. ethnic minorities, rural communities, are underrepresented in the training data, the model's predictions for those groups may be biased. Actuarial practice thus demands carefully validated frameworks for models across subpopulations and adjusting training data. In practice, this often means retraining or recalibrating models for each target population or scenario, and combining multiple models to hedge against misspecification.

The challenge of complex data situations

Real-world mortality data rarely meet the assumptions of classical stochastic mortality models. In actuarial and demographic applications, data complexity arises from several sources: limited sample sizes (e.g., small countries or subnational regions), short or interrupted time series, unobserved covariates, reporting errors, and structural heterogeneity across populations. These issues pose serious challenges for model calibration, identifiability, and predictive reliability.

In small or developing countries and in niche insurance portfolios, mortality experience may be insufficient to support stable parameter estimation. While Bayesian hierarchical models and multi-population frameworks (Li and Lee, 2005; Cairns et al., 2011a; Dowd et al., 2020) allow for partial pooling of information, their empirical application remains largely confined to settings with relatively complete and homogeneous data. Evidence on their effectiveness under extreme data sparsity or fragmented information is still limited.

In related statistical and ML fields, transfer learning and domain adaptation have shown promise for knowledge sharing across data-rich and data-poor contexts (Pan and Yang, 2009; Weiss et al., 2016). However, these approaches have not been systematically explored for mortality data, where age, cohort, and temporal structures impose constraints that differ fundamentally from other predictive domains. The absence of studies applying transfer learning within actuarial or demographic modeling highlights a significant methodological gap.

Similarly, fragmented and aggregated mortality datasets, where only marginal information by age or gender is available, are typically treated with basic smoothing or interpolation. Although IPF and synthetic population generation methods are well established in survey statistics and transportation modeling (Agresti, 2012; Beckman et al., 1996; Lomax and Norman, 2016), their systematic evaluation for reconstructing joint mortality structures is virtually absent in the demographic literature.

Population heterogeneity across countries or subgroups further complicates mortality modeling. While multi-population approaches (e.g., Renshaw and Haberman, 2006; Cairns et al., 2011a) partially address this issue, they often assume consistent structural relationships or proportional trends, which are rarely verified empirically. There is limited historical evidence on how to quantify or propagate structural heterogeneity in mortality forecasts, particularly when combining actuarial and demographic populations.

Finally, extreme events, such as pandemics, wars, and environmental crises, introduce nonstationarity and structural breaks. While changepoint and regime-switching models have been proposed in other time-series domains (Robben et al., 2022; Shiferaw, 2021), there remains no unified framework for integrating sudden mortality shocks into long-term stochastic projections. The Short-Term Mortality Fluctuations (STMF) database (Human Mortality Database, 2024) has only recently enabled such analyses, and empirical methods remain fragmented.

Overall, while these individual statistical and computational techniques exist across disciplines, the literature lacks a coherent, empirically validated framework for handling complex data situations in mortality modeling. There is no historical evidence of an integrated approach that simultaneously addresses data sparsity, aggregation, heterogeneity, extreme shocks, and missingness within a single modeling paradigm. Filling this gap constitutes the central motivation and contribution of this dissertation to that evolving literature by synthesizing such complex structures using a combination of flexible modern statistical models and ML tools tailored for heterogeneous populations.

Application perspectives in public health and financial planning

Forecasts of future mortality and life expectancy are critical for public health and social policy planning. Governments and international agencies use projected death rates to anticipate needs for healthcare, pensions, and long-term care. For example, mortality projections feed into national population forecasts, such as those produced by the United Nations (2022) or World Health Organization (2024), that estimate the number of elderly and dependent individuals over time. These demographic projections influence decisions on retirement age, pension funding, and healthcare infrastructure. In particular, rapidly aging populations worldwide, given for example UN projections, which show that the population over age 60 will nearly double by 2050, make accurate mortality forecasts essential for avoiding resource shortfalls (World Health Organization, 2024). Mortality modeling also underpins evaluations of public health interventions and disease burden. For instance, forecast models for cause-specific mortality help estimate how many future deaths could be prevented by medical advances or prevention programs. During pandemics and health crises, mortality forecasting is indispensable. In the COVID-19 pandemic, for example, analysts compared observed deaths to model-based baselines to estimate "excess mortality" and assess the pandemic's true toll. By quantifying expected versus actual deaths, these models informed public health responses, including vaccine rollout, lockdown policies, and assessed the impact on life expectancy thus testifying that sophisticated mortality forecasts enable policymakers to allocate medical and social resources efficiently and to evaluate the potential impact of health policies.

These limitations are not merely theoretical. They directly impact high-stakes domains such as public health resource allocation, and insurance pricing. Therefore accurate and adaptive mortality forecasts are essential for managing long-term population risks today.

In the insurance and pension industries, mortality models directly affect financial decision-making. Life insurers and annuity providers use projected mortality rates to price products and set technical reserves for future payouts. Pension funds use forecasts of future lifespans to calculate liabilities for retirement benefits and to set contribution rates. Regulatory and accounting frameworks explicitly incorporate longevity risk. Under Solvency II (in Europe) and the global Insurance Capital Standard, insurers must account for adverse longevity scenarios when determining capital requirements. For example, standard shocks (on the order of 10-15% additional life expectancy) are prescribed to test solvency under faster mortality improvement of Insurance Supervisors (IAIS). Similarly, IFRS 17 accounting requires companies to regularly update their best-estimate mortality assumptions and to disclose the uncertainty of those assumptions (IASB). In practice, insurers use stochastic mortality models, such as LC, CBD or more advanced frameworks, within their asset-liability models and scenario generators. More accurate and flexible modeling of mortality thus reduces the risk of underestimating future payouts or over-reserving capital, directly impacting an insurer's profitability and solvency.

1.4. Notation

The following table summarizes the relevant notations used throughout this thesis. While not exhaustive, it includes the most frequently used symbols and terms essential for understanding the methodology and results.

i	Flat index identifying unique demographic groups in aggregated datasets (e.g., age, period, gender, country).
μ_i	Mortality rate for demographic group i
D_i	Observed number of deaths in demographic group i
E_i	Exposure (population or life years at risk) for demographic group i
$f_{ap}(a, p)$	Smooth tensor-product spline surface over age a and period p, used in GAMs
β_0	Intercept term in regression models
$covid_i$	Binary indicator (0/1) for pre-/post-COVID years, used in pandemic modeling
$q(\cdot)$	Prediction function from the global model in hierarchical boosting
$h_j(\cdot)$	Country-specific local model prediction function for country j
$X_{i,j}$	Feature vector for group i in country j
X_i^{global}	Subset of X_i containing global features shared across countries
$X_{i,j}^{\text{local}}$	Country-specific local features for country j
θ_k	Coefficient or learning rate of the k-th weak learner in boosting
$u_k(X)$	Base learner applied to feature set X in the k-th boosting step
δ_i	Drift adjustment factor between transferred and expected mortality
$D_{i,M}^{\text{pretrained}}$	Predicted deaths in the target country M using the pretrained global model
$D_{i,M}^{\text{specialized}}$	Refined predictions in target country M using country-specific adaptation

2. GAM in Semiparametric APC for Cross-Country COVID-19 Forecast

2.1. Research question

Firstly, we tackle the complexities brought about by the pandemic by comparing traditional actuarial models with modern ML techniques. The study introduces an enhanced cross-country coherent framework for modeling and multipopulational projection, specifically focusing on the effects of COVID-19 on mortality forecasting.

The primary research question addressed is how to effectively capture the mortality-related extreme event at the edge of a time series, specifically the impact of COVID-19 on future mortality forecasting. This involves developing a reliable model and detailed scenarios that are crucial for policymakers and stakeholders, providing guidance in navigating the uncertainties brought about by the pandemic, especially focusing on life insurance and pension funds, sectors significantly affected by mortality trends. Existing research in mortality modeling spans from traditional stochastic models to ML and neural networks. While these approaches deliver high predictive power, they typically demand strong computational resources and may not offer the interpretability and ability to measure confidence that traditional models provide. This is particularly true for applications in COVID-related studies, where complex data situations often fail to adequately evaluate or explain the factors influencing mortality trends (Ioannidis, 2021). Our goal is to strike a balance between these aspects.

2.2. Literature review and research gap

The literature presents a diverse range of approaches to mortality modeling that can be used for extrapolation. Traditional stochastic models, as discussed previously, provide a foundational framework for mortality forecasting. Recent studies have demonstrated the efficacy of ML, such as GBM and Random Forest, in enhancing predictive performance (Richman and Wüthrich, 2023). Hierarchical approaches building upon simpler LC models have also gained traction, offering improved forecasts (Li and Lee, 2005; Cairns et al., 2011b). The COVID-19 pandemic has exposed the limitations of conventional mortality models that were primarily designed under assumptions of gradual demographic transitions and stable trends. Classical actuarial models typically extrapolate historical mortality patterns and therefore lack mechanisms to accommodate abrupt shocks. Neural networks, including the Common-Age-Effect Model (Kleinow, 2015), have been applied to mortality modeling, providing enhanced accuracy but often lacking interpretability. GAMs, introduced by Hastie and Tibshirani (1990), offer a flexible framework for mortality analysis. Recent advancements include Bayesian APC models with autoregressive priors (Clements et al., 2005)

and bivariate spline functions within GAMs (Weigert et al., 2021, 2022). The literature on accounting for COVID-19 in mortality projections is growing, with studies using stochastic models and polynomial basis decompositions (Barigou et al., 2020; Robben et al., 2022; Schnürch et al., 2022). While these models address various aspects of mortality forecasting, none have extrapolated GAMs within the APC framework in a multi-populational context to evaluate post-pandemic scenarios.

Overall, the literature indicates a methodological divide between interpretable, theory-based models and flexible, data-driven algorithms. While traditional stochastic models offer transparency and coherence, they often fail to capture sudden, large-scale disruptions such as the COVID-19 pandemic. Conversely, ML methods can flexibly fit complex patterns but lack interpretability and formal uncertainty quantification. Few studies have yet achieved a coherent integration of these approaches capable of producing reliable long-term mortality scenarios under extreme shocks. Addressing this gap, particularly through frameworks that support scenario design, uncertainty assessment, and communication of model confidence, remains a central challenge for actuarial science and longevity risk management.

Therefore, the specific contribution of the GAM approach within the APC framework in this research lies in its novel application to mortality trend forecasting, in particular, its use to incorporate the impact of COVID-19 in a cross-country, multipopulation context as well as the evaluation of it in a post-pandemic fashion. Our research demonstrates the reliability of GAMs within the APC framework and their capacity to extrapolate mortality forecasts, compared to other models, accounting for the impact of the COVID-19 pandemic. GAMs address the identifiability issue inherent in APC analysis by circumventing the linear dependency between age, period, and cohort effects. This semiparametric approach estimates a two-dimensional interaction surface, representing all three temporal dimensions, using a tensor product spline basis. This methodology implicitly deals with the linear dependency of temporal dimensions, offering a nonlinear, highly flexible approach.

2.3. Contribution and prospects

Observed developments of a process of interest can be associated with a person's life cycle (age effect), changes affecting the whole population over some time period, like macro-economic developments or scientific progress in medicine (period effect), or structural differences between members of different generations like socialization or exposure processes (cohort effect). Nowadays, such research questions appear frequently not only in demographic or epidemiological contexts but also in economic, social, or general medical sciences. Flexible and precise approaches are required both for the estimation and visualization of nonlinear associations with said temporal dimensions.

The central challenge for statistical approaches in APC analysis is to circumvent the identification problem, which describes the linear dependency of the temporal dimensions. Many regression-based approaches solve this problem by estimating linear effects while putting hard constraints on specific parameters. When trying to differentiate between the individual temporal dimensions, statistical methods have to account for their linear dependency:

$$\text{cohort} = \text{period} - \text{age} \tag{2.1}$$

Due to this identification problem, a perfect separation of the temporal effects is not possible. Still, different techniques were developed that aim to circumvent this issue. Descriptive approaches typically focus on jointly visualizing all three temporal dimensions, usually building on the concept of Lexis diagrams where age and period groups are depicted in horizontal and vertical direction, respectively, so that individual cohorts are represented along the diagonals (Carstensen, 2007). To estimate individual association structures with the temporal dimensions, we utilize a regression model-based approach. Weigert et al. (2021) outlines a semiparametric approach based on the estimation of a two-dimensional interaction surface to represent all three temporal dimensions. Following Clements et al. (2005), we estimate a GAM with the following structure: $g(\mu_i) = \beta_0 + f_{ap}(\text{age}_i, \text{period}_i) + \eta_i, \quad i = 1, \ldots, n$ with observation index i, μ_i the expected value of an exponential family response, link function $g(\cdot)$, and the intercept β_0. The interaction surface $f_{ap}(\text{age}_i, \text{period}_i)$ is represented by a two-dimensional tensor product spline basis based on two marginal P-spline bases. η_i represents an optional linear predictor that contains further covariates. In contrast to alternative approaches, which are often based on the estimation of linear parameters under specific hard constraints, this semiparametric estimation approach has the main benefit that the linear dependency of the temporal dimensions is dealt with implicitly. Instead of using explicit constraints, a nonlinear, highly flexible approach is used where one temporal dimension (typically cohort) is naturally represented as the interaction of the other two dimensions (typically age and period).

The study relies on the Human Mortality Database (HMD, 2024), which provides yearly mortality rates μ_i, death counts D_i, and population sizes E_i for selected countries. These are categorized by a demographic index i, where each i corresponds to a unique combination of age (a), year (t), gender (g), and country (j). The selection of countries is based on their geographic context and the contrasting impact of the COVID-19 pandemic. Specifically, Finland (until 2019), Germany (until 2017), Italy (until 2018), and the Netherlands (until 2019) represent Europe, while the United States (until 2019) represents North America. For the most recent years up until 2023, the STMF series offers partial information on mortality rates and population, grouped by week and age buckets. However, the study requires data on a yearly basis and in a metric age scale. To address this, we employ a methodology to construct mortality data on a yearly basis and in a metric age scale, which are then combined with the original HMD data (Li and Lee, 2005).

For forecasting purposes, the impact of COVID-19 on mortality rates is incorporated into the mortality model using an additional variable. This variable is specific to each country c and takes the value of 0 for years until 2019 and 1 for the years 2020 and 2021, representing the period during which the COVID-19 pandemic has proven to have a strong impact. Under assumption of a Poisson distribution for death counts

$$D_i \sim \text{Poisson}(\mu_i \cdot E_i),$$

this leads to the final model specification:

$$\log(\mu_i) = \beta_0 + f_{g,j}(\text{age}_i, \text{time}_i) + \beta_{c,j} \cdot (covid \times country)_i + \log(E_i)$$

.

$\log(\mu_i)$ is the logarithm of the expected mortality rate for demographic index i, where i represents a unique combination of demographic characteristics, including age, gender, country, and year. β_0 it

the intercept term in the model, representing the baseline level of mortality rate on the logarithmic scale. $\beta_{c,j} \cdot (covid \times country)_i$ is the interaction term that captures the country-specific impact of the COVID-19 pandemic on mortality rates. The interaction between binary COVID-indicator c and country j allows the model to consider variations in how different countries experienced the pandemic during the years 2020 and 2021, and project these effects into the future. $\log(E_i)$ is the logarithm of the exposure for demographic index i, representing the population at risk or person-years at risk, which scales the mortality rate. The spline function $f_{g,j}(\text{age}_i, \text{time}_i)$ is a function that captures the bivariate tensor effect of age and time on mortality rates, varying by gender g and country j. It can be represented using basis functions: $f(\text{age}_i, \text{time}_i) = \sum_{k=1}^{K} \alpha_k B_k(\text{age}_i, \text{time}_i)$ with basis functions $B_k(\text{age}_i, \text{time}_i)$, which form the building blocks of the spline. α_k are the estimated coefficients for the basis functions, which determine the contribution of each basis function to the overall spline for each k-th basis function.

The values taken for future predictions are subject to the scenario assumptions that offer insights into public health strategies and their implications for mortality modeling during the pandemic: Scenario 1 (no COVID effect) models the situation assuming the virus has no impact, providing a baseline understanding of mortality without the influence of the pandemic. Scenario 2 (COVID full effect) examines the effects of the virus assuming no intervention, highlighting the potential spread and impact on the healthcare system. Scenario 3 (flattening COVID effect) involves measures aimed at reducing the virus's impact, such as social distancing and partial lockdowns, aiming to control transmission while maintaining some social and economic activity. Lastly, Scenario 4 (excess mortality) assesses the impact of the pandemic on mortality rates beyond typical expectations, considering factors like healthcare system strain and indirect consequences of the virus.

First, we compare the predictive performance of four models, conventional models such as LC and APC; a contemporary two-step approach utilizing GBM; and, lastly, the proposed GAM integrated within the APC framework, for modeling and projecting future mortality rates for five countries. We find that the most promising approach is based on a GAM, where cohorts are represented as an interaction between age and period, performing noticeably better than alternative methods in capturing complex mortality dynamics. This framework, adaptable for both aggregated and individual survival data, introduces a state-of-the-art method for the field of multi-populational cross-country mortality research. GAM-APC is the most effective approach for mortality forecasting: It has superior forecast performance, is computationally efficient, and allows interpretability. Furthermore, this work is the first to apply partial APC plots in mortality research, providing useful visual insights into APC effects. This tool aids in communicating complex temporal patterns and highlights gender-specific and cross-country differences as shown in Figure 2.1. The framework can be extended to further features, such as socio-economic variables, and can be used for interpreting each contributing driver to mortality predictions.

Finally, we provide insights into the factors driving the impact of the COVID-19 pandemic on mortality for the five countries. Through analyzing age, period, and cohort associations in a multi-population context using a GAM within an APC framework, we extrapolate mortality rates into the future. Four scenarios, representing varying degrees of pandemic impact, are evaluated against observed mortality data post-pandemic to identify the most accurate scenario. Among the four pandemic impact scenarios, Scenario 3, which assumes a diminishing impact of COVID-19, is the most consistent, particularly for middle-aged and elderly populations. In general, when evaluating the future scenarios based on the two years of the "future" set that we meanwhile

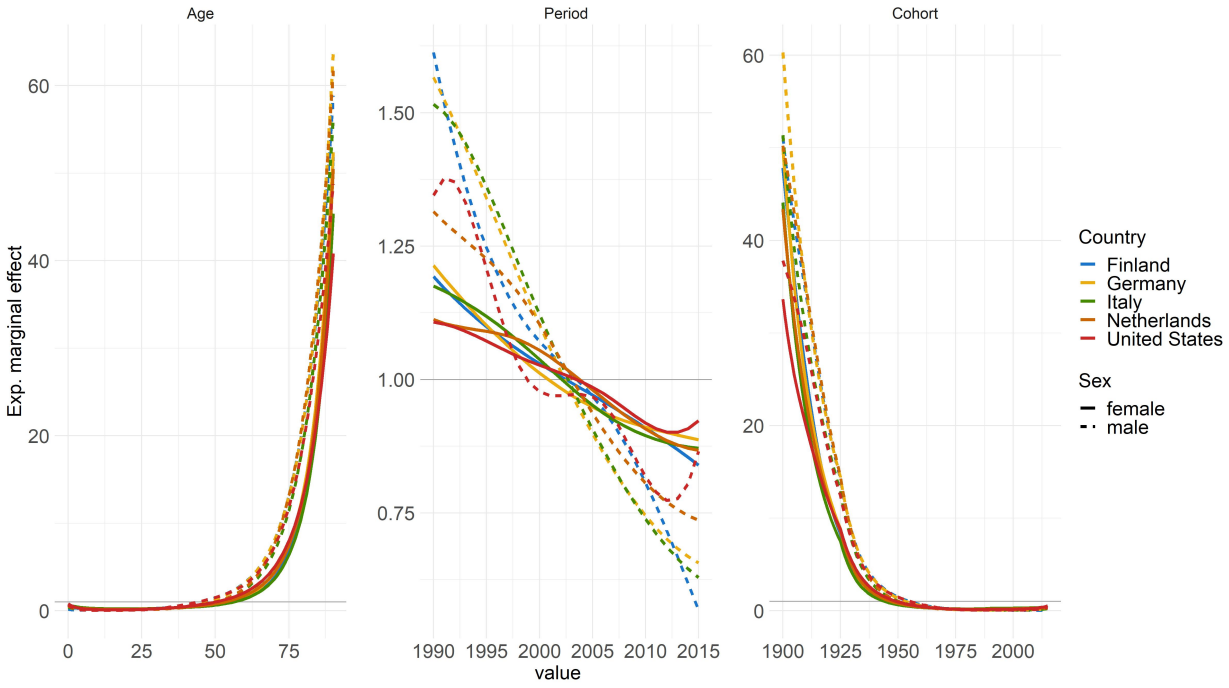


Figure 2.1.: Estimated marginal effects of age, period, and cohort on mortality rates, differentiated by countries and genders. The horizontal lines represent the level of no effect. The GAM model was fitted for the years 1990–2015 and ages 0–90.

have the observed data for from STMF, Scenario 3 (flattening COVID effect) tends to perform well for middle-aged individuals and, in addition, also Scenario 2 (COVID full effect) for older ages. Substantial variations in scenario performance are observed across different countries and age groups. For younger age groups, Scenario 1 (no COVID effect) performs best in Italy and the Netherlands, where substantial COVID impact was observed. Middle-aged groups demonstrate similarly high performance across all scenarios. Older age groups show stronger scenario differences with a clear preference for Scenarios 2 and 3, indicating a better fit. Scenarios do not perform well for those under 19, possibly due to the unique characteristics and weak impact of COVID in this age group.

Future research may explore seasonality in the HMD STMF data, examine the impact of socioeconomic variables, and extend extrapolation with GAMs over longer time ranges. Adapting scenarios to new developments and societal attitudes will enhance the goodness of mortality forecasts. The formulation of potential future scenarios related to the pandemic is a complex undertaking, shaped by a multitude of factors, including political decisions and societal acceptance. Despite the inherent challenges, these presented scenarios offer a valuable foundation for mortality forecasts, taking into account the evolving attitudes of life insurers and contributing to the ongoing discourse surrounding the impact of the pandemic.

3. Regional Mortality Support with Hierarchical Boosting

3.1. Research question

Mortality rates exhibit considerable variation not only across countries but also within regions, revealing pronounced subnational disparities in life expectancy. These differences challenge the actuarial fairness of annuities and public pension schemes (Sánchez-Romero et al., 2019). Reliable regional mortality projections are therefore essential for both insurers and policymakers. However, such projections are frequently constrained by limited or inconsistent subnational data, particularly in developing countries and within insurance portfolios, where mortality records are often sparse or incomplete. The heterogeneous and dynamic nature of regional data availability necessitates a reliable modeling framework that can flexibly adapt to diverse contexts.

The central research question guiding this study is: *How can regional and subnational mortality be modeled accurately in data-scarce environments while balancing global patterns with local variations?* This question holds practical significance for enhancing the equity of pension systems and improving the pricing fairness of global life insurance products across heterogeneous populations.

Existing mortality models often struggle to reconcile global and local dynamics, resulting in frameworks that either overgeneralize or fail to capture region-specific characteristics. To address this gap, the study proposes a hierarchical modeling framework based on Poisson-distributed Gradient Boosting Machines (GBMs). This approach integrates local and global data sources to enhance predictive accuracy, improve computational efficiency, and provide accurate handling of missing or incomplete mortality information.

3.2. Literature review and research gap

Bayesian hierarchical models (Congdon, 2007) and generalized linear models (GLMs) (Roux, 2002) have long been used for multilevel modeling in mortality and epidemiology. These approaches typically rely on random effects to capture unobserved heterogeneity across populations. More recent studies have extended such frameworks with ML models, including random forests and gradient boosting methods, for mortality prediction (Yakovyna et al., 2024). While these modern approaches provide state-of-the-art predictive performance, they often lack interpretability and theoretical transparency, which are crucial for actuarial and demographic applications.

Despite these advances, current models face persistent challenges. They often suffer from overdispersion (Van Buuren and Van Buuren, 2018) and computational inefficiency when processing large or incomplete datasets containint missing data (Banerjee et al., 2003). Mortality data availability

varies widely across regions, especially at subnational levels, which limits model accuracy and comparability (Murray et al., 2007). Moreover, few studies fully exploit the hierarchical structure of mortality data, where regional or demographic layers can be leveraged to improve predictive stability, leaving a gap between global modeling and local relevance.

To address these issues, this study introduces a hierarchical gradient boosting framework inspired by knowledge distillation. The approach leverages global mortality patterns from data-rich regions to inform predictions in data-scarce areas while preserving regional specificity. It employs sequential residual learning and multiplicative retraining to refine local estimates across diverse feature sets. Building on the strengths of Poisson regression for modeling count data (Cameron and Trivedi, 2013) and the efficiency of gradient boosting algorithms (Ke et al., 2017), the proposed model enhances both interpretability and scalability in multilevel mortality forecasting.

3.3. Contribution and prospects

Our study introduces a novel approach to mortality risk modeling, integrating local and global data through a hierarchical two-step GBM-based model, inspired by the Poisson Regression model and knowledge distillation techniques. This innovative blend of structured probabilistic models with ML techniques is derived from the Cox Proportional Hazards (Cox PH) model aimed at estimating mortality rates (Cox, 1972). We assume a Poisson distribution for deaths, crafting a Poisson log-likelihood assumption as the loss function. The adoption of LightGBM for simplifying implementation highlights the proposed method's strength in leveraging existing tools to enhance efficacy and accessibility (Ke et al., 2017; Shi et al., 2025).

Data for the study was collected in a pseudonymized form from eight different operating units of a global primary insurance company, each representing a distinct country. The dataset includes policy data that remained active during this period, even if initially issued before the earliest year studied. In total, the dataset encompasses nearly 10 million life-years of exposure and close to 10,000 recorded insurance claims (deaths). The data underwent analysis in an aggregated form, grouped into (N = 16,689,304) unique combinations of feature values. Specifically, the feature set $X_{i,j}$, where group i ranges from 1 to N and j ranges from 1 to 8, representing the eight countries, consists of a total of 26 features. Among these features, 9 are global, and up to 17 are local features, encompassing information about policyholders, insurance policies, and claims. Given these potential risk factors, our target is to model the number of deaths $D_{i,j}$ in relation to the life years of risk exposure $E_{i,j}$.

Our primary goal is to accurately evaluate mortality rates for all countries and feature combinations. We aim to estimate the conditional expectation of death counts, denoted as $D_{i,j}$, given the available information summarized in the feature set $X_{i,j}$ and the exposure in life years at risk $E_{i,j}$. Assuming a Poisson distribution again:

$$D_{i,j} \sim \text{Poisson}(\mu_{i,j} \cdot E_{i,j}),$$

the expectation for death counts is:

$$E[D_{i,j}|X_{i,j}, E_{i,j}] = \mu_{i,j} \cdot E_{i,j} = \exp\left(X_{i,j}^{\intercal}\beta_j\right) \cdot E_{i,j}.$$

The Poisson log-likelihood is defined as:

$$l(\beta_j|X_{i,j}, D_{i,j}) = \sum_{i=1}^{N_j} \left(D_{i,j} \cdot \log(\hat{D}_{i,j}) - \hat{D}_{i,j} \right),$$

where $D_{i,j}$ denotes the observed death counts, $\hat{D}_{i,j} = \hat{\mu}_{i,j} E_{i,j}$ denotes the predicted death counts, and $\boldsymbol{\beta}_j$ is the parameter vector.

This formulation assumes that deaths follow a Poisson distribution and by assuming piecewise constant hazard rates over time, the likelihood of the Cox PH model coincides with the likelihood of the Poisson GLM when employing $log(E_{i,j})$ as an offset parameter. An advantage of simplifying the Cox PH model into a Poisson GLM is its adaptability to the ML realm, where optimization using Poisson log-likelihood allows for defining offsets or observation weights. ML models, which typically do not assume specific additive relationships between features and targets, can leverage this flexibility, yielding in the formulation: $E[D_{i,j}|X_{i,j}] = \mu_{i,j} \cdot E_{i,j} = \exp(f(X_{i,j})) \cdot E_{i,j}$. Here, $f(X_{i,j})$ within the exponential function demonstrates the ability to capture interactions and non-linear relationships without explicit specification, offering a key advantage over traditional GLMs. This transition from GLMs to ML models provides additional benefits, including integrated variable selection mechanisms and enhanced ability to model complex interactions.

To distinguish between local and global features and ensure high accuracy in each country, we propose a two-step LightGBM model approach. This approach involves two distinct modeling steps: The first model identifies global patterns and uses a training set that includes data from all countries, focusing solely on "global" features. These global features are those where data across countries is comparable, such as age. In contrast, factors like postal code, which lack comparability between regions, are excluded. In the second step, we calculate one Local model per country, totaling eight Local models. The regional models extend the global framework by incorporating local features, effectively distilling information from a large model into smaller, specialized ones. This hierarchical transfer preserves model accuracy, reduces deployment costs, and facilitates the use of heterogeneous data while maintaining interpretability across scales. Each Local model takes the output of the Global model and adjusts it to the specific circumstances of the respective country. Specialized Local models use all global factors plus the country-specific local factors. The distinction of the feature set into global and local features is based on the availability of data across countries as well as domain-specific expert knowledge. This final estimates combine multiplicatively the estimates from both the global and specialized Local models as illustrated in Figure 3.1.

Mathematically, we can express the process of estimating death counts for a policy with given characteristics as follows: $E[D_{i,j}|X_{i,j}] = \mu_{i,j} \cdot E_{i,j} = q(X_{i,j}^{\text{global}}) \cdot h_j(X_{i,j}^{\text{all}}) \cdot E_{i,j}$ where $D_{i,j}$ represents the expected number of deaths given a set of features $X_{i,j}$ for group i and country j; $q(\cdot)$ represents the Global model's prediction function; $h_j(\cdot)$ represents the Local model's prediction function for country j; $X_{i,j}^{\text{global}}$ represents a set of factor values for group i and country j, containing only global factors; $X_{i,j}^{\text{all}}$ represents a set of factor values for country j, containing both global and local factors. In technical terms, the predicted mortality rates from the first Global model are used to initialize the second specialized Local model. Accordingly, the model continues to work on the resulting residuals and iteratively optimizes the second model through boosting, but now with the extended feature set including localized characteristics. The final predicted number of deaths results from

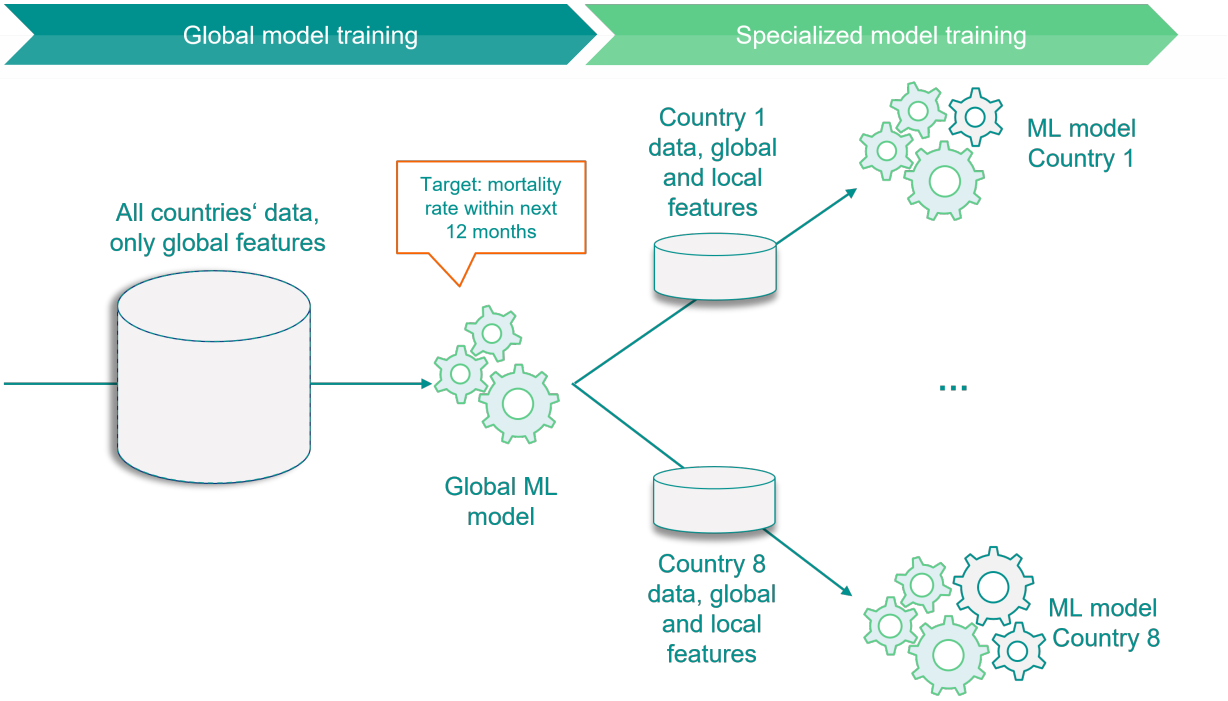


Figure 3.1.: Qualitative illustration of proposed methodology. Gearwheels illustrate the features.

the multiplication of the predictions from the Global model (first step), the predictions from the specialized Local model (second step), and the exposure. The following derivation shows that the multiplication is justified by the nature of the boosting algorithm and the exponentiation by the log link of the Poisson distribution:

$$\begin{aligned}\mu_{i,j} &= \exp\left(\sum_{k=1}^{K} \theta_k \cdot u_k(X_{i,j})\right) \\ &= \prod_{k=1}^{K} \exp\left(\theta_k \cdot u_k(X_{i,j})\right) \overset{g := \exp(\theta \cdot u(X))}{=} \underbrace{\prod_{k=1}^{P} g_k(X_{i,j})}_{\text{Global model} = q(.)} \cdot \underbrace{\prod_{l=P+1}^{K} g_l(X_{i,j})}_{\text{Local model} = h_j(.)}\end{aligned}$$

The starting point is formulated as a linear combination of the base learners with θ_k is the weight of the k-th tree, and $u_k(X_{i,j})$ represents the tree associated with the set of leaves of the k-th tree. Splitting the modeling into two steps offers the advantage of cleanly separating effects into local and global categories. It also optimizes model performance for each market by tailoring the model to local patterns while allowing knowledge sharing across countries via the Global model. When onboarding a new country, we now have the possibility to keep the existing Global model and efficiently compute a new Local model for the country using less data.

The results are reported across several metrics and benchmark methods for all eight countries involved in the study, providing a comprehensive view of the method's effectiveness. All four benchmarked methods are as follows:

- **Local models**: For each country, we take this country's data and run the model separately. This is only applicable if we have enough claims and exposure available for a given country as a solid foundation for training. The information contained in each other country about certain features and their correlation patterns to mortality rates remain unseen for each model.
- **Two-step approach**: This is the proposed novel approach that combines global features in the first step model, using common features across countries. In the second step, a Local model is trained to capture each country's specificities based on residuals from the first step.
- **Global one-step with single value imputation**: Data from different countries are first aggregated through an early data fusion process to create a unified dataset. Discrepancies in feature sets and variable coverage across countries lead to missing data blocks, which were imputed using a single-value imputation strategy.
- **Global one-step with bootstrapped multiple imputation**: Similar to the third approach, this method involves early data fusion by combining datasets from all countries. In this case, we use Bootstrapped Multiple Imputation with Decision Tree as an imputation technique for missing values.

This framework outperforms both purely local models and standard imputation techniques when pooling together different sets of features, particularly in data-scarce regions, by leveraging global patterns to improve generalization. The model is computationally efficient and accurate in handling missing values, making it adaptable for other domains requiring integration of multi-regional data. Overall, our proposed two-step hierarchical modeling approach achieves superior predictive performance for nearly all countries, outperforming Local models and the MICE method, with log-likelihood proving to be a more reliable measure than RMSE due to the distributional assumptions of the data generation process. The two-step model significantly enhances the generalization for smaller countries by protecting local specifics and improves performance even stronger compared to larger countries.

Future research could examine the ethical and policy implications of applying ML in mortality risk estimation, particularly in the context of this modeling approach. Since mortality models influence life insurance pricing and public health policy, it is important to assess their societal impact, fairness, and potential algorithmic biases. Future work may therefore focus on developing fairness-aware ML techniques that mitigate bias while accommodating regional differences and local regulatory requirements. Although this study's methodology is designed to be interpretable by definition, its interpretability could not be fully demonstrated due to data privacy and compliance constraints. Subsequent research could extend this work by incorporating explainable ML methods, such as SHAP value visualizations, to highlight influential features and cross-country differences (Lundberg and Lee, 2017).

4. Transfer Learning for No Data Scenario

4.1. Research question

The third study addresses an extreme data-scarce scenario with regions or countries where no mortality data are available. The core research question guiding this work is:

How can transfer learning be effectively applied to mortality modeling when no local mortality experience data exist?

This question is particularly relevant for emerging life insurance markets or newly introduced products that lack sufficient historical experience data. Addressing it is essential to enable accurate mortality estimation, equitable pricing, and risk management also in data-deficient environments.

To tackle this challenge, the study proposes a transfer learning framework that leverages information from data-rich countries, guided by a country similarity index to ensure meaningful knowledge transfer. A case study based on the UK demonstrates the validity and evaluated the goodness of the proposed method, illustrating how transfer learning can be used to construct reliable mortality estimates and facilitate the development and pricing of new insurance products even in the complete absence of local experience data.

4.2. Literature review and research gap

The concept of transfer learning has been adapted to mortality modeling, particularly in settings where data are limited but not entirely absent. Vincelli (2019) proposed an innovative framework that effectively "reverse-engineers" the structure of an industry mortality table into a set of learned, high-dimensional features. These features serve as transferable representations of mortality dynamics, which can then be fine-tuned to a company's own sparse experience data using nonlinear models such as neural networks. This approach enables knowledge transfer from well-populated industry tables to smaller, company-specific datasets, thereby stabilizing estimates across age-duration cells. Building on this foundation, Lim and Shyamalkumar (2021) incorporated monotonicity constraints to ensure that such data-driven models remain consistent with actuarial principles of smoothness and interpretability.

Our study extends this line of research to a more extreme setting in which the target population lacks any observed mortality data. This represents a non-trivial generalization of prior work, as it requires generating synthetic observations that act as proxies for empirical experience. Within our framework, synthetic data generation is coupled with a boosting-based fine-tuning process, allowing the pretrained model to gradually adapt to the inferred characteristics of the target population.

In addition, we introduce a drift model to explicitly capture demographic and epidemiological divergences between source and target populations, thereby providing a principled explanation for risk differentials and pricing variations across heterogeneous contexts. Together, these contributions offer both methodological advancement and practical value for actuarial modeling, enhancing interpretability, supporting cross-population calibration, and addressing the limitations of existing mortality modeling frameworks that often neglect cross-regional information integration.

4.3. Contribution and prospects

Building on global and local datasets from prior research, we employ a pretrained GBM trained on pooled insurance data from eight countries (excluding the UK) using a common feature set. This model is then used to generate mortality predictions for synthetic populations representing countries most similar to the UK. Figure 4.1 illustrates the target population segments for estimating UK insured population mortality. Although UK-specific insured mortality data is unavailable, we have access to publicly available overall population data categorized by age and gender. The objective is to use this data to bridge the gap and project correlations between insured and overall population mortality from these countries to the UK scenario. This dataset includes common global characteristics shared across different countries, such as age, gender, and sum assured, allowing for cross-country data comparison, plus the overall population mortality for all countries, yielding a total of nine global features.

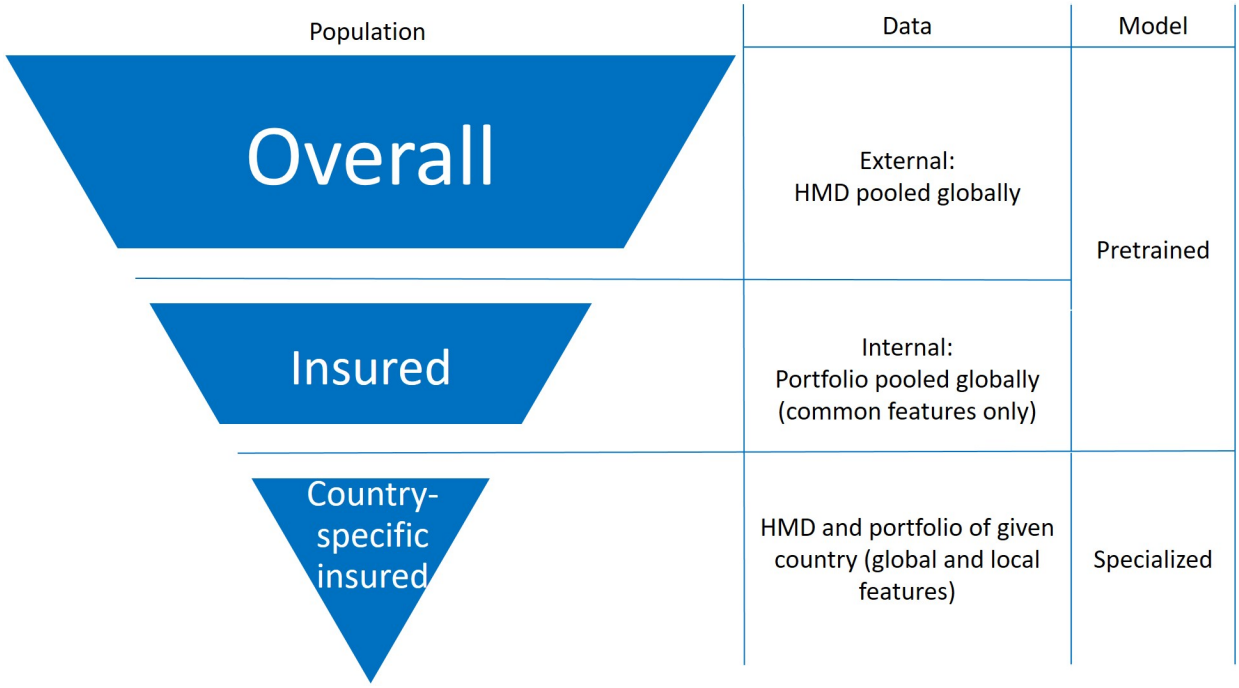


Figure 4.1.: Illustration of targeted population segments across different datasets and models.

The feature set $X_{i,j} \in \mathbb{R}^{N \times p}$ consists of two components: global features $X_{i,j}^{\text{global}}$, which are comparable and available across countries (including data such as the overall population from the HMD (HMD, 2024)), and local features $X_{i,j}^{\text{local}}$, which are specific to each country j and unique

feature combination i, representing demographic characteristics. Our challenge lies in estimating mortality rates $D_{i,M}$ for country M, where we lack internal portfolio data but publicly available population mortality rates exist for this and other countries. To address this, we leverage external population-level information, along with internal experience data from other markets, to infer company-specific mortality estimates for the target country M. In this study, M corresponds to the UK.

We generate synthetic data through proportional resampling and data augmentation, central to the study's findings, offering detailed instructions for its creation to enhance reproducibility and allow other researchers to build upon this work. To measure how similar the target country M is to the K source countries, we create a country similarity index based on Q external insurance and mortality specific demographic items $X^{\text{ext}} \in \mathbb{R}^{(K+1)\times Q}$, with K number of source countries and 1 target country. In our application case, Q is equal to 13, larger than $K+1=9$. After centering and scaling, the Manhattan distance between vectors X_j^{ext} of each source $j=1,\ldots,K$ and X_M^{ext} of target country M is calculated as the sum of the absolute differences between corresponding components of vectors: $d(X_j^{\text{ext}}, X_M^{\text{ext}}) = \|X_j^{\text{ext}} - X_M^{\text{ext}}\|_1$ (Cover and Hart, 1967). This results in a k-dimensional vector, representing the sum of item-wise distances between $j=1,\ldots,K$ and M across all Q items. The summation of distances over the countries is then transformed into the normed similarity score $s(X_j^{\text{ext}}, X_M^{\text{ext}}) = e^{-d(X_j^{\text{ext}}, X_M^{\text{ext}})}$ using the exponential function, so that the value range changes from $[0,\infty)$ to $(0,1]$. For detailed construction of this index, please refer to Chapters IV. Assuming the known age and gender distribution for M, we resample feature combinations (rows) from the K datasets, encompassing both global and local features, along with the number of deaths, proportional to each similarity score $s(X_{i,j}^{\text{ext}}, X_{i,M}^{\text{ext}})$ for $i=1,\ldots,N$ and $j=1,\ldots,K$. After substituting the feature of overall population mortality with that of country M obtained from the HMD, we utilize the global features of the synthetic dataset $X_{i,M}^{\text{global}}$ to generate preliminary predictions $\hat{D}_{i,M}^{\text{pretrained}}$ using a pretrained model. Subsequently, we enhance these predictions by employing specialized GBM models tailored for countries $j=1,\ldots,K$. Through iterative boosting, the specialized model adjusts to the characteristics of the countries according to their similarity, refining the mortality rate predictions. The final mortality rate predictions $\hat{D}_{i,M}$ are determined by combining the specialized predictions $\hat{D}_{i,M}^{\text{specialized}}$ and the pretrained predictions $\hat{D}_{i,M}^{\text{pretrained}}$ for all countries. Using several metrics, we evaluate the agreement of transferred mortality rates $\hat{\mu}_{i,M} = \hat{D}_{i,M}/E_{i,M}$ with the CMI mortality rates $\mu_{i,cmi}$, as a proxy for expected insured population mortality in the UK. We validate results using well-suited metrics such as Spearman correlation, cosine similarity, and R-squared. We propose a drift model to evaluate the remaining disagreement by identifying and quantifying the impact of the drift drivers between the target country's expected mortality and the mortality rates transferred from other countries to M. The inclusion of a drift model to assess residual discrepancies between predicted and expected mortality rates is particularly noteworthy and adds depth to the analysis. We assume again a Poisson distribution for mortality counts in country M, denoted as $D_{i,M} \sim \text{Poisson}(\mu_{i,M} \cdot E_{i,M})$. Our study focuses on examining the discrepancy between the predicted mortality rate $\hat{\mu}_{i,M}$ and the actual rate $\mu_{i,cmi}$ across various features or feature categories. This discrepancy, denoted as δ_i, serves as an indicator of the quality of transfer learning. We adopt the two-stage or residual model proposed by Levantesi and Pizzorusso (2019) to estimate δ_i:

$$D_{i,M} \sim \text{Poisson}(\underbrace{\delta_i \cdot \mu_{i,cmi}}_{=\mu_{i,M}} \cdot E_{i,M}). \tag{4.1}$$

A GLM is used with new exposure $D_{i,cmi} = \mu_{i,cmi} \cdot E_{i,M}$, target $\hat{D}_{i,M} = \hat{\mu}_{i,M} \cdot E_{i,M}$, and model specification as follows (Fahrmeir et al., 2013):

$$\log(\delta_i) = \beta_0 + \beta_{age} \cdot age_{i,M} + \ldots + \beta_{gender} \cdot gender_{i,M} + \log(D_{i,cmi}) \tag{4.2}$$

In the Poisson case, Yan et al. (2009) demonstrated that the method is mathematically equivalent to using the ratio $\frac{D_{i,M}}{D_{i,cmi}}$ as target and $D_{i,cmi}$ as weights:

$$\log\left(\frac{D_{i,M}}{D_{i,cmi}}\right) = \beta_0 + \beta_{age} \cdot age_{i,M} + \ldots + \beta_{gender} \cdot gender_{i,M} \tag{4.3}$$

Additionally, validation techniques, including the inclusion of confidence intervals via bootstrapping methods, are applied to provide a more detailed assessment of the uncertainty associated with the predictions. This is essential to solidify the reliability of the results, especially given the reliance on synthetic datasets. We rely exclusively on anonymized and publicly available data, ensuring compliance with ethical research standards. The lack of reliance on proprietary UK-specific portfolio data also strengthens the ethical standing of the work.

The results indicate that the confidence interval mostly contains the CMI, particularly for males, solidifying the reliability of the results, especially given the reliance on synthetic datasets. Regarding gender-specific mortality risks, while both the transferred results and CMI indicate higher mortality rates for males than females, the transferred estimations may show slight discrepancies: males are slightly overestimated and females underestimated compared to the average mortality risk. However, these deviations appear minor and likely stem from cohort distinctions between CMI and internal data, as well as cultural differences between the primary reference countries and the UK's insurance mortality data, possibly reflecting subtle cultural influences and evolving gender roles in different countries. With the CMI serving as the insurer's base table, the exponentiated effects estimated by the drift model for additional variables provide direct insight to insurers. This allows them to assess the potential impact of including these variables in the pricing model and to determine possible loadings or discounts accordingly. Practical benefits include not only strong predictive performance, but also reduced reliance on local data, and lower computational demands, making it efficient for multi-center studies. It simplifies the development and deployment of ML models by eliminating the need for extensive training data in each new country. Our findings suggest that transfer learning is particularly effective for factors that are less influenced by cultural differences, although it may experience drift when capturing local specificities.

In summary, this framework enables accurate mortality predictions in data-scarce markets, and the findings demonstrate strong agreement with the CMI mortality tables for the UK. The transfer learning leverages a pretrained model from eight other countries due to a lack of local mortality portfolio data, while refining the model using open-source UK total population mortality rates and data synthesized from the available countries according to their similarity degree. While the model performs well with less culture-specific risk factors, discrepancies with CMI mortality tables highlight the need for evaluation using the drift model. This is essential for comprehensive risk assessment and to inform pricing strategies by quantifying expected discrepancies.

While the transfer learning framework holds considerable promise, its generalizability in the absence of local data depends on the similarity between source and target countries. By incorporating mechanisms like the country similarity index, drift model, and bootstrapping confidence

intervals, we facilitate more informed and reliable applications in regions with differing cultural, demographic, or economic conditions, even when local data is completely missing. Further research could explore the generalizability of this approach, specifically examining its performance in regions with differing cultural, demographic, or economic conditions. Local covariates being as homogeneous as possible across the target and source countries would be beneficial. Otherwise, adjustments are necessary when creating the synthetic dataset if the source datasets' local covariates do not apply to the target country at all. Future research could examine possibilities of overcoming this challenge by homogenizing either data collection procedures or post-processing techniques.

5. IPF Simulation for Segmented Mortality

5.1. Research question

Accurate mortality estimates are essential for insurers and pension funds to set fair premiums, manage longevity risk, and design products that reflect the demographic composition of their insured populations. However, access to granular, individual-level mortality data remains highly restricted. Insurers and pension funds hold most detailed records, but privacy regulations, data protection laws, and competitive considerations limit data sharing. As a result, researchers and policymakers often rely on aggregated or anonymized sources that fail to capture the full demographic heterogeneity underlying mortality trends. This reliance can introduce systematic biases in mortality estimation and limit the precision of risk assessments.

To address these limitations, this study develops a simulation-based framework that generates synthetic yet statistically consistent mortality datasets using Iterative Proportional Fitting (IPF). The framework enriches conventional mortality tables with key demographic covariates, such as smoker status, gender, and segmented population counts, enabling finer-grained mortality analysis. By combining IPF-based data synthesis, confidence interval estimation, and an interactive Shiny dashboard, the study provides open-source tools and datasets for both insured and general populations.

This framework offers a practical and replicable solution for enhancing mortality modeling in data-scarce environments by bridging the gap between the need for granular mortality inputs and the constraints imposed by privacy and data availability, ultimately supporting both actuarial and public health applications where demographic differentiation is crucial.

5.2. Literature review and research gap

Existing research on mortality estimation provides important insights into demographic and temporal mortality patterns but remains constrained by data limitations and narrow methodological scope. Prior work has explored small-area mortality estimation (Denecke et al., 2023), mortality prediction within restricted age ranges (Goldstein et al., 2023), and COVID-related mortality forecasting (Duchemin et al., 2022). These studies demonstrate the value of targeted estimation techniques but also highlight the persistent lack of comprehensive, high-resolution mortality data. The need for standardized and accessible data has been consistently emphasized in demographic and epidemiological research (RKI, 2014; Nusselder and Mackenbach, 1997), as well as in recent studies focused on state-level and population-segmented mortality analyses (Smith et al., 2021; Johnson et al., 2023).

Despite these advances, researchers and actuaries remain constrained by reliance on aggregated data, which obscures within-group variation and limits the ability to model demographic heterogeneity. Addressing this gap, the present study introduces a simulation-based framework for mortality data generation that produces synthetic but statistically valid datasets with embedded demographic structure. By enriching mortality tables with demographic covariates and ensuring internal consistency through IPF, this approach facilitates high-resolution mortality modeling and enhances the methodological capacity of subsequent actuarial and public health analyses.

5.3. Contribution and prospects

We propose IPF to calculate multi-dimensional distributions for both deaths and population simulations. Given that mortality data comprises populations and deaths within each subgroup, our objective is to determine the joint distribution for each additional variable. For instance, knowing the age and state population distributions, we aim to compute the joint distribution across age and state categories. Consider a multiway table in N dimensions, each representing a sociodemographic variable. For illustrative purposes, assume N = 3. The multiway table π_{ijk} contains unknown components, subject to constraints defined by marginal distributions $x_{ij\cdot}, x_{i\cdot k}, x_{\cdot jk}$. The constraints ensure that the sum of observations in each category matches the known marginals and the total number of observations n.

The IPF process begins with an initial estimate $\pi_{ijk}^{(0)}$ and iteratively adjusts the values in the deaths or populations table or both based on the specified marginal totals. The algorithm can be extended to higher dimensions, facilitating the synthesis of population data at varying resolutions. For instance, when considering three demographic variables, one iteration of the IPF process can be represented as follows:

$$\pi_{ijk}^{(1)} = \frac{1}{n} \frac{x_{ij\cdot} \pi_{ijk}^{(0)}}{\pi_{ij\cdot}^{(0)}} \tag{5.1}$$

$$\pi_{ijk}^{(2)} = \frac{1}{n} \frac{x_{i\cdot k} \pi_{ijk}^{(1)}}{\pi_{i\cdot k}^{(1)}} \tag{5.2}$$

$$\pi_{ijk}^{(3)} = \frac{1}{n} \frac{x_{\cdot jk} \pi_{ijk}^{(2)}}{\pi_{\cdot jk}^{(2)}} \tag{5.3}$$

Each equation represents an update step where the estimated cell probability π_{ijk} is iteratively adjusted to match the given marginals. Specifically, equation (1) adjusts the initial estimate $\pi_{ijk}^{(0)}$ to align with the marginal totals $x_{ij\cdot}$, ensuring consistency along the first dimension. Equation (2) further refines π_{ijk} using the marginal totals $x_{i\cdot k}$ from the second dimension. Equation (3) completes the iteration by incorporating $x_{\cdot jk}$, ensuring alignment with the third dimension. This iterative process continues until convergence, ensuring that the synthesized dataset accurately represents the given marginal distributions across all dimensions (Agresti, 2012).

Incorporating additional variables, like smoker status, into mortality risk assessments involves disaggregating mortality data to account for distinct risks associated with these variables, while keeping other characteristics constant, similar to the process of disaggregating population data. By applying known hazard ratios or marginal death counts, we can refine mortality tables to reflect these differences accurately. To obtain reliable estimates in countries where insured mortality rates are unavailable, we train the GAM specified with Poisson distributional assumption and log-link on data from the most similar country where insured rates exist. We assume that the ratio between insured and general population mortality rates remains constant across comparable demographic variables between these two countries.

$$D_i \sim \text{Poisson}(\mu_i \cdot E_i), \tag{5.4}$$

Thus, the proposed model for expected insured mortality rates $\hat{\mu}_i$ is as follows:

$$\log(\mu_i) = f_a(\text{age}_i) + f_{g,s}(D_i^P) + \log(E_i) \tag{5.5}$$

The application of IPF allows us to disaggregate the population count while maintaining correct multidimensional distributions in the exposure structure.

Scenario 1 serves as a proof of concept, both for the application of IPF in enriching demographic resolution by including an additional variable (in this case, regional segmentation), and for introducing a Monte Carlo layer of uncertainty given minimal data input. It focuses on enhancing demographic precision while assuming uniform mortality rates across states, applying IPF only to population counts to refine segmentation while keeping mortality fixed. This scenario highlights what can be achieved when only marginal population distributions are available. While aggregating over states leads to the same overall mortality rates as in the DAV tables (DAV, 2022), the value added lies in the disaggregation step: we now obtain a joint population structure including state, which allows us to reflect demographic heterogeneity in the simulated exposure. This is particularly useful when modeling local effects, insurance penetration, or regional portfolio dynamics. Even without direct state-level mortality data, applying uniform mortality rates to demographically distinct state populations yields differentiated death counts. In this sense, the structure alone becomes informative. Monte Carlo simulations add a quantification of uncertainty, showing how mortality outcomes may vary due to differences in exposure structure alone, especially relevant for small regions or sub-portfolios.

Scenarios 2 and 3 build on the same foundational idea: the use of IPF to construct joint demographic distributions, but now extended to incorporate variation in mortality rates as well. Scenario 2 includes smoker status as an additional risk dimension, using hazard ratios to differentiate mortality between smokers and non-smokers. Scenario 3 further applies this framework to general population data, inferring insured mortality using regression-based adjustments. In both cases, IPF is applied not only to population segmentation but also to disaggregate mortality across subgroups. This enables the modeling of more realistic mortality structures when additional marginal information, such as prevalence or hazard ratios, is available. These scenarios illustrate how a common foundation (joint distribution reconstruction via IPF and uncertainty quantification via Monte Carlo) can be extended from pure population modeling to fully differentiated mortality modeling.

In summary, we offer disaggregated mortality data, including both population and death counts, for Germany, Italy, and Switzerland, taking into account demographic distributions like age, gender, smoker status, and state, along with their interactions. Our findings show that the simulated mortality rates closely match the base tables when aggregated at a higher level. They also provide significant insights into demographic impacts on mortality at a more granular level, generating synthetic insured and general populations while preserving realistic distributional assumptions.

As a prototypical framework, this study introduces a statistically rigorous and privacy-compliant methodology that advances both mortality research and actuarial practice. The approach is modular and extensible: each simulation scenario can be expanded to incorporate additional countries, demographic variables, or higher-dimensional interactions. All tools and datasets developed within this framework are made available through an open-source interactive dashboard, alongside the accompanying codebase, to promote transparency, reproducibility, and future methodological extensions. The proposed methodology improves mortality modeling by explicitly accommodating insured-specific marginal distributions where such data exist, thereby enabling precise demographic representation without assuming equivalence to general populations. In cases where insured-level data are unavailable, general population marginals are used as a proxy, acknowledging that certain selection effects may not be fully captured. To mitigate this limitation, we introduce refinements such as integrating empirical data on insured smokers to adjust hazard ratios and disentangle selection effects from broader demographic differences. Furthermore, we employ a GAM with Poisson regression to statistically estimate these selection effects, ensuring that mortality differentials between insured and general populations are appropriately represented. Future work could extend the framework by incorporating insured amounts as an additional stratification variable, enabling the segmentation of risk by coverage level and capturing the impact of benefit size on selection effects and mortality outcomes. Such extensions would further enhance model precision and the interpretive value of simulated mortality datasets.

6. Summary and outlook

The contributions of this dissertation can be summarized as a multifaceted methodological advancement in mortality modeling, integrating classical actuarial models with modern statistical and ML techniques. In particular, it develops four novel approaches: (1) a semiparametric GAM-APC model to incorporate pandemic-related mortality shocks, (2) a hierarchical gradient boosting framework that leverages data-rich populations to improve estimates in data-scarce countries, (3) a transfer-learning scheme using synthetic data and drift correction to model completely unobserved insurance populations, and (4) an IPF-based synthetic data simulator to generate realistic subpopulation mortality datasets. Together, these methods extend mortality forecasting to settings marked by extreme events and sparse or fragmented data, contributing new tools to the demographic and actuarial modeling literature.

All of the proposed innovations have direct real-world relevance. For example, the GAM-APC approach provides accurate post-pandemic extrapolations mortality projections with penalized smoothing splines revealing the mortality patterns across multiple countries, which are critical for pension planning and public-health decision-making in the wake of COVID-19. The hierarchical boosting and transfer learning models enhance actuarial and insurance analytics by improving longevity risk estimates in emerging markets or small portfolios where traditional country-specific models often fail. The IPF simulation framework enables analysis of segmented populations (e.g. insured vs. general) when detailed data are unavailable, supporting research advancements for difficulty accessible data sources and thus informed decision-making in health policy and insurance planning. All methods are accompanied by implementations in open-source software to ensure reproducibility and facilitate adoption by practitioners.

Looking ahead, this work opens several general future directions. One extension is to enrich these models with additional covariates (for example socioeconomic, health or behavioral indicators) to capture more granular demographic risk factors. Another is to incorporate fairness-aware and privacy-preserving techniques (for example via federated learning) to ensure equitable and secure mortality predictions across subpopulations. Finally, leveraging emerging cause-of-death databases, future work can develop models for specific diseases or death causes, enabling finer analysis of longevity trends. These future directions are aligned with existing best practices and are achievable using current technologies, offering a clear pathway for advancing mortality modeling in complex real-world scenarios.

Part II.

Generalized Additive Model in Semiparametric APC Framework

7. Forecasting mortality trends: Advanced techniques and the impact of COVID-19

Contributing article

Nalmpatian, A., Heumann, C., and Pilz, S. (2024). Forecasting Mortality Trends: Advanced Techniques and the Impact of COVID-19. *Stats.*

Code and data repository

https://doi.org/10.5281/zenodo.13905807

Copyright information

Author contributions

The idea to integrate COVID-19 as a temporal shock within an Age-Period-Cohort (APC) framework using Generalized Additive Models (GAMs) was developed by Asmik Nalmpatian. Asmik Nalmpatian carried out the literature review and proposed the incorporation of pandemic effects via scenario-based extrapolation in GAM-APC models. She conducted all statistical modeling, evaluation of forecast performance, and visualization of results. Asmik Nalmpatian also wrote the initial draft of the paper. Christian Heumann and Stefan Pilz provided methodological guidance throughout, contributed to the design of the APC extensions, and thoroughly reviewed and revised the manuscript. The content of this paper was presented by Asmik Nalmpatian at the 5th Insurance Data Science Conference, held on 15–16 June 2023 at Bayes Business School, City, University of London.

stats

MDPI

Article

Forecasting Mortality Trends: Advanced Techniques and the Impact of COVID-19

Asmik Nalmpatian *, Christian Heumann and Stefan Pilz

Department of Statistics, Ludwig Maximilian University of Munich, 80539 Munich, Germany; christian.heumann@lmu.de (C.H.); stefan.pilz@posteo.de (S.P.)
* Correspondence: asmik.nalmpatian@campus.lmu.de

Abstract: The objective of this research is to evaluate four distinct models for multi-population mortality projection in order to ascertain the most effective approach for forecasting the impact of the COVID-19 pandemic on mortality. Utilizing data from the Human Mortality Database for five countries—Finland, Germany, Italy, the Netherlands, and the United States—the study identifies the generalized additive model (GAM) within the age–period–cohort (APC) analytical framework as the most promising for precise mortality forecasts. Consequently, this model serves as the basis for projecting the impact of the COVID-19 pandemic on future mortality rates. By examining various pandemic scenarios, ranging from mild to severe, the study concludes that projections assuming a diminishing impact of the pandemic over time are most consistent, especially for middle-aged and elderly populations. Projections derived from the superior GAM-APC model offer guidance for strategic planning and decision-making within sectors facing the challenges posed by extreme historical mortality events and uncertain future mortality trajectories.

Keywords: mortality modeling; COVID impact; multi-populational; cross-country; generalized additive models; partial APC plots; APC; machine learning; excess mortality

Citation: Nalmpatian, A.; Heumann, C.; Pilz, S. Forecasting Mortality Trends: Advanced Techniques and the Impact of COVID-19. *Stats* **2024**, *7*, 1172–1188. https://doi.org/10.3390/stats7040069

Academic Editor: Wei Zhu

Received: 15 September 2024
Revised: 10 October 2024
Accepted: 12 October 2024
Published: 16 October 2024

1. Introduction

In recent decades, life expectancy in the developed world has shown a substantial increase, exemplified by a notable 56.7% reduction in the mortality rate of 80-year-old men in the USA from 1933 to 2019. This trend, reflecting societal advancements and improved healthcare systems, underscores the importance of accurately predicting mortality trends for informed decision-making by policymakers, pension schemes, insurers, and social security systems. The emergence of the COVID-19 pandemic has further underscored the need to understand its impact on mortality trends over the short- to mid-term. Our study focuses on assessing the impact of COVID-19 on mortality trends, aiming to enhance existing mortality models while maintaining explainability.

The literature presents a diverse range of approaches to mortality modeling, from traditional stochastic models like those discussed by [1] to modern methods such as the use of machine learning (ML) techniques. Recent studies have shown that methods like pure Gradient Boosting or Random Forest perform exceptionally well [2], while others have explored hierarchical approaches with ML building upon simpler LC models [3]. Furthermore, recent advancements include the application of neural networks to enhance mortality models, such as the Common-Age-Effect Model proposed by [4] and the extension of LC models for multiple populations demonstrated by [5]. Generalized Additive Models (GAMs), a well-established model class, was first introduced by [6] and has been applied in the mortality context; [7] describes a Bayesian APC model with an autoregressive prior on the age, period, and cohort terms; and [8] proposed the use of a bivariate spline function within a GAM to effectively capture two-dimensional cohort information. A similar model was applied by [9] for projecting cancer incidence and mortality in Finland, by [10] for

mortality in the UK, and by [11] for projecting breast cancer mortality in Spain. However, none of them model and extrapolate GAM in the APC framework with a tensor product in a multi-populational fashion. The literature on accounting for COVID in mortality projections is also growing, e.g., [12], which uses the stochastic Li and Lee model; [13], which proposes parsimonious decomposition of the mortality surface on a polynomial basis with regularization and cross-validation; and [4], which quantifies the impact of the 2020 mortality shock by calibrating the Lee–Carter model. However, none of the aforementioned studies employ GAM in APC and evaluate scenarios post-pandemic.

To compare with the traditional stochastic mortality models, LC [1] and age–period–cohort (APC) [14], alongside contemporary ML methodology proposed by [3], this study introduces a cross-country GAM within an APC framework, utilizing a smoothed second-order spline with penalty points. To our knowledge, this research is the first to integrate the GAM method into the APC framework in a multi-population context and employ it to extrapolate the impact of COVID-19 on future mortality trends. We examine Germany, Finland, the Netherlands, Italy (representing Europe), and the United States (representing North America) using data from the Human Mortality Database [15]. We employ a cross-country approach, enabling the model to learn from multiple countries concurrently, thereby capturing both universal trends and country-specific variations in mortality patterns.

Our research makes three key contributions, which are visually summarized in Figure 1. First, we compare the predictive performance of four models, including traditional single-population and contemporary multi-populational models, for modeling and projecting future mortality rates for five countries. We find that the most promising approach is based on a GAM, where cohorts are represented as an interaction between age and period. This framework, adaptable for both aggregated and individual survival data, introduces a state-of-the-art method for the field of multi-populational cross-country mortality research. Secondly, we introduce partial APC plots as a novel graphical tool in mortality research, enabling the analysis of specific APC structures. This tool aids in communicating complex temporal patterns and highlights gender-specific and cross-country differences. Finally, we provide fresh insights into the factors driving the impact of the COVID-19 pandemic on mortality for the five countries. Through analyzing age, period, and cohort associations in a multi-population context using a GAM within an APC framework, we extrapolate mortality rates into the future. Four scenarios, representing varying degrees of pandemic impact, are evaluated against observed mortality data post-pandemic to identify the most accurate scenario.

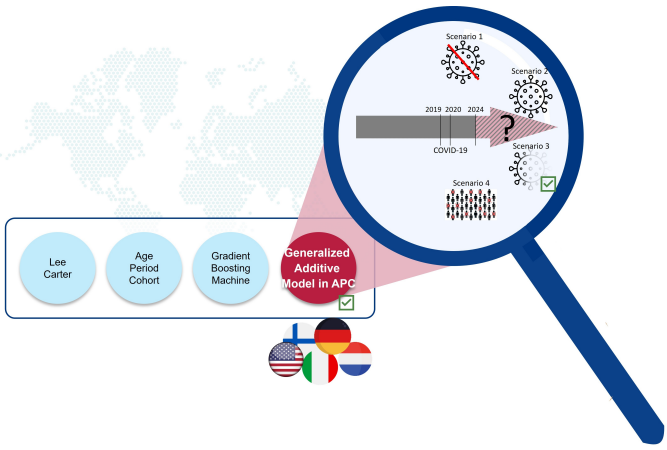


Figure 1. Schematic illustration of the key findings: (1) GAM-APC is the most effective approach for mortality forecasting. (2) The model provides multi-populational and cross-country insights. (3) Among the four pandemic impact scenarios, Scenario 3, which assumes a diminishing impact of COVID-19, is the most consistent, particularly for middle-aged and elderly populations.

Stats **2024**, *7* 1174

The practical implications of our findings are considerable. Our research demonstrates the efficacy of the GAM within the APC framework and its capacity to extrapolate mortality forecasts, accounting for the impact of the COVID-19 pandemic. This offers invaluable insights for policymakers and stakeholders, providing guidance in navigating the uncertainties brought about by the pandemic, with a particular emphasis on matters pertaining to life insurance and pension funds.

Our study follows a structured approach, beginning with an overview of the database and methodology. We then compare the predictive performance of benchmark methods across countries, offering insights into optimal trend forecasting techniques. Additionally, we conduct scenario analyses to evaluate the impact of COVID-19 on mortality trends. Finally, we conclude by summarizing our key findings and implications.

2. Data and Methods

2.1. Data

The study relies on the Human Mortality Database (HMD) [15], which provides mortality rates $\mu_{a,t,g,c}$, death counts $D_{a,t,g,c}$, and population sizes $E_{a,t,g,c}$ categorized by age a, year t, gender g, and country c. The selection of countries is based on their geographic context and the contrasting impact of the COVID-19 pandemic. Specifically, Finland (until 2019), Germany (until 2017), Italy (until 2018), and the Netherlands (until 2019) represent Europe, while the United States (until 2019) represents North America.

For the most recent years up until 2023, the Short-Term Mortality Fluctuations (STMF) [16] series offers partial information on mortality rates and population, grouped by week and age categories. However, the study requires data on a yearly basis and in a metric age scale. To address this, we employ a methodology to construct mortality data on a yearly basis and in a metric age scale, which are then combined with the original HMD data. This process is based partly on the proposal by [12] and further detailed in Appendix A.

Figure 2's heatmaps visualize mortality rate changes in the United States, showing a decreasing trend over the years. Darker red colors indicate higher mortality rates, with females generally exhibiting lower rates than males, especially in older age categories. Infant mortality rates have strongly improved, transitioning from red to white. The diagonal lines symbolize distinct birth cohorts, highlighting the three primary effects examined in this study: age, period, and cohort.

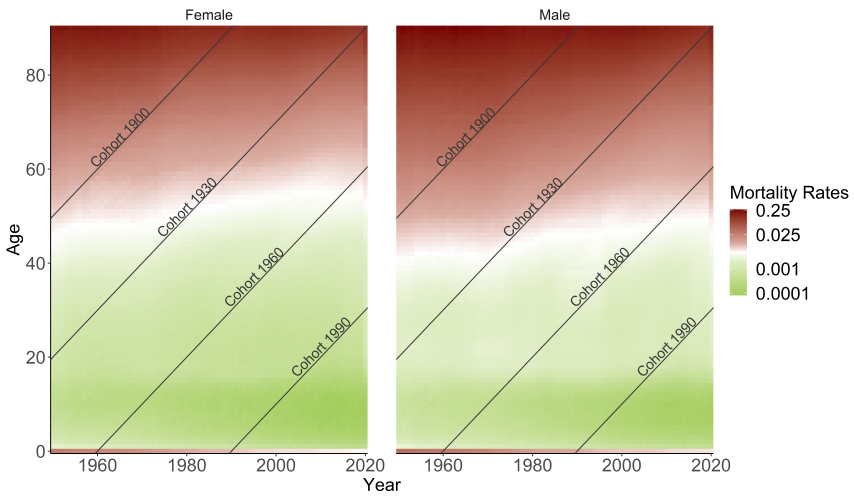


Figure 2. Heatmaps of mortality rates for the US population are shown, with age groups and periods represented horizontally and vertically, respectively. The diagonal lines display unique cohorts.

2.2. Methods

The aim of the developed and applied methodology is to enhance mortality modeling by prioritizing predictive performance and future trend forecasting. To achieve this goal, four different methods have been benchmarked and outlined below: traditional ones such as Lee–Carter (LC) and age–period–cohort (APC); a more modern two-step approach using Gradient Boosting Machine (GBM); and finally, Generalized Additive Model (GAM) within the APC framework.

2.2.1. Lee–Carter (LC)

The classical LC model, initially proposed by Lee and Carter (1992) [1], estimates and forecasts mortality rates $\mu_{a,t}$ at age a in year t. This model is applicable on single-populational data, considering one gender category for each country:

$$\log \mu_{a,t} = \alpha_a + \beta_a \kappa_t + \epsilon_{a,t} \tag{1}$$

Here, α_a represents the age-specific average over time, β_a describes the rate of mortality improvement at age a, and κ_t denotes the general trend of mortality at time t. The error terms $\epsilon_{a,t}$ reflect the residual age- and year-specific historical influence on mortality rates that the model cannot capture. The authors suggest using the singular value decomposition (SVD) method to find the least squares solution to minimize the residual sum of squares. The original method is embedded in a single-populational Poisson regression model by [17]:

$$D_{a,t} \sim \text{Poisson}(E_{a,t} \cdot \mu_{a,t}) \tag{2}$$

The expected number of deaths according to the LC fit can be calculated as $D_{a,t} = E_{a,t} \cdot \mu_{a,t}$ using linear predictor $\eta_{a,t} = \log \mu_{a,t} = \alpha_a + \beta_a \kappa_t$ with constraints like $\sum_a \beta_a = 1, \sum_t \kappa_t = 0$ [18].

2.2.2. Age–Period–Cohort (APC)

The APC model extends the LC model by including a cohort effect γ_{t-a} and omitting the age-specific improvement rates, yielding in the following linear predictor: $\eta_{a,t} = \alpha_a + \kappa_t + \gamma_{t-a}$. The cohort is generally computed by $cohort = year - age$. This model is applicable on single-populational data and has its origins in the fields of medicine and demography, going back a long way ([14,19]). However, [20] was the first who considered this type of model in the actuarial field. With the Poisson distribution assumption and the log link function remaining the same, it can be traced back to the general shape of Generalized APC models [18]. The identifiability is ensured with the following constraints: $\sum_t \kappa_t = 0$, $\sum_{c=t_{min}-90}^{t_{max}-0} \gamma_c = 0$, and $\sum_{c=t_{min}-90}^{t_{max}-0} c\gamma_c = 0$, indicating that the cohort effect oscillates around zero with no apparent linear trend.

2.2.3. Two-Step Approach with Gradient Boosting Machine (GBM)

In this study, we adopt the two-step approach proposed by [3] to refine mortality rate estimations derived from the LC model using Gradient Boosting Machine (GBM).

Firstly, we employ the LC model to estimate mortality rates $\mu^{LC}_{a,t,g,c}$ separately for each country c and gender g. Secondly, we leverage GBM to adjust these estimations by estimating the improvement factor $q_{a,t,g,c}$, which corrects for underestimations (values greater than 1) or overestimations (values smaller than 1). This step involves training GBM with hyperparameter optimization on a multi-populational level using age, year, cohort, gender, and country as features, with death counts $D_{a,t,g,c}$ as the target variable:

$$D_{a,t,g,c} \sim \text{Poisson}(\underbrace{q_{a,t,g,c} \cdot \mu^{LC}_{a,t,g,c}}_{=\mu_{a,t,g,c}} \cdot E_{a,t,g,c}). \tag{3}$$

The adapted exposure $\tilde{E}_{a,t,g,c}$ is calculated as $\mu^{LC}_{a,t,g,c} \cdot E_{a,t,g,c}$. Although the software does not permit the direct inclusion of $\tilde{E}_{a,t,g,c}$ as an offset, we circumvent this limitation by scaling the death counts by exposure $D_{a,t,g,c}/\tilde{E}_{a,t,g,c}$ and using $\tilde{E}_{a,t,g,c}$ as weights, a method shown to be mathematically equivalent to using death counts as target and exposure as offset in the Poisson case by [21]. Details on the GBM methodology are given in Appendix B. Finally, the refined mortality rates are obtained as $\mu_{a,t,g,c} = \mu^{LC}_{a,t,g,c} \cdot q_{a,t,g,c}$.

2.2.4. Generalized Additive Model (GAM) in APC Framework

We integrate the APC framework into GAM in a multi-populational setting, allowing for the modeling of nonlinear, smooth effect structures and facilitating the capture of complex relationships.

In the traditional regression framework, collinearity arises among the three temporal components (age, period, and cohort) leading to identification problems. To mitigate this, we employ a bivariate tensor product with penalized B-splines between age and period, establishing a two-dimensional interaction surface. This surface inherently incorporates cohort information along its diagonals, as illustrated in Figure 2. By doing so, we address the identification problem without imposing restrictive assumptions or constraints [8,14].

By employing multi-populational modeling, within the same GAM framework, we estimate a separate bivariate function for each country and gender interaction. This enables the capture of specific mortality patterns within each subpopulation while allowing countries to learn from each other's experiences through the intercept term.

Finally, we fit a semiparametric additive Poisson regression with a log link, using death counts $D_{a,t,g,c}$ as the target and offsetting for population size $E_{a,t,g,c}$. The model structure is formulated as follows:

$$\log(\mu_{a,t,g,c}) = \beta_0 + f_{a,t,g,c}(age, time) + \log(E_{a,t,g,c}) \tag{4}$$

Here, $f_{a,t,g,c}$ represents the bivariate function for the interaction of age a and period t, specific to each gender g and country c combination.

To visualize the marginal effects of each component effectively, temporal developments are condensed in one specific dimension (either age, period, or cohort) and averaged over the respective other component. This approach allows for examining the effects of age or period by country and gender while considering the cohort values as post-stratification [22–24]:

$$\begin{aligned} f_{a,g,c}(age) &= \frac{1}{T} \sum_{time \in T} f_{a,t,g,c}(time \,|\, age) \\ f_{t,g,c}(time) &= \frac{1}{A} \sum_{age \in A} f_{a,t,g,c}(age \,|\, time) \\ f_{c,g,c}(cohort) &= \frac{1}{A \cdot T} \sum_{age \in A} \sum_{time \in T} f_{a,t,g,c}(age, time \,|\, cohort) \end{aligned} \tag{5}$$

For forecasting purposes, the impact of COVID-19 on mortality rates is incorporated into the mortality model using an additional variable called *covid*. This variable is specific to each country c and takes the value of 0 for years until 2019 and 1 for the years 2020 and 2021, representing the period during which the COVID-19 pandemic has proven to have a strong impact. The values taken for future predictions are subject to the scenario assumptions elaborated in Section 3.

$$\log(\mu_{a,t,g,c}) = \beta_0 + f_{a,t,g,c}(age, time) + \beta_{covid} covid^{*}c + \log(E_{a,t,g,c}) \tag{6}$$

The interaction $covid^{*}c$ allows the model to capture the country-specific impact of the pandemic on mortality rates during the years 2020 and 2021 and, thus, project into the future.

To forecast the mortality rates into the future, we follow a differentiated approach, as outlined in Figure 3. For the time-dependent components (period and cohort) of the LC, APC as well as the final rates of the two-step approach with GBM ARIMA model as random walk with linear drift and automated parameter estimation are selected [25–27]. The forecast of mortality rates for GAM is based on extrapolation of the spline fit, assuming a globally quadratic structure and a persistent curvature outside the observed data. The choice of degrees of freedom for the covariates can be either predetermined or estimated automatically. We caution against extrapolating too far into the future [24].

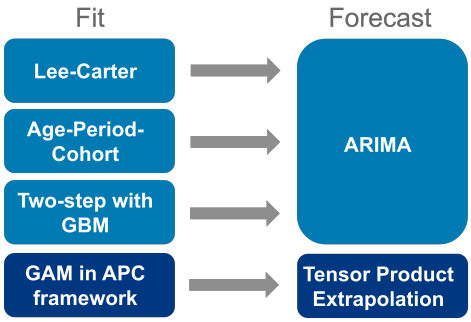


Figure 3. Illustration of the models used for fit and forecast.

We employed R for the technical implementation of the LC and APC models [28], and their forecasting [25], as well as for GAM [29]. Additionally, for the two-step approach, we utilized Python 3.11.10 to implement a Light GBM algorithm [30] and Hyperopt [31] for hyperparameter optimization.

To evaluate the accuracy of the models, the root-mean-square error (RMSE) is calculated, which measures the average difference between the mortality rates $\hat{\mu}_{a,t,g,c}$ and the observed mortality rates $\mu_{a,t,g,c}$ across all ages a, years t, genders g, and countries c. The RMSE is computed as follows:

$$RMSE_{a,t,g,c} = \sqrt{\frac{\sum_{a\in A}\sum_{t\in T}\sum_{g\in G}\sum_{c\in C}(\hat{\mu}_{a,t,g,c} - \mu_{a,t,g,c})^2}{n}} \tag{7}$$

where n represents the total number of observations.

3. Results

This section begins with an evaluation of predictive performance, both in-sample and out-of-sample, for all four models. Table 1 gives an overview of the different training and test sets used in the analysis. It should be noted that for the purposes of benchmarking, years up to and including 2019 are used on the assumption that they are not affected by the impact of the COVID-19 pandemic. This approach eliminates any year-specific artifacts that might otherwise affect the assessment of the predictive performance of the models themselves. The single-populational models LC, APC, and GBM (based on LC in the first step) are limited to using data from Germany from 1990 onwards due to data inconsistencies before reunification. In order to maintain comparability across the tensor–product spanned by years and ages, it is necessary to ensure that the multi-population GAM is coherent. This is because GAM requires a joint coverage of years—a prerequisite for coherent modeling. As a result, available years for all countries will be restricted to 1990–2015 to ensure sufficient training for capturing current effects on mortality rates and projecting them into the future for 2016–2019. Conversely, to ensure comparability among the LC, APC, and GBM models, we permit a shorter training period for Germany (until 2010), resulting in a longer testing period. This approach is especially appropriate for LC, which, due to its linear nature, does not require extensive historical data to accurately forecast future trends.

Table 1. Model training and test periods for benchmarking.

Model	Country	Training Set (Fitting Period)	Test Set (Forecast Period)
LC, APC, GBM	Finland, Italy, Netherlands, US	1950–2010	2011–2019
LC, APC, GBM	Germany	1990–2010	2011–2019
GAM	Finland, Italy, Netherlands, US, Germany	1990–2015	2016–2019

Focusing on in-sample RMSE for single-populational LC, APC, multi-populational GBM, and GAM models, we fit the training periods range from 1950 to 2010 for Finland, Italy, the Netherlands, and the US, and from 1990 to 2010 for Germany. GAM models are fitted from 1990 to 2015 for all countries. In-sample error is calculated within the same range as the model training period and reveals that the two-step approach with GBM and GAM in the APC framework exhibits superior performance over traditional stochastic mortality models LC and APC. There is no clear preference between GBM and GAM, with both achieving strong reductions in RMSE compared to LC and APC for in-sample predictions. Table A1 in Appendix D provides an overview of the goodness-of-fit for information reasons.

GBM demonstrates proficiency in adaptive learning, which enables it to discern intricate, non-linear relationships that markedly enhance model fit. However, this also gives rise to the possibility of overfitting, whereby the learning process is influenced by noise rather than the underlying trends, thereby enhancing training performance but not ensuring better long-term generalization. Table 2 highlights the differences in computational efficiency. The APC model, being slightly larger and taking longer to run than the LC model, reflects its added complexity due to the cohort dimension. As expected, LightGBM (GBM) takes significantly longer to train and uses substantially more memory than GAM. This is because gradient boosting involves iterative training and optimization across many decision trees, which requires more computational resources than fitting a GAM model in Appendix C. The model size of GBM is also larger than that of GAM, reflecting its complexity and the higher number of parameters.

Table 2. Computational efficiency for training, summarized for all 10 single populations for LC and APC.

Metric	LC	APC	GBM	GAM
Runtime (s)	63.15	87.31	799.75	412.49
Memory (MB)	17.9	20.3	301	76.2
Storage (MB)	0.0274	0.0308	72.10	11.78

Table 3 presents the out-of-sample results for different models, indicating the forecasting quality across countries and genders. Out-of-sample RMSE is calculated based on forecast periods ranging from 2011 to 2019 for LC, APC, and GBM models, while GAM forecasts span from 2016 to 2019 for all countries.

Table 3. Out-of-sample RMSE comparison for LC, APC, GBM, and GAM models. Forecast for LC, APC, and GBM based on ARIMA (2011–2019), while GAM extrapolates tensor product (2016–2019).

Country	Female				Male			
	LC	APC	GBM	GAM	LC	APC	GBM	GAM
FIN	0.0021	0.0029	0.0028	0.0012	0.0029	0.0029	0.0027	0.0015
DE	0.0048	0.0046	0.0039	0.0021	0.0052	0.0045	0.0045	0.002
ITA	0.0045	0.0025	0.0044	0.0016	0.0042	0.0021	0.0026	0.0013
NLD	0.003	0.0020	0.0024	0.0013	0.0035	0.0038	0.0027	0.0011
US	0.0023	0.0018	0.0014	0.0010	0.0054	0.0020	0.0031	0.0016

The differing training periods for the multi-populational GAM model and single-populational models were necessary to ensure coherent modeling based on historical data availability. We used RMSE as a consistent metric to measure mean forecast error across models, and despite concerns regarding GAM's shorter training period, outlier analysis confirmed the reliability of its RMSE mean.

While GBM shows improved fit and forecast performance, GAM exhibits stronger improvement in forecast accuracy, especially for short-term forecasts within a few years. The GAM-based APC model achieves notable reductions not only in fit but also in forecast errors compared to the classical APC model, implying improved accuracy of mortality rate predictions. The choice of GAM for further analysis is justified based on its superior forecast performance.

One key highlight of the GAM in APC framework is its multi-populational nature, enabling the interpretation of exponential marginal effects, with age, period, and cohort being the components analyzed further. Figure 4 displays the effects of the model based on these components: Both age and cohort effects conform to expectations; thus, higher ages correspond to higher mortality rates while cohort effects reflect variations stemming from individuals' unique experiences based on their birth year [32–34]. Conversely, similar reverse effects are observable for age and cohort.

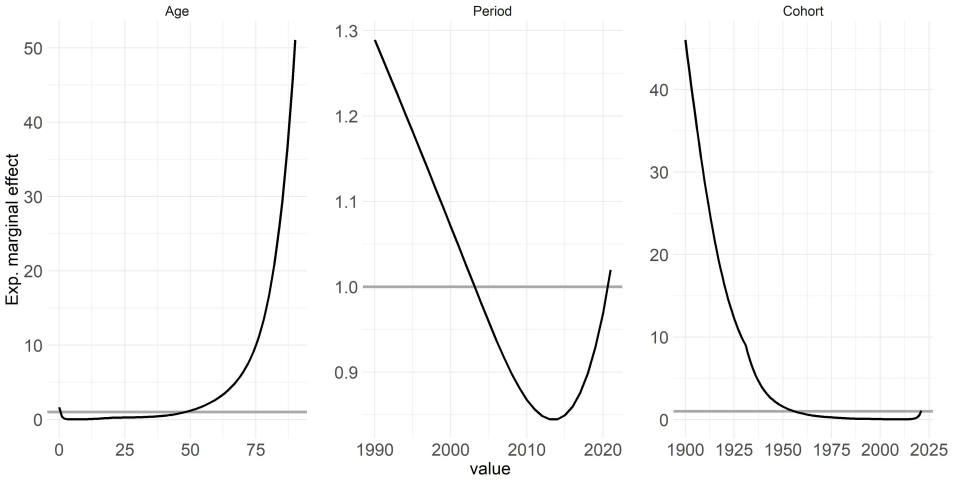


Figure 4. The figure illustrates the estimated marginal effects of age, period, and cohort on mortality rates across multiple countries and genders based on GAM fitted for years 1990–2020.

The period effect, which indicates the improvement of mortality over time and is influenced by external factors affecting all age groups equally at a given point in time, exhibits a notable increase leading up to 2020 [35,36]. However, the period effect notably spikes, particularly approaching 2020, signifying a strong influence of this year on mortality rates. Specifically, the strides made in improving mortality rates over preceding years or even decades appear to have been offset by the effect of COVID-19, resulting in a regression to levels observed around 2003. Appendix D contains this figure stratified by countries and genders for more detailed interpretation.

Following the benchmarking of the four models, we delve into an analysis of the effects of each temporal component on mortality rates throughout the considered time period. Finally, based on GAM in APC framework, we assess the trend forecast into the future considering the impact of COVID-19.

Even though the impact of COVID-19 is, fortunately, diminishing in the present time, its impact on historical (and future) data and the persisting uncertainties in the future cannot be overlooked. These factors necessitate continued attention for many years to come. It is important to note that the idea and methodology employed in this study extend beyond COVID-19 and encompass other events, especially those occurring at the edge of time series, which can present challenges for standard breakpoint analyses.

The scenarios depicted herein must be viewed in light of a meticulous plausibility assessment and the underlying assumptions. To validate the scenario-based findings, we engaged with epidemiological experts. This collaboration is paramount for ensuring the reliability and robustness of the analysis, especially given the complexities inherent in such events. Comparing our framework with expert opinions in the literature, as conducted by citetelenti2021after, reveals a high level of agreement.

Scenario 1: In this scenario, the assumption is made that COVID-19 will disappear in the future. The model is trained using data up to 2019 only, excluding the years 2020 and 2021. The predictions are then made for the years 2020–2025, assuming no long-term effects of COVID-19 on mortality. This approach treats COVID-19 as a special event that does not have any influence on mortality in the upcoming years. The model focuses on the underlying mortality trend without considering the impact of COVID-19 and, thus, without the COVID-indicator.

Scenario 2: In Scenario 2, the expectation is that the full effect of COVID will persist in the future. The model is trained on mortality data up to and including 2021, encompassing years impacted by COVID-19. The indicator variable *covid* is incorporated, set to 0 for years before 2019 and 1 for 2020 and 2021. Predictions are then made for subsequent years, assuming that *covid* remains set to 1 to indicate the ongoing presence of COVID-19. This scenario assumes that the COVID-related situation will continue similarly as it did until 2021 and that it will have a consistent effect on mortality over the coming years.

Scenario 3: In this scenario, the assumption is made that the COVID effect will flatten over time. Similar to Scenario 2, the model is trained using mortality data up to and including 2021. However, in this case, the COVID-19 effect is assumed to decrease exponentially over time. The predictions take into account the diminishing impact of COVID-19 in the future, reflecting the belief that the effect of COVID-19 on health and mortality will slowly flatten out and eventually disappear after a few years. Therefore, the *covid* indicator takes exponentially decreasing values between 1 and 0 for each year.

Scenario 4: In Scenario 4, the focus is on adjusting for excess mortality associated with COVID-19. The years 2020 and 2021 are treated as outliers, but the excess mortality is explicitly considered. The model calculates the difference between the expected death counts and the actual mortality counts for these two years to account for the excess mortality. It is assumed that the excess mortality will not average out over the coming years and must be explicitly accounted for. The baseline mortality, representing the mortality trend without the influence of COVID-19, remains unchanged. This scenario allows for separate consideration of the excess mortality caused by COVID-19 while keeping the baseline mortality unchanged.

The formulation of potential future scenarios related to the pandemic is a complex undertaking, shaped by a multitude of factors, including political decisions and societal acceptance. Despite the inherent challenges, these presented scenarios offer a valuable foundation for mortality forecasts, taking into account the evolving attitudes of life insurers and contributing to the ongoing discourse surrounding the impact of the pandemic. A summary of the key assumptions underlying each scenario is provided in Table 4.

Table 5 presents the different training and test periods used in the scenario analysis, now also considering years after 2019.

Table 4. Discussion of assumptions underlying COVID-19 mortality impact scenarios.

Assumption	Description
Scenario 1: Complete disappearance of COVID-19	
No long-term effects	Assumes no long-term health consequences for recovered individuals, despite evidence of "Long Covid" [37]. Assumes no lasting psychological or social impacts from lockdowns [38].
Vaccination effectiveness	Assumes widespread vaccination will lead to the abrupt disappearance of the pandemic, despite uncertainties about long-term vaccine efficacy.
Excess mortality	Assumes excess mortality will average out in the coming years, with no rapid population reductions.
Scenario 2: Continuous COVID-19 impact	
Viral variants	Acknowledges that while vaccines reduce infection risk [39], rising incidence rates suggest ongoing challenges [40]. Considers the potential for emerging variants to undermine vaccine effectiveness.
Consistent mortality impact	Assumes the impact on mortality will remain unchanged over the next years, despite short-term decreases and uncertainties as well as advancements in science and medicine [41].
Economic and health consequences	Recognizes the negative economic and health impacts of prolonged lockdowns and containment measures.
Scenario 3: Gradual decline in COVID-19 impact	
Medical progress and behavioral changes	Credits medical advancements, behavioral changes, and herd or vaccine immunity for the reduced impact.
Residual effect	Recognizes a residual effect of the pandemic but anticipates it will diminish over time.
Scenario 4: Adjustment for 2020/2021 excess mortality	
Disappearance of adverse effects	Assumes the adverse health effects of the pandemic will disappear with no long-term consequences.
Explicit excess mortality accounting	Assumes excess mortality from 2020 and 2021 will not average out and must be explicitly accounted for.
Unchanged baseline mortality	Assumes baseline mortality remains unchanged, discounting behavioral changes (e.g., reduced traffic fatalities, fewer influenza deaths due to hygiene, and quarantine measures).

Table 5. Overview of different scenario periods.

Scenario	Fitting Period	Forecast Period	Validation Period
1—Without COVID-effect	1990–2019	2020–2025	2022–2023
2—Full COVID-effect	1990–2021	2022–2025	2022–2023
3—Flattening COVID-effect	1990–2021	2022–2025	2022–2023
4—Excess mortality	1990–2019	2020–2025	2022–2023

Figure 5 presents outcomes for four scenarios across various countries and genders, focusing on 80-year-olds. We chose age 80 for illustration, but the overall structure is similar for other ages, albeit with less intense COVID-19 effects for younger age groups. Notable high-value outliers for Italy, the US, and the Netherlands in 2020 and 2021 indicate a pronounced impact of COVID-19. Different trend forecasts capture varying effects of COVID-19 on mortality rates. Scenario 1 represents a milder assumption, while Scenario 2 depicts a more severe projection. Future forecasts vary by country and age group, influenced by past behaviors and responses to the pandemic. Trend forecasts in different scenarios generally align with plausibility. Excluding 2020 and 2021 in Scenario 1 results in lower mortality rates, while adjusting for excess mortality in Scenario 4 leads to even lower rates, considering the population shift due to previous deaths. Scenario 2 with full COVID effect shows the highest mortality trend, which is particularly evident for older age groups. However, younger populations appear less affected. Scenario 3 starts similar to Scenario 2 but gradually decreases over time. Different countries show distinct trends, likely influenced by COVID's demographic and political impact. The Netherlands' observed rates in years 2022 and 2023 align with Scenario 2, whereas Italy and the US

show patterns more consistent with the flattening effect. German scenarios show less differentiation, aligning closely with observations, while Finland's forecast suggests lower mortality rates than observed.

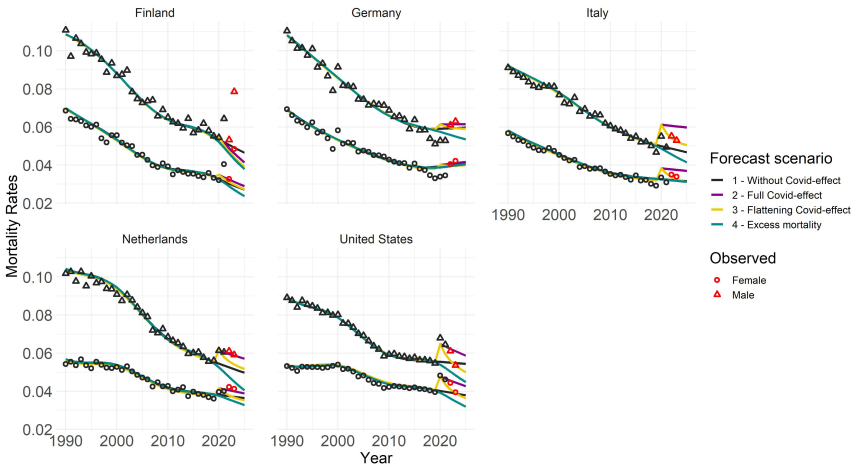


Figure 5. Trend forecasts for 80-year-olds across four distinct scenarios. Training data span from 1990 to a maximum of 2021, depending on the scenario, with forecasts projected up to 2025. Circles and triangles represent observed rates, with red markers indicating those used for validation purposes.

The heatmap (Figure 6) illustrates both a cross-country and country-specific perspective on the y-axis, while age groups are delineated on the x-axis. Colors within the heatmap indicate the normalized RMSE (NRMSE) values for the years 2022 and 2023 when compared with the observed mortality rates from STMF and processed in accordance with Appendix A.

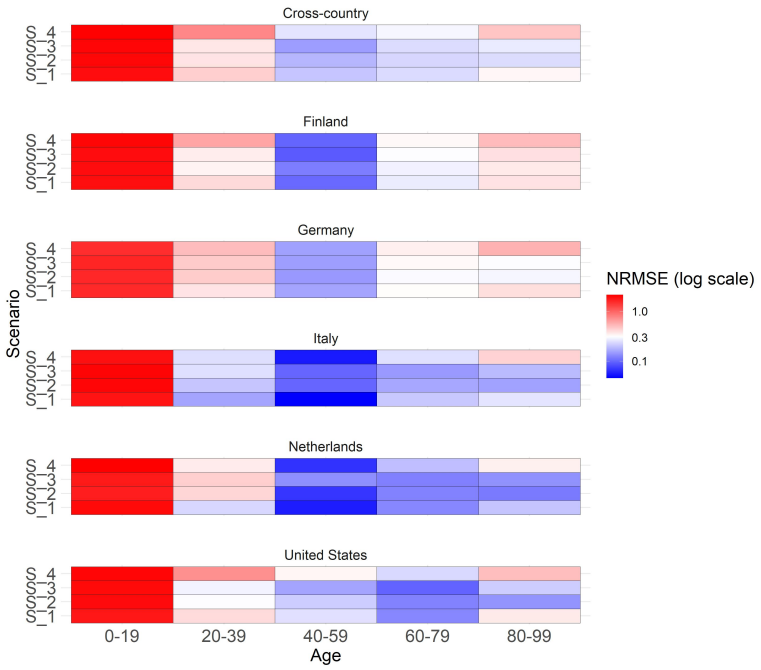


Figure 6. Heatmap showing the normalized RMSE of scenarios for extrapolated years 2022 and 2023 for males across different countries and age groups. The cross-country section presents a summary of the NRMSE across all countries.

NRMSE is calculated by dividing RMSE by the mean of observed mortality rates in a specific category with values ranging usually from 0 to 1. A value of 0 means perfect predictions, while 1 suggests predictions are as accurate as predicting the mean. Values above 1 often suggest that the model's performance may not be optimal. We prioritize the analysis on males since previous findings suggest a more pronounced emphasis on the COVID effect for this gender, although the overall patterns for females exhibit similarity. For males in the 20-year age brackets, the graph shows a generally good overall forecast accuracy, especially for ages over 20, as colors tend towards blue for middle and older ages, indicating smaller NRMSE values closer to 0 and suggesting better forecast accuracy. No clear scenario preference is evident in the cross-country view across all age groups. In general, Scenario 3 (flattening COVID effect) tends to perform well for middle-aged individuals and, in addition, also Scenario 4 (COVID full effect) for older ages. Substantial variations in scenario performance are observed across different countries and age groups. For younger age groups, Scenario 1 (no COVID effect) performs best in Italy and the Netherlands, where substantial COVID impact was observed. Middle-aged groups demonstrate similarly high performance across all scenarios. Older age groups show stronger scenario differences with a clear preference for Scenarios 2 and 3, indicating better fit. Scenarios do not perform well for those under 19, possibly due to the unique characteristics and weak impact of COVID in this age group. Appendix D contains the same graph with individual ages instead of grouped age buckets for a more detailed overview.

4. Conclusions

To summarize, this research work focused on addressing the challenge of capturing the mortality-related extreme event at the edge of a time series—in particular, COVID-19's effect on future mortality forecasting.

The key findings of our research include identifying the GAM within APC framework as the most effective method for forecasting future mortality rates across multiple countries. This innovative approach, utilizing a smoothed second-order spline, surpasses traditional stochastic models (e.g., LC) and machine learning techniques (e.g., GBM) in predictive accuracy. By applying the GAM-APC model to data from Germany, Finland, Italy, the Netherlands, and the United States, the study provides valuable cross-country and multi-populational insights into mortality trends. This enables the capture of both universal mortality patterns and country-specific variations, offering a comprehensive understanding of global and localized mortality dynamics. The research develops and evaluates four pandemic impact scenarios (ranging from mild to severe) to forecast the impact of COVID-19 on future mortality rates. It concludes that scenarios assuming a diminishing impact of the pandemic over time are the most consistent, especially for middle-aged and elderly populations.

To ensure a rigorous assessment, these scenarios and their underlying assumptions were thoroughly evaluated and discussed in collaboration with epidemiological experts. This approach including the content of scenarios aligns with existing literature and enhances the credibility of the forecast analysis [42].

Overall, this work contributes to the existing literature by introducing traditional, enhanced, and novel models, comparing different approaches and providing insights into future trends in mortality rates while considering the impact of COVID-19 in a cross-country context. The specific contribution of the GAM approach with the APC framework in this research lies in its novel application for mortality trend forecasting, particularly incorporating the impact of COVID-19 in a multi-populational cross-country fashion.

Despite the current waning impact of COVID-19, it is crucial to acknowledge the lasting importance of historical data and the persisting uncertainties that lie ahead. These factors emphasize the need for ongoing attention in the years to come. It is important to acknowledge that the concept and methodology utilized in this study extend beyond COVID-19, encompassing other events that occur at the edges of time series data. Looking ahead to future research directions, the GAM with APC framework has a promising

potential for expanding the feature set by the inclusion of socioeconomic status, income, and education as additional factors, allowing for a more complete understanding of mortality trends.

Author Contributions: Conceptualization, A.N., C.H. and S.P.; methodology, A.N., C.H. and S.P.; formal analysis, A.N.; writing—original draft preparation, A.N.; writing— review and editing, A.N., C.H. and S.P.; visualization, A.N.; supervision, C.H.; funding acquisition, A.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original data and the code are available at https://doi.org/10.5281/zenodo.13905807.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Data Preparation

This section discusses the methodology employed to enrich existing mortality data obtained from [16], focusing on the number of deaths and population size for recent years absent in the [15] dataset. The primary challenge addressed is the aggregation of data into rough age categories, while the study requires a metric age scale.

The methodology involves several steps. Firstly, weekly population sizes are derived from the mortality dataset followed by extrapolation to annual levels. Using mortality rates and death counts, the weekly population size can be calculated. These weekly data are then aggregated to annual figures. Similarly, weekly death counts are summed to obtain annual totals. To construct annual death counts and populations for individual ages for the aforementioned years, specific procedures were applied, as described below. The methodology ensures that the derived data align with observed mortality patterns within each age group.

Once the weekly population is extrapolated to the annual level by multiplying by a factor of 52, the approach leverages cohort-wise population patterns from previous years (2015–2019) and assumes a similar age distribution for 2020–2023. The initial population course, i.e., for 2020, is created by shifting the population size pattern of 2019 one year forward. This shift leads to an initial gap at age 0 in 2020, which is linearly extrapolated based on data from 2018 and 2019. The resulting population values are adjusted to match observed data within age groups.

A three-stage approach is employed to distribute death counts from grouped to metric age scale on an annual basis. Firstly, averaged weights for each age in each age bucket $[l, u]$ are computed based on data from the previous five years (2015–2019): $w_{\{l,u\}} = \frac{D_{\{l,u\}}}{\frac{1}{u-l+1} \cdot D_{[l,u]}}$. Secondly, these weights are applied to the averaged death counts in each age group to correct for deviations from the mean: $D^*_{\{l,u\}} = (\frac{1}{5} \sum_{j=2015}^{2019} w_{\{l,u\},j}) \cdot \frac{1}{u-l+1} \cdot D_{[l,u],2020}$. Finally, the corrected death counts are adjusted to ensure equal counts in both grouped and metric versions within each age group: $D_{\{l,u\},2020} = k_{\{l,u\}} \cdot D^*_{\{l,u\}}$, with $k_{\{l,u\}} = \frac{D_{[l,u],2020}}{\sum_{i=l}^{u} D^*_{i,2020}}$.

The resulting mortality rates are computed by dividing death counts by population size for each individual age and subpopulation. These enriched mortality data are used to impute the [15] dataset for the years 2020–2023. The same procedure is applied across all subpopulations and missing years.

Appendix B. Details on Gradient Boosting Machine

Gradient Boosting is another form of an ensemble learner that is based on the weighted combination of weak predictive learners such as Decision Trees, usually outperforming Random Forest. The model is built stepwise and optimized by a differentiable loss function,

minimizing the in-sample loss [43]. It builds the model stepwise, like other boosting methods, and generalizes by allowing optimization of any differentiable loss function. Whereas multiple samples of the original training dataset are used to fit a separate decision tree to each one independent of the others and to combine all trees into a single predictive model in bagging, boosting grows the trees sequentially, meaning the information gained from the previous trees is used to grow the current one. This helps to overcome the major issue of training a single large Decision Tree by possibly resulting in an overfitting problem. The gradient boosting algorithm instead learns by constructing a new model based on the previous one and adding the ith base learner $h^{(i)}_{a,t,g,c}$:

$$\hat{q}^{(i)}_{a,t,g,c} = \hat{q}^{(i-1)}_{a,t,g,c} + \lambda_i h^{(i)}_{a,t,g,c} \tag{A1}$$

The model is improved in such a way that the current residual will be used as an outcome to fit a new Decision Tree and to add this into the originally fitted function with the notion to update the residuals. So, the gradient boosting algorithm fits the new predictor to the residual errors made by the previous predictor. The shrinkage parameter λ_i helps to run the process even slower, allowing for more trees and more detailed enhancement of the residuals. All parameters of the Decision Trees undergo optimization through the training of Poisson boosted trees, with the objective of minimizing the negative log-likelihood associated with the Poisson distribution, serving as the designated loss function. Overall, in contrary to the bagging methodology, each tree depends on the previous ones [44]. Even though the gradient boosting keeps on minimizing the errors, this can cause overfitting in cases where there is a lot of noise in the data and is computationally time and memory expensive, especially because trees are built sequentially (not in parallel as Random Forest does). Due to the high flexibility, the gradient boosting algorithm also tends to be harder to tune than Random Forest [43]. In this study, we specifically utilized LightGBM [45], employing Microsoft's library for implementing these models, which have demonstrated high accuracy in various scenarios [30].

Appendix C. Details on Hyperparameter Optimization

For hyperparameter optimization of the GBM, we used Hyperopt, a Python library that employs the Tree Parzen Estimator (TPE) algorithm. TPE is an efficient method that utilizes a probabilistic model to guide the search for optimal hyperparameters. The TPE workflow can be summarized as follows: First, TPE begins by randomly sampling a few hyperparameter combinations to create an initial set of observations, serving as the starting point for optimization. Next, TPE models the relationship between hyperparameter values and the performance metric (e.g., loss or accuracy), estimating the probability that a configuration will yield better results.

TPE then balances exploration, trying new configurations, and exploitation, focusing on promising configurations based on probabilistic models. It emphasizes configurations likely to lead to better results, similar to how gradient descent focuses on the gradient of the loss function. As TPE evaluates more configurations, it iteratively refines its probabilistic models, making more informed decisions.

By continually balancing exploration and exploitation, TPE efficiently navigates the hyperparameter space, eventually converging on an optimal set of hyperparameters for a given machine learning model. For a deeper understanding of the TPE algorithm and its practical application, refer to [46,47].

As for GAMs, each smooth term has associated smoothing parameters that control the trade-off between fit and smoothness. The mgcv package's gam() function automates the process of smoothing parameter optimization when the method is set to Restricted Maximum Likelihood (REML). After specifying the model, the mgcv package automatically selects the optimal smoothing parameters by maximizing the restricted likelihood, using an internal Newton–Raphson numerical optimization algorithm. The result is a model with

optimally chosen smoothing parameters that balance fit and smoothness according to the restricted likelihood criterion [24].

Appendix D. Additional Results

The analysis in Table A1 highlights the superior in-sample predictive performance of the two-step GBM and GAM models within the APC framework over traditional LC and APC models, with no clear preference between GBM and GAM, across various training periods for different countries.

Table A1. In-sample RMSE comparison for LC, APC, GBM, and GAM models. LC, APC, and GBM are fitted from 1950 to 2010 (Finland, Italy, Netherlands, US) and 1990 to 2010 (Germany). GAM is fitted from 1990 to 2015 for all countries.

Country	Female				Male			
	LC	APC	GBM	GAM	LC	APC	GBM	GAM
FIN	0.0045	0.0015	0.0035	0.0011	0.0072	0.0027	0.0066	0.0013
DE	0.0015	0.0022	0.0004	0.001	0.0021	0.0021	0.0007	0.001
ITA	0.0025	0.0021	0.001	0.001	0.0012	0.0025	0.0008	0.001
NLD	0.0019	0.0032	0.0014	0.001	0.0017	0.0015	0.0014	0.001
US	0.0014	0.0033	0.0005	0.0013	0.0017	0.0028	0.0004	0.0011

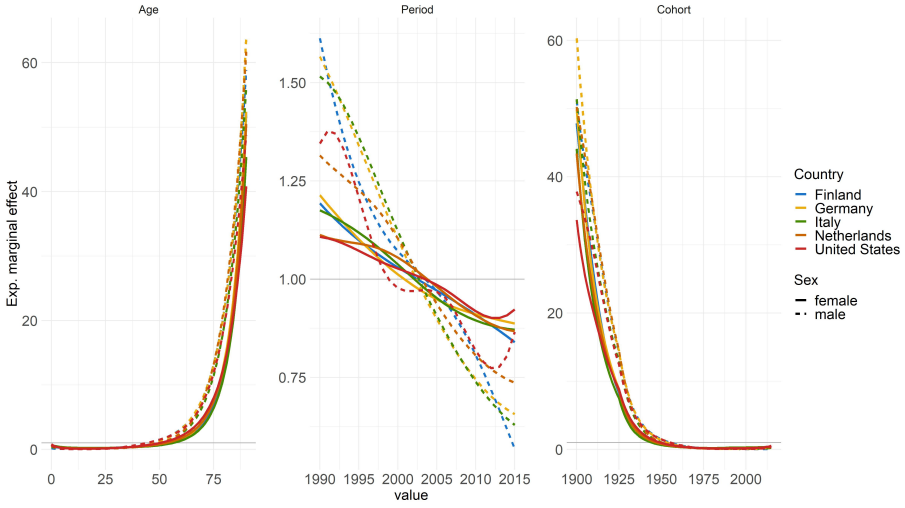


Figure A1. Estimated marginal effects of age, period, and cohort on mortality rates, differentiated by countries and genders. The horizontal lines represent the level of no effect. The GAM model was fitted for the years 1990–2015 and ages 0–90.

GAM enables the interpretation of exponential marginal effects, with age, period, and cohort being the components analyzed and differentiated by countries and gender. Notably, while the descending trend in period effect for women is relatively consistent and shallow across all countries, men exhibit a much steeper decline, indicating a stronger improvement in mortality rates over the years. There are noticeable increases in mortality rates for Italy and the US in recent years, particularly for US males, which may be associated with factors such as the opioid crisis.

The heatmap depicted in Figure A2 offers a detailed examination on an individual age basis for assessing the scenario analysis across the years 2022 and 2023.

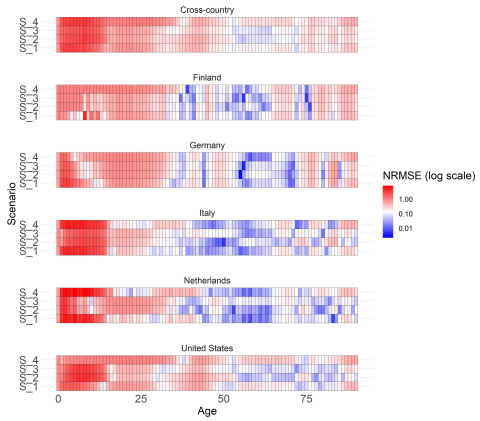


Figure A2. Heatmap showing the normalized RMSE of scenarios for extrapolated years 2022 and 2023 for males across different countries and individual ages.

References

1. Lee, R.D.; Carter, L.R. Modeling and forecasting US mortality. *J. Am. Stat. Assoc.* **1992**, *87*, 659–671.
2. Bjerre, D.S. Tree-based machine learning methods for modeling and forecasting mortality. *ASTIN Bull. J. IAA* **2022**, *52*, 765–787. [CrossRef]
3. Levantesi, S.; Pizzorusso, V. Application of machine learning to mortality modeling and forecasting. *Risks* **2019**, *7*, 26. [CrossRef]
4. Schnürch, S.; Kleinow, T.; Korn, R.; Wagner, A. The impact of mortality shocks on modelling and insurance valuation as exemplified by COVID-19. *Ann. Actuar. Sci.* **2022**, *16*, 498–526. [CrossRef]
5. Richman, R.; Wüthrich, M.V. A neural network extension of the Lee–Carter model to multiple populations. *Ann. Actuar. Sci.* **2021**, *15*, 346–366. [CrossRef]
6. Hastie, T.; Tibshirani, R. Generalized additive models: Some applications. *J. Am. Stat. Assoc.* **1987**, *82*, 371–386. [CrossRef]
7. Bray, I. Application of Markov chain Monte Carlo methods to projecting cancer incidence and mortality. *J. R. Stat. Soc. Ser. C Appl. Stat.* **2002**, *51*, 151–164. [CrossRef]
8. Clements, M.S.; Armstrong, B.K.; Moolgavkar, S.H. Lung cancer rate predictions using generalized additive models. *Biostatistics* **2005**, *6*, 576–589. [CrossRef]
9. Bashir, S.A.; Estève, J. Projecting cancer incidence and mortality using Bayesian age-period-cohort models. *J. Epidemiol. Biostat.* **2001**, *6*, 287–296.
10. Dodds, S.; Williams, L.J.; Roguski, A.; Vennelle, M.; Douglas, N.J.; Kotoulas, S.-C.; Riha, R.L. Mortality and morbidity in obstructive sleep apnoea–hypopnoea syndrome: Results from a 30-year prospective cohort study. *ERJ Open Res.* **2020**, *6*, 00057-2020. [CrossRef]
11. Clèries, R.; Ribes, J.; Esteban, L.; Martinez, J.M.; Borras, J.M. Time trends of breast cancer mortality in Spain during the period 1977–2001 and Bayesian approach for projections during 2002–2016. *Ann. Oncol.* **2006**, *17*, 1783–1791. [CrossRef] [PubMed]
12. Robben, J.; Antonio, K.; Devriendt, S. Assessing the impact of the COVID-19 shock on a stochastic multi-population mortality model. *Risks* **2022**, *10*, 26. [CrossRef]
13. Barigou, K.; Loisel, S.; Salhi, Y. Parsimonious predictive mortality modeling by regularization and cross-validation with and without Covid-type effect. *Risks* **2020**, *9*, 5. [CrossRef]
14. Clayton, D.; Schifflers, E. Models for temporal variation in cancer rates. II: Age–period–cohort models. *Stat. Med.* **1987**, *6*, 469–481. [CrossRef] [PubMed]
15. Human Mortality Database; University of California, Berkeley, CA, USA; Max Planck Institute for Demographic Research, Max Planck Society for the Advancement of Science e.V., Munich, Germany, 2024. Available online: https://www.mortality.org/ (accessed on 1 March 2024).
16. Short-term Mortality Fluctuations (STMF); University of California, Berkeley, CA, USA; Max Planck Institute for Demographic Research, Max Planck Society for the Advancement of Science e.V., Munich, Germany, 2024. Available online: https://www.mortality.org/ (accessed on 1 March 2024).
17. Brouhns, N.; Denuit, M.; Vermunt, J.K. A Poisson log-bilinear regression approach to the construction of projected lifetables. *Insur. Math. Econ.* **2002**, *31*, 373–393. [CrossRef]
18. Villegas Ramirez, A. Mortality: Modelling, Socio-Economic Differences and Basis Risk. Ph.D. Dissertation, City University London, London, UK, 2015.
19. Hobcraft, J.; Menken, J.; Preston, S. *Age, Period, and Cohort Effects in Demography: A Review*; Springer: Berlin/Heidelberg, Germany, 1985.
20. Currie, I.D.; Durban, M.; & Eilers, P.H. Smoothing and forecasting mortality rates. *Stat. Model.* **2004**, *4*, 279–298. [CrossRef]

21. Yan, J.; Guszcza, J.; Flynn, M.; Wu, C.S.P. Applications of the offset in property-casualty predictive modeling. *Casualty Actuar. Soc.-Forum* **2009**, *1*, 366–385.
22. Weigert, M.; Bauer, A.; Gernert, J.; Karl, M.; Nalmpatian, A.; Küchenhoff, H.; Schmude, J. Semiparametric APC analysis of destination choice patterns: Using generalized additive models to quantify the impact of age, period, and cohort on travel distances. *Tour. Econ.* **2022**, *28*, 1377–1400. [CrossRef]
23. Bauer, A.; Weigert, M.; Jalal, H. APCtools: Descriptive and Model-based Age-Period-Cohort Analysis. *J. Open Source Softw.* **2022** , *7*, 4056. [CrossRef]
24. Wood, S.N. *Generalized Additive Models: An Introduction with R*; CRC Press: Boca Raton, FL, USA, 2017.
25. Hyndman, R.J.; Khandakar, Y. Automatic time series forecasting: The forecast package for R. *J. Stat. Softw.* **2008**, *27*, 1–22. [CrossRef]
26. Bai, J.; Perron, P. Estimating and testing linear models with multiple structural changes. *Econometrica* **1998**, *66*, 47–78. [CrossRef]
27. Zeileis, A.; Kleiber, C.; Kramer, W.; Hornik, K. Testing and Dating of Structural Changes in Practice, Computational Statistics and Data Analysis. *Comput. Stat. Data Anal.* **2003**, *44*, 109–123. [CrossRef]
28. Ramirez Villegas, M.A.; Millossovich, P.; Kaishev, V. StMoMo: An R Package for Stochastic Mortality Modelling. 2016. Available online: Https://cran.r-project.org/web/packages/StMoMo/StMoMo.pdf (accessed on 25 July 2024).
29. Wood, S.N. *Generalized Additive Models: An Introduction with R*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2006; R package version 1.8-23, 2015. Available online: https://cran.r-project.org/web/packages/mgcv/index.html (accessed on 1 October 2023).
30. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
31. Bergstra, J.; Komer, B.; Eliasmith, C.; Yamins, D.; Cox, D.D. Hyperopt: A python library for model selection and hyperparameter optimization. *Comput. Sci. Discov.* **2015**, *8*, 014008. [CrossRef]
32. Hamilton, A.D.; Jang, J.B.; Patrick, M.E.; Schulenberg, J.E.; Keyes, K.M. Age, period and cohort effects in frequent cannabis use among US students: 1991–2018. *Addiction* **2019**, *114*, 1763–1772. [CrossRef]
33. Crimmins, E.M.; Shim, H.; Zhang, Y.S.; Kim, J.K. Differences between men and women in mortality and the health dimensions of the morbidity process. *Clin. Chem.* **2019**, *65*, 135–145. [CrossRef]
34. Trovato, F.; Lalu, N.M. Narrowing sex differentials in life expectancy in the industrialized world: Early 1970's to early 1990's. *Soc. Biol.* **1996**, *43*, 20–37. [CrossRef]
35. Rosella, L.C.; Calzavara, A.; Frank, J.W.; Fitzpatrick, T.; Donnelly, P.D.; Henry, D. Narrowing mortality gap between men and women over two decades: A registry-based study in Ontario, Canada. *BMJ Open* **2016**, *6*, e012564. [CrossRef]
36. Perls, T.T.; Fretts, R.C. Why Women Live Longer than Men-What gives women the extra years? *Sci. Am.* **1998**, *2*, 100–103.
37. Sudre, C.H.; Murray, B.; Varsavsky, T.; Graham, M.S.; Penfold, R.S.; Bowyer, R.C.E.; Pujol, J.C.; Klaser, K.; Antonelli, M.; Canas, L.S.; et al. Attributes and predictors of Long-COVID. *Nat. Med.* **2021**, *27*, 626–631. [CrossRef]
38. Kunzler, A.M.; Röthke, N.; Günthner, L.; Stoffers-Winterling, J.; Tüscher, O.; Coenen, M.; Rehfuess, E.; Schwarzer, G.; Binder, H.; Schmucker, C.; et al. Mental burden and its risk and protective factors during the early phase of the SARS-CoV-2 pandemic: Systematic review and meta-analyses. *Glob. Health* **2021**, *17*, 1–29. [CrossRef] [PubMed]
39. Polack, F.P.; Thomas, S.J.; Kitchin, N.; Absalon, J.; Gurtman, A.; Lockhart, S.; Perez, J.L.; Pérez Marc, G.; Moreira, E.D.; Zerbini, C.; et al. Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine. *N. Engl. J. Med.* **2020**, *383*, 2603–2615. [CrossRef] [PubMed]
40. Johns Hopkins University. COVID-19 Dashboard. 2021. Available online: https://coronavirus.jhu.edu/map.html (accessed on 1 October 2024).
41. Boudourakis, L.; Uppal, N. Decreased COVID-19 mortality—A cause for optimism. *JAMA Intern. Med.* **2021**, *181*, 478–479. [CrossRef] [PubMed]
42. Telenti, A.; Arvin, A.; Corey, L.; Corti, D.; Diamond, M.S.; Garcìa-Sastre, A.; Garry, R.F.; Holmes, E.C.; Pang, P.S.; Virgin, H.W. After the pandemic: Perspectives on the future trajectory of COVID-19. *Nature* **2021**, *596*, 495–504. [CrossRef] [PubMed]
43. Hastie, T.; Tibshirani, R.; Friedman, J.H. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*; Springer: New York, NY, USA, 2009; Volume 2.
44. Deprez, P.; Shevchenko, P.V.; Wüthrich, M.V. Machine learning techniques for mortality modeling. *Eur. Actuar. J.* **2017**, *7*, 337–352. [CrossRef]
45. Oram, E.; Dash, P.B.; Naik, B.; Nayak, J.; Vimal, S.; Nataraj, S.K. Light gradient boosting machine-based phishing webpage detection model using phisher website features of mimic URLs. *Pattern Recognit. Lett.* **2021**, *152*, 100–106. [CrossRef]
46. Bergstra, J.; Yamins, D.; Cox, D. Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. In Proceedings of the International Conference on Machine Learning, Atlanta, GA, USA, 17–19 June 2013.
47. Bergstra, J.; Bardenet, R.; Bengio, Y.; Kégl, B. Algorithms for hyper-parameter optimization. In Proceedings of the 25th Annual Conference on Neural Information Processing Systems 2011, Granada, Spain, 12–14 December 2011.

Part III.

Regional Mortality Support with Hierarchical Boosting

8. Local and global mortality experience: A novel hierarchical model for regional mortality risk

Contributing article

Nalmpatian, A., Heumann, C., Alkaya, L., and Jackson, W. (2026). Local and global mortality experience: A novel hierarchical model for regional mortality risk. *PLoS One* (in press).

Copyright information

Author contributions

The hierarchical Gradient Boosting Machine (GBM) model for leveraging data-rich countries to improve mortality modeling in data-scarce regions was conceptualized by Asmik Nalmpatian. She designed and implemented the two-step global-local modeling framework. Levent Alkaya supported the development and optimization of the software components, while William Jackson provided key access to regional datasets and country-specific actuarial knowledge. Asmik Nalmpatian curated and harmonized the data, conducted all analyses, and drafted the manuscript. Christian Heumann contributed to the formulation of the hierarchical modeling structure and reviewed all stages of the work. William Jackson assisted in interpreting the results and revised the manuscript together with the full author team. The work was presented by Asmik Nalmpatian, William Jackson, and Levent Alkaya at the Annual Meeting of the German Actuarial Society (DAV) and the German Society for Insurance and Financial Mathematics (DGVFM), held from 26 to 28 April 2023 in Dresden, Germany.

Local and global mortality experience: A novel hierarchical model for regional mortality risk

Asmik Nalmpatian[1*], Christian Heumann[1], Levent Alkaya, William Jackson

1 Department of Statistics, LMU Munich, Munich, Bavaria, Germany

* asmik.nalmpatian@campus.lmu.de

Abstract

Accurate mortality risk assessment is critical for decision-making in life insurance, healthcare, and public policy. Regional variability in mortality, driven by diverse local factors and inconsistent data availability, presents significant modeling challenges. This study introduces a novel hierarchical mortality risk model that integrates global and local data, enhancing regional mortality estimation across diverse regions. The proposed approach employs a two-stage process: first, a global Light Gradient Boosting Machine model is trained on globally shared features; second, region-specific models are developed to incorporate local characteristics. This framework outperforms both purely local models and standard imputation techniques, particularly in data-scarce regions, by leveraging global patterns to improve generalization. The model is computationally efficient, scalable, and robust in handling missing values, making it adaptable for other domains requiring integration of multi-regional data. This method enhances predictive accuracy across various regions and provides a more reliable approach for mortality risk estimation in data-scarce environments.

Introduction

Mortality risk assessment plays a crucial role in various sectors, including life insurance, healthcare, and public policy. Reliable estimates of mortality rates are essential for strategic planning, policy formulation, and ensuring the financial stability of life insurance systems. However, accurately estimating mortality risk presents an essential challenge due to the diverse and dynamic nature of regional data availability and factors that affect mortality rates.

Hierarchical models have been utilized in mortality studies to account for variations at different levels, including regional, individual and national. Originally developed in fields like education, sociology, and demography, these models have gained significant traction in public health and epidemiology. By generalizing the classical pooling of group estimates, hierarchical or multilevel models offer a flexible framework for analyzing mortality data [50]. This flexibility allows researchers to better understand and interpret the complex factors influencing mortality rates across different populations.

Existing models in hierarchical mortality modeling include Bayesian approaches, generalized linear models, and machine learning (ML) techniques. Bayesian hierarchical models estimate mortality rates by incorporating prior distributions to handle uncertainty [48]. Generalized linear models, including multilevel Poisson regression, have been applied to mortality data to account for overdispersion and hierarchical

structure [49]. Although the existing literature predominantly employs random effects for both methodologies, our approach diverges by sequentially processing the residuals. Recent studies have also explored ML methods such as random forests and gradient boosting for COVID-19 mortality modeling [58].

Studies have highlighted the importance of balancing global patterns with local specifics in mortality modeling to ensure both generalizability and relevance [56, 57]. However, the availability of mortality data varies widely across regions, posing challenges for model accuracy and reliability [54]. Poisson regression is commonly used for modeling count data, including mortality rates [47], whereas Light Gradient Boosting Machine (LightGBM) has been recognized for its efficiency and accuracy in handling large datasets, making it suitable for hierarchical mortality modeling [52].

Existing mortality models often struggle to balance global trends and local variations, leading to models that either overgeneralize or fail to capture region-specific nuances. Furthermore, inconsistent and sparse data availability across regions intensifies these challenges, reducing the reliability of predictions, especially in data-scarce environments [54]. Current approaches often suffer from overdispersion [46] or are computationally inefficient when handling large datasets [53] or missing data [53]. These limitations underscore the need for a more flexible and scalable solution.

To address these challenges, this study introduces a novel hierarchical mortality modeling approach that integrates both global and local data. By using a two-stage process, our model first captures global patterns through a LightGBM model with a Poisson regression objective and then refines these predictions with region-specific models that account for local characteristics. While the first step includes shared variables that apply to all countries, such as age and gender, the country-specific models capture unique regional characteristics by incorporating additional region-specific factors, such as lifestyle habits and environmental conditions. This method markedly improves predictive performance, particularly in data-sparse regions, by leveraging global insights while remaining adaptable to the unique conditions of each region. Additionally, the model is computationally efficient, scalable, and capable of handling missing values, making it superior to traditional pooling methods. Beyond mortality risk estimation, this hierarchical modeling framework is applicable to other domains requiring multi-regional data integration, such as public health planning, epidemiological forecasting, and financial risk assessments. Its ability to generalize well across different regions makes it particularly valuable in scenarios where data sparsity or inconsistency is a common obstacle.

The structure of this paper is as follows: Section 2 provides a brief overview of our database and Section 3 presents our proposed methodology in detail. Section 4 examines the effectiveness of our methodology by presenting and discussing the results. Finally, Section 5 concludes by summarizing the main findings and suggesting research and industry perspectives.

Database

Data for the study was collected in a pseudonymised form from eight different operating units of a global primary insurance company, each representing a distinct country. Data privacy regulations prohibit the disclosure of these countries' names, keeping the focus on the technical aspects of the model evaluation and comparison, rather than on potential privacy breaches. The chosen organizations were based on two key factors: having relevant data available of high quality and representing diverse geographic regions.

The dataset includes policy data that remained active during this period, even if initially issued before the earliest year studied. In total, the dataset encompasses nearly

10 million life-years of exposure and close to 10,000 recorded insurance claims (=deaths).

The data underwent analysis in an aggregated form, grouped into $N = 16.689.304$ unique combinations of feature values. Specifically, the feature set $X_{i,j}$, where group i ranges from 1 to N and j ranges from 1 to 8 - representing the eight countries, consists of a total of 26 features. Among these features, 9 are global, and up to 17 are local features, encompassing information about policyholders, insurance policies, and claims. Given these potential risk factors, our target is to model the number of deaths $D_{i,j}$ in relation to the life years of risk exposure $E_{i,j}$. To facilitate model training and evaluation, an artificial variable was constructed before aggregating to create an 80-20 train-test split, ensuring that all unique combinations are adequately represented in both the training and test sets.

Table 1 provides an overview of $D_{i,j}$, $E_{i,j}$, and the total number of years included $T_{i,j}$ for group i in country j, thereby facilitating a comprehensive understanding of the dataset's key characteristics and distribution.

Table 1. Overview of death counts D_j, exposure in life years E_j, unique feature combination N_j, and observed years T_j for each country j.

Country j	D_j	E_j	N_j	T_j
1	1699	1295299	1880792	2013–2020
2	1291	1686299	2190943	2010–2020
3	494	815795	1868691	2010–2020
4	1225	1347150	1572539	2017–2020
5	1816	1825901	4825792	2016–2020
6	2132	1548157	3852306	2016–2020
7	458	498560	207951	2017–2020
8	297	99473	290290	2015–2020
Total	9412	9116634	16689304	2010–2020

Methodology

The foundation of our approach is rooted in the Cox Proportional Hazards model (Cox PH), a class of survival models in statistics that aligns with our objective of estimating mortality rates [2]. To simplify the complexity of Cox PH model calculations, we leveraged the connection between Cox PH and a Poisson Generalized Linear Model (GLM). Assuming piecewise constant hazard rates over time, the likelihood of the Cox PH model coincides with the likelihood of the Poisson GLM when we employ $log(E_{i,j})$ as an offset parameter, as detailed by [29] who noted, "we do not assume [the Poisson model] is true, but simply use it as a device for deriving the likelihood". Independent of [29], [45] published a similar insight, emphasizing that the piece-wise proportional hazards model is equivalent to a specific Poisson regression model.

Our primary goal is to accurately evaluate mortality rates. We aim to estimate the conditional expectation of death counts, denoted as $D_{i,j}$, given the available information summarized in the feature set $X_{i,j}$ and the exposure in life years at risk $E_{i,j}$. Assuming that $D_{i,j} \overset{\text{ind.}}{\sim} \text{Poisson}(\mu_{i,j} \cdot E_{i,j})$, the expectation according to the Poisson distributional assumption is:

$$\mathbb{E}[D_{i,j}|X_{i,j}, E_{i,j}] = \mu_{i,j} \cdot E_{i,j} = \exp\left(X_{i,j}^{\intercal}\beta_j\right) \cdot E_{i,j}$$

The Poisson log-likelihood is defined:

$$l(\beta_j|X_{i,j}, D_{i,j}) = \sum_{i=1}^{N_j} \left(D_{i,j} \cdot \log(\hat{D}_{i,j}) - \hat{D}_{i,j}) \right)$$

where $D_{i,j}$ denotes the observed death counts, $\hat{D}_{i,j} = \hat{\mu}_{i,j} E_{i,j}$ denotes the predicted death counts, and β_j is the parameter vector.

This formulation assumes that deaths follow a Poisson distribution. An advantage of simplifying the Cox PH model into a Poisson GLM is its adaptability to the ML realm, requiring optimization using Poisson log-likelihood and the ability to define an offset or observation weights. ML models, which generally do not assume specific (i.e. additive) relationships between features and targets, can leverage this flexibility:

$$E[D_{i,j}|X_{i,j}] = \mu_{i,j} \cdot E_{i,j} = \exp(f(X_{i,j})) \cdot E_{i,j}$$

This transition from GLMs to ML models offers additional benefits, including integrated variable selection mechanisms and the ability to capture interactions without explicit specification.

To implement this approach, we employ the LightGBM algorithm [52], a popular ML technique based on boosting. LightGBM iteratively builds an ensemble of decision trees to model the relationship between features and the target variable, optimizing the model to minimize the negative log-likelihood of the Poisson distribution [25]. Trees are fit to residuals derived from the loss function, and the model is updated iteratively to minimize this loss. The prediction is formulated as a linear combination of the base learners:

$$\mu_{i,j} = \exp(f(X_{i,j}|\boldsymbol{\theta})) = \exp\left(\sum_{k=0}^{K} \theta_k \cdot u_k(X_{i,j})\right)$$

where θ_k is the weight of the k-th tree, and $u_k(X_{i,j}) = \sum_{l \in V_k} b_l \cdot \mathbb{I}[X_{i,j} \in R_l]$ represents the tree associated with V_k as set of leaves of the k-th tree, b_l as the predicted value in the l-th leaf, and R_l as the region defined by disjoint partitions of the training set associated with the l-th leaf [28]. LightGBM uses a leaf-wise growth strategy, splitting the leaf with the highest loss reduction first, and adopts a histogram-based algorithm to improve the efficiency and speed of building decision trees. This approach results in efficient and accurate models, particularly for datasets with complex or imbalanced relationships. Mechanisms we employ to control overfitting and ensure robust performance are detailed in S2 Appendix.

Two-step model: To distinguish between local and global features and ensure high accuracy in each country, we propose a Two-step model approach. This approach involves two distinct modeling steps:

Step 1: Global model: The first model identifies global patterns and uses a training set that includes data from all countries, focusing solely on "global" factors. These global factors are those where data across countries is comparable, such as age. In contrast, factors like postal code, which lack comparability between regions, are excluded.

Step 2: Specialized Local model: In the second step, we calculate one Local model per country, totaling eight Local models. Each Local model takes the output of the Global model and adjusts it to the specific circumstances of the respective country. Specialized Local models use all global factors plus the country-specific local factors. The distinction of the feature set into global and local features is based on the availability of data across countries as well as domain-specific expert knowledge.

This approach combines the estimates from both the global and specialized Local models as illustrated in Fig 1.

Mathematically, we can express the process of estimating death counts for a policy with given factors as follows:

$$E[D_{i,j}|X_{i,j}] = \mu_{i,j} \cdot E_{i,j} = q(X_{i,j}^{\text{global}}) \cdot h_j(X_{i,j}^{\text{all}}) \cdot E_{i,j}$$

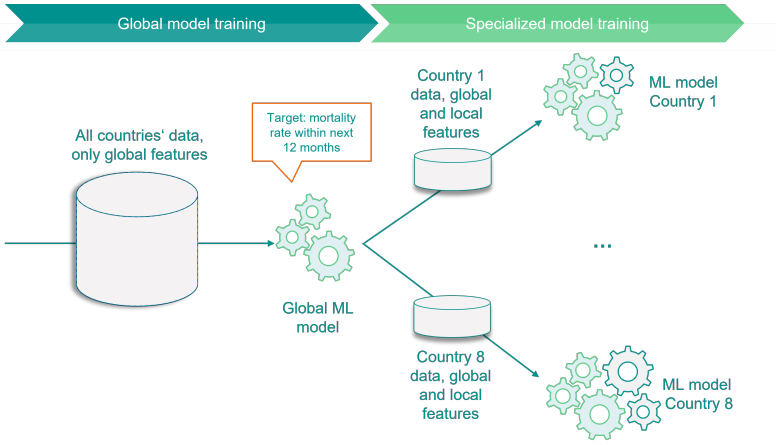


Fig 1. Qualitative illustration of proposed methodology. Gearwheels illustrate the features.

where $D_{i,j}$ represents the expected number of deaths given a set of features $X_{i,j}$ for group i and country j; $q(\cdot)$ represents the Global model's prediction function; $h_j(\cdot)$ represents the Local model's prediction function for country j; $X_{i,j}^{\text{global}}$ represents a set of factor values for group i and country j, containing only global factors; $X_{i,j}^{\text{all}}$ represents a set of factor values for country j, containing both global and local factors.

In technical terms, the predicted mortality rates from the first Global model are used to initialise the second specialized Local model. Accordingly, the model continues to work on the resulting residuals and iteratively optimises the second model - but now with the broader, localised data set. The final predicted number of deaths results from the multiplication of the predictions from the Global model (first step), the predictions from the specialised Local model (second step) and the exposure. The following derivation shows that the multiplication is justified by the nature of the boosting algorithm and the exponentiation by the log link of the Poisson distribution:

$$\begin{aligned}\mu_{i,j} &= \exp\left(\sum_{k=1}^{K} \theta_k \cdot u_k(X_{i,j})\right) \\ &= \prod_{k=1}^{K} \exp\left(\theta_k \cdot u_k(X_{i,j})\right) \overset{g := \exp(\theta \cdot u(X))}{=} \underbrace{\prod_{k=1}^{P} g_k(X_{i,j})}_{\text{Global model} = q(.)} \cdot \underbrace{\prod_{l=P+1}^{K} g_l(X_{i,j})}_{\text{Local model} = h_j(.)}\end{aligned}$$

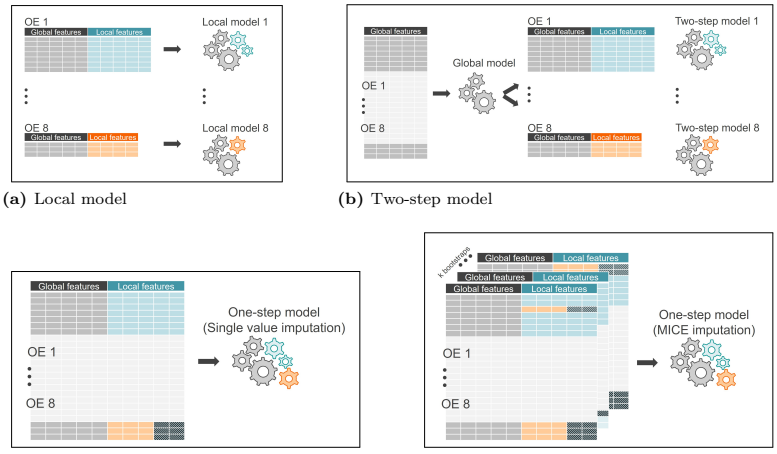
Splitting the modeling into two steps offers the advantage of cleanly separating effects into local and global categories. It also optimizes model performance for each market by tailoring the model to local patterns while allowing knowledge sharing across countries via the Global model. Additionally, when onboarding a new country, we can choose to retain the existing Global model and calculate a new Local model for this new country.

We employ Microsoft's ML library "LightGBM" for implementing these models, which have demonstrated high accuracy in various scenarios. As the software does not

allow the inclusion of an offset, we utilize observed mortality rates as the target variable, thus the death counts are scaled by exposure $D_{i,j}/E_{i,j}$ and exposure $E_{i,j}$ is used as weights, a method demonstrated to be mathematically equivalent in the Poisson case by [33]. These residuals $R_{i,j}$ represent the deviation of the observed deaths from the expected deaths $\hat{D}_{i,j}^{\text{global}}$ predicted by the first step, and are calculated as follows: $R_{i,j} = \frac{D_{i,j}}{\hat{D}_i^{\text{global}}}$. In the second step, these residuals serve as the target variable for further modeling. The new weights for this step are the expected deaths from the first step, $\hat{D}_i^{\text{global}}$. It is important to note that in the second step, we use the complete feature set of a single country, whereas in the first step, we utilize pooled data with global features only. Details on prediction calibration are provided in S3 Appendix.

Benchmarking results

Our objective is to benchmark the proposed methodology against three other approaches using specific evaluation metrics. This aims to determine the predictive performance and computational efficiency of the proposed model compared to the alternatives. All these methods are based on the model specification proposed in the previous section, where death counts are estimated in relation to exposure using the ML model LightGBM, optimizing the Poisson log-likelihood assumption. The differences among these methods are outlined below and illustrated in Fig 2:



(a) Local model

(b) Two-step model

(c) One-step model with single value imputation

(d) One-step model with MICE

Fig 2. Comparison of benchmarked models and their frameworks. Gearwheels represent features: grey for global features, blue and orange for local features specific to different countries, and patterned dark cells indicate missing values.

1.Local models for individual countries: For each country, we take this country's data and run the model separately. This is, of course, only applicable if we have enough claims and exposure available for a given country as a solid foundation for training. The information contained in the each other countries about certain features and their correlation patterns to mortality rates remain unseen for each model.

2. Two-step approach: As detailed in the previous section, this approach combines

global features in the first step model, using common features across countries. In the second step, a Local model is trained to capture also each country's specificities based on residuals from the first step.

3. Global one-step with single value imputation: All datasets from different countries are combined in this early data fusion technique. The discrepancy in feature sets and values across countries results in missing blocks, as shown in Table 2.

Country	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	F16	F17	F18	F19	F20	F21	F22	F23	F24	F25	F26
1	0	0	0	0	0	0	0	0	0	0	0	33	0	72	0	0	0	0	0	0	0	72	72	72	72	0
2	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	48	48	48	48	48	48
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	33	33	0	0	0	0	33	5	6	0	0
4	0	0	0	0	0	0	0	0	0	0	0	28	0	0	0	28	0	28	0	0	0	28	0	28	28	28
5	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	72	0	100	100	0
6	0	0	0	0	0	0	0	0	0	0	0	37	0	0	0	62	0	0	0	0	62	58	56	0	62	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	2	8	0	6	0	8	8	8	8	8	8
8	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	6	0	0	0	6	6	0	0	6	6	6

Table 2. Percentage of missing values in each feature by country

For all three model types, missing values are imputed based on feature type: categorical features receive "Missing" and metric features receive "-1". This approach retains information from non-missing values and identifies missing values during interactions for local features. In contrast, global features are free from missing values due to the design of the data collection process.

In cases where a local model cannot be trained due to small data size, the One-Step approach may be the only viable option, but it results in missing blocks that must be imputed. The Two-Step model offers a valuable alternative by providing flexibility: if a local feature is entirely missing, it can be dropped, similar to local models, while global features are retained based on global patterns. For partially missing local features, single value imputation is applied, and the researcher has the option to drop or keep the imputed feature for a specific country. We chose to retain all features that are not completely missing within a country to ensure no information is lost.

4. Global one-step with bootstrapped multiple imputation: Similar to the previous approach, this method involves early data fusion by combining datasets from all countries. In this case, we use Bootstrapped Multiple Imputation with Decision Tree as imputation technique for missing values that arise due to the synthetic dataset creation. The procedure is as follows:

- First draws k bootstrap samples from the combined dataset including missing values.
- Fit a classification or regression tree by recursive partitioning, variable by variable.
- After fitting a tree for the missing value based on the other values of the variable from the corresponding leaf, a value is randomly drawn.

This ensures that we can use it properly for multiple imputation, so that we are inducing some variation and not just the randomness in the leaf. The implementation was done in Python [30] with an adapted version of IterativeImputer [43], using 4 bootstrap samples and 2 imputations iterations each. We refer to [32] for further algorithm details. The number of iterations was determined based on a trial-and-error approach, as higher numbers had no significant impact on the final model results due to the dataset's size. Based on each dataset resulting from the bootstrapped iteration, we trained the proposed model and finally pooled the eight predictions by averaging.

Evaluation criteria: To evaluate our proposed methodology, we place a strong emphasis on two critical dimensions: predictive accuracy and computational efficiency.

To gauge the predictive performance of our models, we employ two essential metrics: Root Mean Square Error (RMSE) for both in-sample and out-of-sample assessments. For a given country j it is calculated as follows:

$$RMSE_j = \sqrt{\sum_{i=1}^{N_j} (\hat{D}_{i,j} - D_{i,j})^2}$$

Additionally, we utilize the Poisson log-likelihood, which serves a dual role as a loss function and evaluation metric:

$$l_j = \sum_{i=1}^{N_j} \left(D_{i,j} \cdot \log(\hat{D}_{i,j}) - \hat{D}_{i,j} \right)$$

In the equations, $\hat{D}_{i,j} = \hat{\mu}_{i,j} \cdot E_{i,j}$ represents the predicted, while $D_{i,j}$ the observed death counts. The in-sample metrics allow us to examine how well the model fits the training data. On the other hand, the out-of-sample metrics serves as a litmus test for the model's ability to generalise to new, unseen data.

A higher log-likelihood and lower RMSE signify a closer fit between the model and the data, indicating superior performance. Conversely, a lower log-likelihood and higher RMSE are indicative of a less suitable model for the given data.

We consider runtime, memory usage, and storage requirements to evaluate the computational efficiency of our models, aiming for lower values to enhance their practical utility. These criteria offer a comprehensive assessment of our models' performance in estimating mortality rates and pricing life insurance.

Outcomes: This section details the benchmarking process for all four models, focusing on key metrics for performance and efficiency assessment. We evaluated the models using multiple metrics, including train and test RMSE and log-likelihood. Although RMSE is reported, log-likelihood is more reliable due to the distributional assumptions of the data. Additionally, we assessed computational efficiency through run time (seconds), memory consumption (megabytes), and storage space of the model object (kilobytes).

In Tables 3 and 4 we present the results exemplarily for country 5 and 7, and in S1 Appendix an overview of all countries as well as the cross-country results. Each table provides an insight into the performance of the four benchmarked models, highlighting their strengths and weaknesses in various aspects. For ease of interpretation, we have used colour coding in dark grey to identify the best model within each row, based on the respective metric. The comparison is based on original values, before rounding for readability reasons.

Table 3. Performance evaluation for country 5

Metric	Local model	Two-step model	One-step model (Single Value)	One-step model (MICE)
RMSE (Train)	1.990×10^{-2}	1.979×10^{-2}	2.009×10^{-2}	2.126×10^{-2}
RMSE (Test)	1.709×10^{-2}	1.706×10^{-2}	1.709×10^{-2}	1.811×10^{-2}
Log Likelihood (Train)	-1.110×10^{4}	-1.069×10^{4}	-1.295×10^{4}	-1.315×10^{4}
Log Likelihood (Test)	-3.429×10^{3}	-3.399×10^{3}	-3.938×10^{3}	-3.998×10^{3}
Runtime (Sec)	1.370×10^{4}	3.970×10^{2}	-	-
Memory (MB)	2.998×10^{3}	1.663×10^{2}	-	-
Storage (KB)	2.174×10^{6}	2.162×10^{6}	-	-

Table 4. Performance evaluation for country 7

Metric	Local model	Two-step model	One-step model (Single Value)	One-step model (MICE)
RMSE (Train)	5.358×10^{-2}	5.061×10^{-2}	5.736×10^{-2}	5.847×10^{-2}
RMSE (Test)	3.542×10^{-2}	3.604×10^{-2}	3.983×10^{-2}	3.714×10^{-2}
Log Likelihood (Train)	-1.821×10^{3}	-1.469×10^{3}	-2.439×10^{3}	-2.682×10^{3}
Log Likelihood (Test)	-5.615×10^{2}	-5.529×10^{2}	-5.682×10^{2}	-5.693×10^{2}
Runtime (Sec)	9.144×10^{2}	1.518×10^{1}	-	-
Memory (MB)	3.041×10^{2}	1.983×10^{2}	-	-
Storage (KB)	9.816×10^{4}	9.376×10^{4}	-	-

Our Two-step modeling approach demonstrates the best predictive performance for nearly all countries, as evidenced by our comprehensive evaluation. This method outperforms Local models in most cases and shows significant advantages over the MICE method. Detailed results can be found in the tables and figures, highlighting the effectiveness of our approach.

The Two-step model shows the most substantial improvements for smaller countries (e.g., countries 7 and 8), compared to larger countries (e.g., countries 4 and 5). This is particularly evident in the test log-likelihood improvements from Local models to the Two-step model. By leveraging a Global model in the first step, we protect local specifics while enhancing the generalization capability, especially for smaller datasets.

Our research compares also one-step models, including single value imputation and MICE, with the proposed two-step approach. The findings consistently show that one-step models underperform and demand substantial computational resources. Specifically, MICE exhibits inferior performance for country-specific results. In terms of storage, single value imputation slightly outperforms the proposed model, if considered both steps. However, the one-step approaches require full retraining when new data becomes available, which can impact results for other countries.

When considering computational efficiency, encompassing aspects like runtime and memory consumption, the two-step approach stands out as the preferred choice. It's important to emphasise that the performance of Local models is closely linked to the availability and quality of data within a given country. While this study has the privilege of using high-quality data with rich claims and exposures, this may not be the case for every country or data source. In such cases, the two-step approach with its cross-country learning capabilities provides a distinct advantage, as we can use the insights gained from the Global model to retrain the second step of the process.

Overall, our proposed two-step hierarchical modeling approach achieves superior predictive performance for nearly all countries, outperforming Local models and the MICE method, with log-likelihood proving to be a more reliable measure than RMSE due to the distributional assumptions of the data generation process. The Two-step model significantly enhances generalization for smaller countries, such as countries 7 and 8, by leveraging a Global model in the first step, which protects local specifics and improves performance even stronger compared to larger countries like countries 4 and 5.

Summary and outlook

This study introduces a novel two-stage hierarchical mortality model that integrates global and local data to improve regional mortality risk estimation, particularly in data-scarce regions. The model leverages a LightGBM [31] in the first stage to capture global patterns, followed by country-specific refinements in the second stage. This approach demonstrated superior predictive accuracy compared to traditional methods and effectively addressed challenges related to missing data, scalability, and overgeneralization, offering a robust solution for mortality risk modeling across diverse regions.

The two-stage hierarchical modeling approach not only enhances predictive performance but also offers practical benefits in fields such as life insurance pricing, risk assessment, and public health planning. By generating more accurate mortality risk estimates, particularly in regions with limited local data, the model supports better-informed decision-making in industries that rely on precise risk evaluations. Its scalability and computational efficiency make it especially valuable in large-scale, multi-regional contexts.

Our model also stands out for its computational efficiency, excelling in runtime, memory usage, and storage requirements, particularly when the first-stage global model

is excluded. This efficiency is advantageous when scaling to new countries, as only the second step requires retraining, leaving existing predictions unaffected. The reduced model size speeds up training times while maintaining high performance, making it suitable for applications where rapid training is essential. Additionally, the model provides an efficient solution for handling missing data, outperforming other methods like single-value imputation or MICE, particularly when working with small datasets where local data alone is insufficient, and the pre-learned knowledge of a larger model becomes critical.

Despite its strong performance across multiple regions, the model's effectiveness depends on the availability and quality of data. In regions with low or inconsistent data quality, future research could explore more advanced imputation techniques or alternative methods for managing missing data. Further work could also investigate optimizing computational efficiency for even larger datasets or extending the model's applicability to domains such as epidemiological forecasting, financial risk modeling, or public health surveillance. Integrating techniques like deep learning could enhance performance for more complex datasets, though this may compromise its interpretability.

The flexibility and robustness of the proposed hierarchical model open up new possibilities for accurate risk estimation, particularly in data-scarce environments. As industries continue to rely on precise mortality estimates for strategic decision-making, this approach sets the foundation for more reliable, scalable, and adaptable models capable of addressing the complexities of regional variability without compromising performance.

Supporting information

S1 Appendix. Rest of country-specific results

Table 5. Cross-country evaluation of computational efficiency

Metric	Local model	Two-step model (Step 1 & 2)	One-step model (Single Value)	One-step model (MICE)
Runtime (Sec)	4.586×10^4	$6.122 \times 10^3 + 8.064 \times 10^2$	4.696×10^4	5.119×10^5
Memory (MB)	1.382×10^4	$5.562 \times 10^2 + 4.909 \times 10^3$	1.069×10^4	2.004×10^5
Storage (KB)	7.903×10^6	$8.816 \times 10^6 + 7.837 \times 10^6$	6.669×10^6	5.785×10^7

Table 6. Performance evaluation for country 1

Metric	Local model	Two-step model	One-step model (Single Value)	One-step model (MICE)
RMSE (Train)	2.509×10^{-2}	2.506×10^{-2}	2.539×10^{-2}	2.614×10^{-2}
RMSE (Test)	2.181×10^{-2}	2.180×10^{-2}	2.210×10^{-2}	2.259×10^{-2}
Log Likelihood (Train)	-6.575×10^3	-6.538×10^3	-7.927×10^3	-7.916×10^3
Log Likelihood (Test)	-2.071×10^3	-2.066×10^3	-2.409×10^3	-2.459×10^3
Runtime (Sec)	3.565×10^3	3.832×10^1	-	-
Memory (MB)	1.301×10^3	7.233×10^1	-	-
Storage (KB)	8.420×10^5	8.371×10^5	-	-

Table 7. Performance evaluation for country 2

Metric	Local model	Two-step model	One-step model (Single Value)	One-step model (MICE)
RMSE (Train)	2.631×10^{-2}	2.628×10^{-2}	2.853×10^{-2}	2.855×10^{-2}
RMSE (Test)	1.872×10^{-2}	1.872×10^{-2}	1.920×10^{-2}	1.941×10^{-2}
Log Likelihood (Train)	-7.877×10^3	-7.673×10^3	-8.017×10^3	-8.201×10^3
Log Likelihood (Test)	-2.152×10^3	-2.151×10^3	-2.542×10^3	-2.537×10^3
Runtime (Sec)	2.759×10^3	0.988×10^2	-	-
Memory (MB)	1.580×10^3	5.512×10^2	-	-
Storage (KB)	9.751×10^5	9.688×10^5	-	-

Table 8. Performance evaluation for country 3

Metric	Local model	Two-step model	One-step model (Single Value)	One-step model (MICE)
RMSE (Train)	1.674×10^{-2}	1.673×10^{-2}	1.764×10^{-2}	1.783×10^{-2}
RMSE (Test)	1.329×10^{-2}	1.328×10^{-2}	1.328×10^{-2}	1.331×10^{-2}
Log Likelihood (Train)	-3.314×10^{3}	-3.119×10^{3}	-3.471×10^{3}	-3.479×10^{3}
Log Likelihood (Test)	-9.515×10^{2}	-9.506×10^{2}	-1.109×10^{3}	-1.210×10^{3}
Runtime (Sec)	2.387×10^{3}	9.700×10^{1}	-	-
Memory (MB)	1.324×10^{3}	9.292×10^{2}	-	-
Storage (KB)	8.326×10^{5}	8.279×10^{5}	-	-

Table 9. Performance evaluation for country 4

Metric	Local model	Two-step model	One-step model (Single Value)	One-step model (MICE)
RMSE (Train)	2.999×10^{-2}	2.219×10^{-2}	3.183×10^{-2}	3.183×10^{-2}
RMSE (Test)	2.219×10^{-2}	2.998×10^{-2}	3.307×10^{-2}	3.307×10^{-2}
Log Likelihood (Train)	-6.842×10^{3}	-6.730×10^{3}	-1.166×10^{4}	-1.167×10^{4}
Log Likelihood (Test)	-1.907×10^{3}	-1.900×10^{3}	-2.275×10^{3}	-2.275×10^{3}
Runtime (Sec)	1.772×10^{3}	1.401×10^{1}	-	-
Memory (MB)	1.084×10^{3}	1.074×10^{3}	-	-
Storage (KB)	7.004×10^{5}	6.954×10^{5}	-	-

Table 10. Performance evaluation for country 6

Metric	Local model	Two-step model	One-step model (Single Value)	One-step model (MICE)
RMSE (Train)	2.404×10^{-2}	2.396×10^{-2}	2.731×10^{-2}	2.846×10^{-2}
RMSE (Test)	2.129×10^{-2}	2.128×10^{-2}	2.565×10^{-2}	2.648×10^{-2}
Log Likelihood (Train)	-1.081×10^{4}	-1.060×10^{4}	-1.191×10^{4}	-1.285×10^{4}
Log Likelihood (Test)	-3.194×10^{3}	-3.186×10^{3}	-3.492×10^{3}	-3.621×10^{3}
Runtime (Sec)	1.093×10^{4}	3.421×10^{3}	-	-
Memory (MB)	2.666×10^{3}	1.331×10^{3}	-	-
Storage (KB)	1.735×10^{6}	1.724×10^{6}	-	-

Table 11. Performance evaluation for country 8

Metric	Local model	Two-step model	One-step model (Single Value)	One-step model (MICE)
RMSE (Train)	3.283×10^{-2}	3.230×10^{-2}	3.304×10^{-2}	3.514×10^{-2}
RMSE (Test)	2.909×10^{-2}	2.906×10^{-2}	2.907×10^{-2}	2.955×10^{-2}
Log Likelihood (Train)	-1.247×10^{3}	-1.546×10^{3}	-1.730×10^{3}	-1.845×10^{3}
Log Likelihood (Test)	-5.275×10^{2}	-5.211×10^{2}	-6.318×10^{2}	-6.398×10^{2}
Runtime (Sec)	5.235×10^{2}	4.571×10^{1}	-	-
Memory (MB)	2.641×10^{2}	8.588×10^{1}	-	-
Storage (KB)	1.334×10^{5}	1.303×10^{5}	-	-

S2 Appendix. Hyperparameter optimization

For the hyperparameter optimization we used Hyperopt, a Python library that uses the Tree Parzen Estimator (TPE) algorithm. TPE is an efficient method leveraging a probabilistic model to guide the search for optimal hyperparameters. TPE workflow can be briefly characterized as follows:

Initialisation: TPE starts by randomly sampling a few hyperparameter combinations to create an initial set of observations. These initial combinations serve as a starting point for the optimisation process.

Probabilistic Modelling: TPE uses probabilistic models to capture the relationship between hyperparameter values and the performance metric (e.g., loss or accuracy). Specifically, it models the probability that a configuration will produce better results.

Exploitation and exploration: Based on the probabilistic models, TPE tries to balance exploration and exploitation. It aims to understand the correlation between hyperparameter values and performance, emphasising configurations that are likely to lead to better results. This process is similar to the gradient descent algorithm, but instead of searching for the gradient of the loss function, it focuses on the probability distribution of the hyperparameters.

Updating the model: As TPE collects more observations and evaluates additional configurations, it updates its probabilistic models. The algorithm iteratively learns and

refines its models to make more informed decisions.

By iteratively balancing exploration and exploitation, TPE efficiently navigates the hyperparameter space, eventually converging on an optimal set of hyperparameters for a given ML model. For a deeper understanding of the TPE algorithm and practical application we refer to [27] and [26]. The optimized hyperparameters of the discussed models are represented in Table 12 (For the sake of clarity and simplicity, MICE has been excluded).

Table 12. Optimal Hyperparameters for Local, Two-step and One-step models

Country	n_estimators	subsample	colsample_bytree	num_leaves	min_child_samples	learning_rate
Local models						
1	1806	0.8247	0.7842	51	76	0.01
2	1452	0.7526	0.5450	6	78	0.01
3	627	0.6610	0.5325	4	49	0.01
4	1118	0.5877	0.5378	4	89	0.01
5	1875	0.9231	0.6487	50	13	0.01
6	1223	0.8147	0.5253	15	41	0.01
7	1197	0.9384	0.8898	51	5	0.01
8	1990	0.7399	0.6098	5	80	0.01
Two-step model						
1^{st} step	441	0.7833	0.8400	27	734	0.1
2^{nd} step - 1	1488	0.6711	0.9983	9	644	0.01
2^{nd} step - 2	322	0.9983	0.8304	98	992	0.01
2^{nd} step - 3	1580	0.5496	0.6262	97	43	0.01
2^{nd} step - 4	895	0.8998	0.8485	44	666	0.01
2^{nd} step - 5	1786	0.9210	0.8615	94	294	0.01
2^{nd} step - 6	1993	0.7889	0.8863	18	735	0.01
2^{nd} step - 7	82	0.7024	0.6998	22	421	0.01
2^{nd} step - 8	1538	0.7188	0.6480	78	166	0.01
One-step model (Single Value)						
–	1355	0.5807	0.8553	17	296	0.05

S3 Appendix. Evaluation of prediction calibration

The balance property is a critical consideration in the context of statistical modeling. It signifies that a well-calibrated model should exhibit no deviations between the mean of observed and predicted values. This property is satisfied for any generalized linear model (GLM) within the exponential dispersion family (EDF) as long as the canonical link function is employed ([40], [42]). However, deviations can arise when the chosen link function departs from the canonical link. In such cases, it's essential to address this issue. [41] provides a relatively straightforward approach to correct for balance property failures, ensuring that the model remains well-calibrated.

To mitigate these issues and achieve calibration, we leverage Gradient Boosting advantages, which allows us to optimize the model with a Poisson loss function, implicitly using a log-link function. This strategic choice corrects any balance property problems and results in a well-calibrated model. Table 13 below illustrates that our models, in general, exhibit either no deviations or very minimal differences between observed and predicted death counts. Calibration is more meaningful when applied to training data, where models are meticulously adapted and fine-tuned. Any minor deviations observed in the test set are typically of negligible significance. For the Two-step model, the final mean predictions are computed by multiplying the means from the first and second steps, ensuring robustness in our approach.

Table 13. Predicted vs. observed avg. death counts (Train/Test)

Country		1	2	3	4	5	6	7	8
Local models									
Predicted Mean (Train)		0.0013	0.0008	0.0006	0.0009	0.001	0.0014	0.0009	0.003
Observed Mean (Train)		0.0013	0.0008	0.0006	0.0009	0.001	0.0014	0.0009	0.003
Predicted Mean (Test)		0.0013	0.0008	0.0006	0.0009	0.001	0.0014	0.0008	0.003
Observed Mean (Test)		0.0012	0.0008	0.0006	0.0009	0.0009	0.0013	0.0008	0.0029
Two-step model	1^{st} step					2^{nd} step			
Predicted Mean (Train)	0.001	1.0302	0.9826	0.9791	0.9813	0.9833	0.9885	1.0031	1.056
Observed Mean (Train)	0.001	1.0337	0.9908	0.98	0.9855	0.9958	0.9892	1.0049	1.0639
Predicted Mean (Test)	0.001	1.0295	0.984	0.9805	0.9826	0.955	0.9952	1.003	1.0566
Observed Mean (Test)	0.001	0.9687	1.0172	0.9351	0.9218	0.9101	0.9094	0.9027	1.0156
One-step model (Single-Value)									
Predicted Mean (Train)		0.0013	0.0008	0.0006	0.0009	0.001	0.0014	0.0009	0.003
Observed Mean (Train)		0.0013	0.0008	0.0006	0.0009	0.001	0.0014	0.0009	0.003
Predicted Mean (Test)		0.0013	0.0008	0.0006	0.0009	0.001	0.0014	0.0009	0.0029
Observed Mean (Test)		0.0012	0.0008	0.0006	0.0009	0.0009	0.0013	0.0008	0.0029

References

1. McKinsey & Company. Digital disruption in insurance: Cutting through the noise. 2017. `https://www.mckinsey.com/~/media/mckinsey/industries/financial%20services/our%20insights/time%20for%20insurance%20companies%20to%20face%20digital%20reality/digital-disruption-in-insurance.ashx`.
2. Cox DR. Regression models and life-tables. J R Stat Soc Series B (Methodological). 1972;34(2):187–202.
3. Anderson BA, Silver BD. Issues of data quality in assessing mortality trends and levels in the New Independent States. Premature Death in the New Independent States. 1997:120–155. Washington, DC: National Academy Press.
4. Mathers C, Boerma T. Mortality measurement matters: improving data collection and estimation methods for child and adult mortality. PLoS Med. 2010;7(4):e1000265.
5. Heuveline P. Global and National Declines in Life Expectancy: An End-of-2021 Assessment. Popul Dev Rev. 2022;48(1):31–50.
6. Reiter JP. Bayesian finite population imputation for data fusion. Stat Sin. 2012;22:795–811.
7. Van Der Putten P, Kok JN, Gupta A. Data fusion through statistical matching. Available at SSRN 297501. 2002.
8. Gilula Z, McCulloch RE, Rossi PE. A direct approach to data fusion. J Mark Res. 2006;43(1):73–83.
9. Moriarity C, Scheuren F. A note on Rubin's statistical matching using file concatenation with adjusted weights and multiple imputations. J Bus Econ Stat. 2003;21(1):65–73.
10. Coppola L, Di Zio M, Luzi O, Ponti A, Scanu M. On the use of Bayesian networks in official statistics. Atti della XLI Riunione Scientifica della Società Italiana di Statistica. 2002:237–240.
11. Rässler S. A non-iterative Bayesian approach to statistical matching. Stat Neerl. 2003;57(1):58–74.

12. Lahat D, Adali T, Jutten C. Multimodal data fusion: an overview of methods, challenges, and prospects. Proc IEEE. 2015;103(9):1449–1477.

13. Khaleghi B, Khamis A, Karray FO, Razavi SN. Multisensor data fusion: A review of the state-of-the-art. Inf Fusion. 2013;14(1):28–44.

14. Rubin DB, Schenker N. Multiple imputation for interval estimation from simple random samples with ignorable nonresponse. J Am Stat Assoc. 1986;81(394):366–374.

15. Martínez HP, Yannakakis GN. Deep multimodal fusion: Combining discrete events and continuous signals. In: Proceedings of the 16th International Conference on Multimodal Interaction. 2014. p. 34–41.

16. Kuncheva LI. Combining pattern classifiers: methods and algorithms. 2014. John Wiley & Sons.

17. Ramachandram D, Taylor GW. Deep multimodal learning: A survey on recent advances and trends. IEEE Signal Process Mag. 2017;34(6):96–108.

18. Neverova N, Wolf C, Taylor G, Nebout F. Moddrop: adaptive multi-modal gesture recognition. IEEE Trans Pattern Anal Mach Intell. 2015;38(8):1692–1706.

19. Joze HRV, Shaban A, Iuzzolino ML, Koishida K. MMTM: Multimodal transfer module for CNN fusion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020. p. 13289–13299.

20. Karpathy A, Toderici G, Shetty S, Leung T, Sukthankar R, Fei-Fei L. Large-scale video classification with convolutional neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014. p. 1725–1732.

21. Pawłowski M, Wróblewska A, Sysko-Romańczuk S. Effective techniques for multimodal data fusion: A comparative analysis. Sensors. 2023;23(5):2381.

22. Trajanoska M, Trajanov R, Eftimov T. Dietary, comorbidity, and geo-economic data fusion for explainable COVID-19 mortality prediction. Expert Syst Appl. 2022;209:118377.

23. Albahri AS, Duhaim AM, Fadhel MA, et al. A systematic review of trustworthy and explainable artificial intelligence in healthcare: Assessment of quality, bias risk, and data fusion. Inf Fusion. 2023.

24. Boerma T. Addressing the gaps in mortality data: A case for national mortality surveillance. Am J Trop Med Hyg. 2023;108(5 Suppl):1.

25. Hastie T, Tibshirani R, Friedman JH. The elements of statistical learning: data mining, inference, and prediction. 2nd ed. 2009. Springer.

26. Bergstra J, Yamins D, Cox D. Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. In: Proceedings of the International Conference on Machine Learning. 2013. p. 115–123.

27. Bergstra J, Bardenet R, Bengio Y, Kégl B. Algorithms for hyper-parameter optimization. Adv Neural Inf Process Syst. 2011;24.

28. James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning. 2013. Springer.

29. Laird N, Olivier D. Covariance analysis of censored survival data using log-linear analysis techniques. J Am Stat Assoc. 1981;76(374):231–240.

30. Ernesti J, Kaiser P. Python 3. Rheinwerk: Bonn, Germany. 2017.

31. Oram E, Dash PB, Naik B, Nayak J, Vimal S, Nataraj SK. Light gradient boosting machine-based phishing webpage detection model using phisher website features of mimic URLs. Pattern Recognit Lett. 2021;152:100–106.

32. Doove LL, Van Buuren S, Dusseldorp E. Recursive partitioning for missing data imputation in the presence of interaction effects. Comput Stat Data Anal. 2014;72:92–104.

33. Yan J, Guszcza J, Flynn M, Wu C-SP. Applications of the offset in property-casualty predictive modeling. In: Proceedings of the Casualty Actuarial Society e-Forum. 2009;1(1):366–385.

34. Renner IW, Louvrier J, Gimenez O. Combining multiple data sources in species distribution models while accounting for spatial dependence and overfitting with combined penalized likelihood maximization. Methods Ecol Evol. 2019;10(12):2118–2128.

35. Lee RD, Carter LR. Modeling and forecasting US mortality. J Am Stat Assoc. 1992;87(419):659–671.

36. Luciano E, Outreville JF, Rossi M. Life insurance ownership by Italian households: A gender-based differences analysis. Geneva Pap Risk Insur Issues Pract. 2016;41:468–490.

37. Feldman R, Dowd B. Risk segmentation: goal or problem?. J Health Econ. 2000;19(4):499–512.

38. Nassar MO, Kanaan G. The factors affecting the performance of data fusion algorithms. In: Proceedings of the 2009 International Conference on Information Management and Engineering. 2009. p. 465–470.

39. Mitchell HB. Data fusion: concepts and ideas. 2012. Springer Science & Business Media.

40. Noll A, Salzmann R, Wüthrich MV. Case study: French motor third-party liability claims. SSRN. 2020.

41. Wüthrich MV. Bias regularization in neural network models for general insurance pricing. Eur Actuar J. 2020;10(1):179–202.

42. Wüthrich MV. The balance property in neural network modelling. Stat Theory Relat Fields. 2022;6(1):1–9.

43. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine learning in Python. J Mach Learn Res. 2011;12:2825–2830.

44. Klomp T. Iterative imputation in Python: A study on the performance of the package IterativeImputer. 2022.

45. Holford TR. The analysis of rates and of survivorship using log-linear models. Biometrics. 1980;36(2):299–305.

46. Van Buuren S. Flexible imputation of missing data. 2018. CRC Press.

47. Cameron AC, Trivedi PK. Regression analysis of count data. 2013. Cambridge University Press.

48. Congdon P. Bayesian statistical modelling. 2001. Wiley.

49. Diez Roux AV. A glossary for multilevel analysis. J Epidemiol Community Health. 2002;56(8):588–594.

50. Gelman A, Hill J. Data analysis using regression and multilevel/hierarchical models. 2006. Cambridge University Press.

51. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian data analysis. 3rd ed. 2013. CRC Press.

52. Ke G, Meng Q, Finley T, et al. LightGBM: A highly efficient gradient boosting decision tree. In: Advances in Neural Information Processing Systems. 2017. p. 3146–3154.

53. Banerjee S, Carlin BP, Gelfand AE. Hierarchical modeling and analysis for spatial data. 2014. CRC Press.

54. Murray CJL, Laakso T, Shibuya K, Hill K, Lopez AD. Can we achieve Millennium Development Goal 4? New analysis of country trends and forecasts of under-5 mortality to 2015. Lancet. 2007;370(9592):1040–1054.

55. Raftery AE, Li N, Ševčíková H, Gerland P, Heilig GK. Bayesian probabilistic population projections for all countries. Proc Natl Acad Sci USA. 2014;111(11):4676–4684.

56. Diez Roux AV. Investigating neighborhood and area effects on health. Am J Public Health. 2001;91(11):1783–1789.

57. Subramanian SV, Jones K, Duncan C. Multilevel methods for public health research. In: Kawachi I, Berkman LF, editors. Neighborhoods and Health. 2003. p. 65–111. Oxford University Press.

58. Yakovyna V, Shakhovska N, Szpakowska A. A novel hybrid supervised and unsupervised hierarchical ensemble for COVID-19 cases and mortality prediction. Sci Rep. 2024;14:Article number: 9782. 10.1038/s41598-024-60637-y.

Part IV.

Transfer Learning for No Data Scenario

9. Transfer learning for mortality risk: A case study on the United Kingdom

Contributing article

Nalmpatian, A., Heumann, C., Alkaya, L., and Jackson, W. (2025). Transfer learning for mortality risk: A case study on the United Kingdom. *PLoS One*, 20(4): e0313378.

Data repository

https://zenodo.org/records/14547227

Copyright information

Author contributions

The idea to extend the hierarchical modeling framework to countries with no direct insured portfolio data using transfer learning was developed by Asmik Nalmpatian and Christian Heumann. Asmik Nalmpatian led the formulation of the transfer learning pipeline, including the synthetic data generation and the drift correction model. She curated and preprocessed population and insurance datasets in collaboration with Levent Alkaya and William Jackson. The pretrained global model was refined and adapted to the UK setting through extensive validation against official data. Asmik Nalmpatian authored the first draft of the paper. Christian Heumann supported the methodological design and validation, while Levent Alkaya and William Jackson contributed to the software implementation and interpretation of the model outputs. All co-authors were involved in the critical revision of the final manuscript.

Supplementary material

Supplementary materials for this article are available online at PLOS ONE (DOI: 10.1371/journal.pone.0313378), including files S1–S5.

PLOS ONE

RESEARCH ARTICLE

Transfer learning for mortality risk: A case study on the United Kingdom

Asmik Nalmpatian[1]*, **Christian Heumann**[1], **Levent Alkaya**[2], **William Jackson**[2]

1 Department of Statistics, LMU Munich, Munich, Bavaria, Germany, **2** Independent Researcher, Munich, Bavaria, Germany

* asmik.nalmpatian@campus.lmu.de

OPEN ACCESS

Citation: Nalmpatian A, Heumann C, Alkaya L, Jackson W (2025) Transfer learning for mortality risk: A case study on the United Kingdom. PLoS One 20(5): e0313378. https://doi.org/10.1371/journal.pone.0313378

Editor: Arne Johannssen, University of Hamburg, GERMANY

Received: October 23, 2024

Accepted: April 08, 2025

Published: May 23, 2025

Data availability statement: The data utilized in this study are owned by a third-party organization (an insurance company), and we do not have the authority to share it. Additionally, the synthetic dataset is derived from real data from other countries, and revealing feature names or data-related details beyond the visual results already shared would breach confidentiality agreements. Consequently, the synthetic dataset and detailed generation scripts cannot be made publicly available. To ensure reproducibility, we have

Abstract

This study introduces a transfer learning framework to address data scarcity in mortality risk prediction for the UK, where local mortality data is unavailable. By leveraging a pre-trained model built from data across eight countries (excluding the UK) and incorporating synthetic data from the countries most similar to the UK, our approach extends beyond national boundaries. This framework reduces reliance on local datasets while maintaining strong predictive performance. We evaluate the model using the Continuous Mortality Investigation (CMI) dataset and a Drift model to address discrepancies arising from local demographic differences. Our research bridges machine learning and actuarial science, enhancing mortality risk prediction and pricing strategies, particularly in data-poor settings.

Introduction

In life insurance, accurate mortality risk prediction is essential for pricing and managing risks. However, this process is often hindered by data scarcity, particularly in underrepresented demographic segments or smaller niches of the market. Mortality events are infrequent, meaning data accumulates slowly, making it difficult for insurers to build robust predictive models. This lack of data can lead to unreliable risk assessments and pricing strategies, ultimately affecting profitability and customer affordability.

Transfer learning offers a promising solution to these challenges by leveraging models trained on data-rich countries and adapting them to data-poor environments. This allows insurers to generate reliable mortality predictions even when local data is unavailable. Previous studies, such as those by [1] and [2], have laid the groundwork for transfer learning in mortality risk prediction, but have primarily focused on scenarios with small volumes of target data. Additionally, much of the research has relied on deep neural networks (DNNs), which, while powerful, can be computationally intensive and require extensive fine-tuning, especially for small datasets [3,4].

In contrast, gradient boosting machines (GBMs) offer a more efficient and interpretable alternative for transfer learning, particularly in cases where no target data is available. Despite their potential, GBMs have received less attention in the context of mortality risk prediction. Inspired by the success of machine learning (ML) models in clinical research [5–7], this study introduces a GBM-based transfer learning framework for predicting mortality rates in

included a comprehensive workflow outlining the synthetic data generation process in Appendix S5 and provided visual and textual descriptions of the dataset in the manuscript's Data Section. We have also properly cited and acknowledged external data sources, such as the Human Mortality Database (HMD) and the OECD datasets, which are thoroughly detailed and referenced in Table 2 of the manuscript. If needed, we are willing to offer additional guidance for researchers interested in the original data from the third-party sources. Here is the minimal anonymized dataset we are able to provide: [DOI 10.5281/zenodo.14546939 or URL https://zenodo.org/records/14547227]. This dataset corresponds to the figure titled "Comparison of UK mortality rates between Transfer Learning and CMI by age and gender." The information included is made public without ethical concerns, as the grouping by age and gender ensures that there is a sufficient amount of data to prevent tracing back to individuals. Extending the dataset beyond this point could raise ethical issues by introducing more granular data, which could potentially be used to identify individuals based on additional characteristics and country information. Such an extension would compromise the ethical integrity of the study. The authors have exclusively used anonymized and publicly available data, ensuring no identifiable personal information is utilized. This approach complies with ethical research standards. Additionally, the absence of reliance on proprietary UK-specific portfolio data further reinforces the ethical soundness of the work. To comply with institutional data availability requirements, we designate the Machine Learning Consulting Unit (MLCU) at Ludwig-Maximilians-Universität München as the official institutional contact for data-related inquiries. The MLCU can be reached at mlcu@stat.uni-muenchen.de, ensuring long-term accessibility for any data requests. The primary contact person is Dr. Andreas Bender, whose details are as follows: Institut für Statistik, Ludwig-Maximilians-Universität München, Ludwigstraße 33, D-80539 München.

Funding: The author(s) received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

the UK, where no local life insurance data is available. By incorporating synthetic data from countries most similar to the UK, this approach demonstrates high predictive accuracy while reducing dependence on local datasets. To further enhance the model, we introduce a Drift model to evaluate and correct discrepancies arising from demographic differences between countries.

This research not only extends the boundaries of transfer learning in actuarial science but also has broader implications for improving mortality risk prediction and pricing strategies in data-poor markets. Our study—the workflow of which is illustrated in Fig 1—is guided by three primary research questions:

(i) *How can we estimate mortality rates in a country with no internal life portfolio data?* This involves implementing a ML-based transfer method, focusing on the UK, and constructing a Country similarity index using external data to identify relevant source countries.
(ii) *How accurate is the model, and how can a Drift model address discrepancies between predicted and expected mortality?* The accuracy of the transfer learning method is assessed using various metrics, with a Drift model employed to explore factors contributing to discrepancies between transferred mortality tables and expected outcomes from the CMI dataset.
(iii) *How can additional variables beyond age and gender improve mortality risk predictions.* We investigate how the inclusion of additional variables can enhance the baseline mortality predictions, providing an application case to demonstrate improvements.

Related work

Previous research has predominantly focused on leveraging DNNs to model mortality data. A notable example is the work by [8], which discussed the integration of Generalized Linear Models (GLMs) within residual networks to capture both linear and nonlinear effects. Despite their potential, these Combined Actuarial Neural Networks (CANNs) face challenges in enforcing monotonicity, which is crucial for mortality data [9]. In contrast, our study explores the use of GBMs, which offer a more flexible, interpretable and computationally efficient alternative, particularly in data-scarce environments. GBMs have shown promise in various actuarial applications, providing a transparent framework for mortality prediction. Our approach extends previous methodologies by incorporating a Drift model to explicitly address demographic discrepancies, enhancing the model's adaptability to different population characteristics.

Numerous studies have aimed to compare health care systems, financing mechanisms and health outcomes across countries. Bauer and Ameringer [10] emphasizes the difficulty of collecting comprehensive data from different countries due to logistical and financial constraints. However, incorporating statistical data from credible sources like the World Health Organization (WHO) and the Organization for Economic Co-operation and Development (OECD), along with a proposed multivariate statistics framework, serves as a valuable supplement. The significance of conducting cross-national research on healthcare system performance is underscored since it is considered crucial for guiding public policy [11]. For example, [12] revealed that the difference in spending between the United States and European countries can be traced back to disparities in diagnosis and treatment rates for certain chronic conditions. Another study explored the impact of culture in forecasting a country's population health, gauged through life expectancy and healthcare spending [13]. Hofstede's influential study on cross-cultural research argues that comprehending a nation's

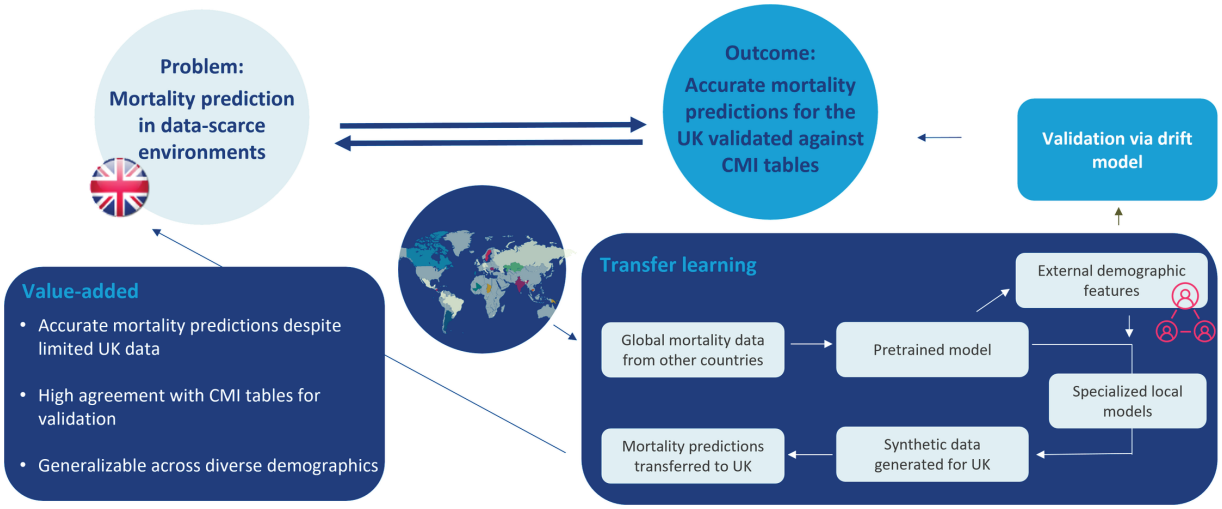


Fig 1. Graphical abstract to summarize our workflow and value-added.

https://doi.org/10.1371/journal.pone.0313378.g001

culture demands exploring dimensions like Power Distance, Individualism-Collectivism, Masculinity-Femininity, and Uncertainty Avoidance [14]. To tackle the issue of determining and gauging population health, [15] suggested two models. The descriptive model assesses population health through indicators like life expectancy, categorized by markers such as socio-economic status or race. Various indices measure similarities between countries across a range of dimensions, yet there is currently a gap in addressing both mortality and life insurance specifics. Our approach involves constructing and optimizing a distance-based index for country similarity. We base this approach on solely external sources. The forthcoming sections outline our proposed method in a reproducible manner.

Database and methodology

Data

In our study, we rely on the open source Human Mortality Database (HMD) as our primary external data source. HMD offers age and gender-specific mortality rates for the overall population across various countries. However, our primary focus is not on estimating the mortality of the overall population in the UK. Instead, our goal is to estimate the mortality rates within the company's own life insurance portfolio in the UK. It's important to note that there are often differences between overall mortality rates and those within a specific portfolio, particularly due to the underwriting process in life insurance. To address this limitation, we leverage data from eight countries and establish connections to capture this discrepancy between overall and portfolio mortality rates. To ensure that the analysis accurately reflects the mortality patterns across different countries and within the company's life insurance portfolio, our approach involves three different populations, as illustrated in Fig 2: the global overall population, the global insured population of the company, and the insured population of the company within a particular country.

Overall population: Age- and gender-specific overall population mortality rates from the HMD are retrieved for all countries. While these represent total population mortality, not

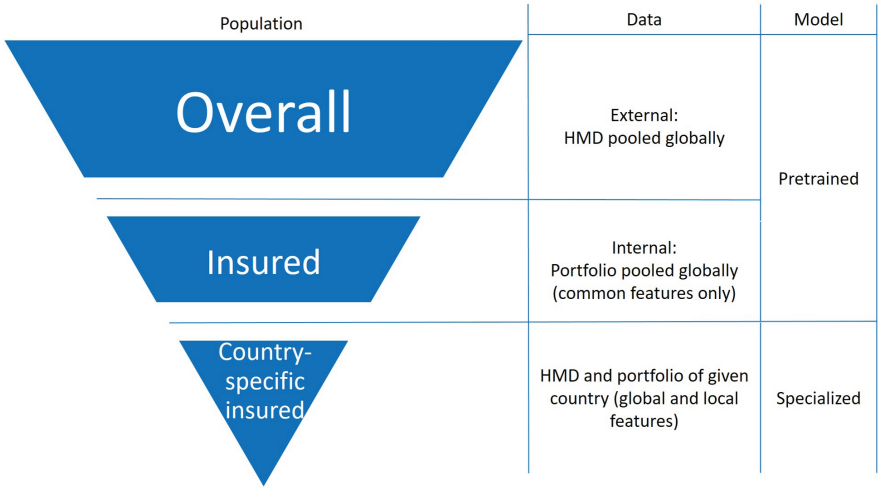


Fig 2. Illustration of targeted population segments across different datasets and models.

https://doi.org/10.1371/journal.pone.0313378.g002

insured population mortality, they bridge the gap between total and insured mortality, as it is the only feature we have available for the target country. To minimize yearly artifacts mortality rates from 2008 to 2018 were projected one year ahead using the Lee Carter model [16] (see Methodology section and S2 Appendix).

Insured population: We utilize a pooled internal portfolio dataset from different countries to pretrain a GBM model [17] for predicting mortality rates for the insured population globally. This dataset incorporates common global characteristics shared across different countries, such as age, gender, sum assured, allowing for cross-country data comparison, and integrates the overall population mortality, yielding in a total of 9 global features, without any country indicator. (see Methodology section and S1 Appendix).

The dataset includes policy data from a global primary insurer that was active during the specified period, totaling almost 10 million life-years of exposure and recording nearly 10,000 insurance claims (deaths). The data analysis was conducted in an aggregated form, grouped into distinct combinations of feature values, summarizing the deaths D_j and exposure E_j data for each unique combination features across all $j = 1, \ldots, K$ countries, in this case $K = 8$, the names of which have been withheld to maintain confidentiality. Four of the countries are located in Western Europe, three in Latin America, and one in Central and Eastern Europe.

Table 1 provides a detailed overview of D_j, E_j and the total number of years T_j in country j, to give the main characteristics and distribution of the pooled dataset. This paper will analyse age and gender as internal features, while keeping other features used in the modeling anonymous for privacy reasons.

Insured population in specific countries: In addition to the global features, including the overall population mortality of these countries, we include 12-16 local features from each country j, depending on local data availability, such as occupational class, which are not comparable across regions. After retraining the specialized GBM models on a total of 21-25 features, initialized by the pretrained model, we predict mortality rates for the portfolio of country M using a synthetic dataset tailored specifically for M. Our method for creating the synthetic dataset combines stochastic and rule-based techniques to bootstrap by

Table 1. Overview of death counts D_j, exposure in life years E_j, and total number of years T_j in country j.

Country j	D_j	E_j	T_j
1	1 699	1 295 299	2013–2020
2	1 291	1 686 299	2010–2020
3	494	815 795	2010–2020
4	1 225	1 347 150	2017–2020
5	1 816	1 825 901	2016–2020
6	2 132	1 548 157	2016–2020
7	458	498 560	2017–2020
8	297	99 473	2015–2020
Total	9 412	9 116 634	2010–2020

https://doi.org/10.1371/journal.pone.0313378.t001

resampling from the internal portfolio of K countries, while introducing variations to account for uncertainty [18] (see Methodology section and S1 Appendix).

Mortality of UK's insurance population for evaluation:

We utilize the '16' series mortality tables from Working Paper 154 [19] for the evaluation purposes and the Drift model, given the absence of an actual UK portfolio for comparison. These tables, derived from data from different UK life insurance companies, offer detailed insights into age, gender, smoking status, and curtate duration. To guarantee an impartial assessment and prevent undue complication, we consolidate the tables according to age and gender categories that correspond to population proportions.

External data for the Country similarity index:

The Country similarity index seeks to measure the similarity between the target country M and the K (= 8) source countries in the internal dataset in terms of mortality and life insurance characteristics. We develop this by considering various indicators, selected based on prior research and expert input, adaptable to specific contexts. These indicators are categorized into three dimensions: Life Insurance Performance Indicators, Healthcare Statistics, and Overall Population Mortality. The details of these indicators are outlined in Table 2, with the methodology for their construction discussed in the subsequent subsection.

Methodology

General setup

Consider a scenario where K source datasets with aggregated sample size n_j are collected from countries $j = 1, \ldots, K$ representing life insurance portfolios. The pooled dataset has total aggregated sample size $N = \sum_{j=1}^{K} n_j$. The objective is to estimate death counts $D \in \mathbb{R}^N$ relative to exposure. The feature set $X \in \mathbb{R}^{N \times p}$ comprises global features X^{global} that are comparable and available across countries including the overall population from HMD and local features X_j^{local} that are specific to each country. Our challenge arises in estimating mortality rates D_M due to the lack of internal data. However, we do have access to external data that provides information about mortality rates in different countries, including M. So, the scenario we are dealing with is comparing what we know from this external data along with some internal data we have (which is not specific to M) to try to estimate mortality rates specifically for country M. Fig 3 is a visual representation of the transfer learning framework: From the pretrained global model to the refined mortality rate predictions for the target county M based on a synthetic dataset.

Pretrained model

Consider a broad category of risk prediction models, where the process of fitting the model involves using a loss function $L(\gamma; D, X)$. With an estimated parameter vector $\hat{\gamma}$

Table 2. Dimensions and items obtained from external sources for the construction of a Country similarity index related to mortality in life insurance.

Item	Description and source
1. Life Insurance Performance Indicators from OECD	
1.1 Life Insurance Share	The ratio of gross life insurance premium to total gross premium, indicating the relative importance of life insurance compared to non-life insurance [20].
1.2 Density	The ratio of life insurance premiums to the whole population, measured in US Dollars [21].
1.3 Penetration	The level of development of the life insurance sector in a country, represented as a percentage [22].
1.4 Total Gross Premiums	Aggregate amount of premiums collected by life insurance companies in US Million Dollars [23].
1.5 Retention Ratio	Percentage of premiums retained by an insurance company rather than being transferred to reinsurers [24].
2. Healthcare Statistics	
2.1 Health Care	Measured by The Health Index by Global Residence Index, providing an overall assessment of the healthcare system and general health of the local population. Ranges from 0 to 1, indicating low to high healthcare levels [25].
2.2 Retirement Pension	Country-specific Minimum Pensionable Age for Men obtained from Indicator Data [26].
2.3 Medical Staff per Capita	Number of physicians, nurses, and midwives per 1,000 people [27].
2.4 Hospital Beds per Capita	Number of hospital beds per 10,000 population [28].
2.5 Access to Basic Healthcare	Percentage of people with access to basic healthcare [29].
2.6 Healthcare Expenditure per Capita	Expenditure on healthcare per capita in US Dollars [30].
2.7 Risk of Impoverishing Expenditure for Surgical Care	Percentage of people at risk [31].
3. Overall Population Mortality	
3.1 HMD Rates	Utilizing also here HMD's age- and gender-specific mortality rates by country [32].

https://doi.org/10.1371/journal.pone.0313378.t002

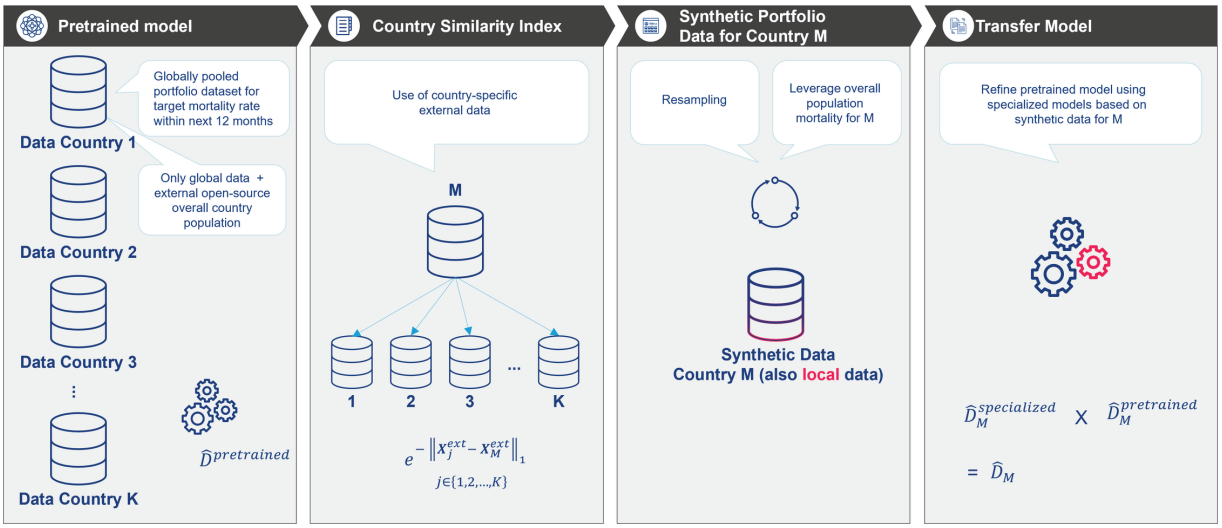


Fig 3. Framework sketch: Synthetic-data-based mortality predictions for target country M using a pretrained global mortality risk model.

https://doi.org/10.1371/journal.pone.0313378.g003

corresponding to the coefficients in a GBM, the predicted outcome is given by $\hat{D} = f(X, \hat{\gamma})$. Specifically, we employ the negative Poisson log-likelihood function with Poisson distributional assumption. By minimizing the expected loss function based on X^{global} we result in the parameter set estimate $\hat{\gamma}^{\text{pretrained}}$ an thus predicted number of deaths $\hat{D}^{\text{pretrained}}$. A detailed methodology for the GBM model is provided in S1 Appendix. Up to this point, a benchmark model has been developed without considering the country *M*. Previous work such as [33], [34], [4] and [35] characterize the similarity between the target model and the source models by a certain distance measure. Based upon this idea, we will generate a synthetic portfolio dataset X_M for country M, leveraging the similarity of the external data between the target population M and the source populations 1 to K (excluding M).

Country similarity index

To measure how similar the target country *M* is to the *K* source countries, we create a Country similarity index based on external insurance and mortality data $X^{\text{ext}} \in \mathbb{R}^{(K+1)\times Q}$, with K number of source countries and 1 target country. In our application case, Q is equal to 13, larger than $K + 1 = 9$. These Q items, which are given in Table 2 apply to the entire population of a country, rather than internal data *X*, which specifically characterizes the country's insured population. After centering and scaling, the Manhattan distance between vectors X_j^{ext} of each source $j = 1, \ldots, K$ and X_M^{ext} of target country *M* is calculated, as the sum of the absolute differences between corresponding components of vectors: $d(X_j^{\text{ext}}, X_M^{\text{ext}}) = \|X_j^{\text{ext}} - X_M^{\text{ext}}\|_1$ [36]. Finally, this results in a k-dimensional vector, representing the sum of item-wise distances between the $j = 1, \ldots, K$ and *M* across all Q items. The summation of distances over the countries is then transformed into the normed similarity score $s(X_j^{\text{ext}}, X_M^{\text{ext}}) = e^{-d(X_j^{\text{ext}}, X_M^{\text{ext}})}$ using the exponential function, so that the value range changes from $[0, \infty)$ to $(0, 1]$. This transformation allows a similarity comparison rather than an absolute measure of distance, and becomes important later in the resampling stage to define the variance of the Gaussian distribution.

Synthetic portfolio data for country M

In countries with no mortality data at all due to portfolio characteristics and size, synthetic data generation offers an efficient solution to address data limitations [37]. The process of producing mortality datasets that closely mimic actual data may comprise stochastic techniques [38], rule-based approaches set by human experts [39] or deep generative models (e.g., [40], [41]).

Assuming the known age and gender distribution for M, we resample feature combinations (rows) from the K datasets, encompassing both global and local features, along with the number of deaths, proportional to each similarity score $s(X_j^{\text{ext}}, X_M^{\text{ext}})$ for $j = 1, \ldots, K$. The overall population mortality of those countries has been substituted with the one of country M obtained from the HMD.

To address potential unknown heterogeneity between j and M, we use a data augmentation technique with noise drawing inspiration from established practices (e.g., [42], [43]):

1. Metric Data: We introduce Gaussian noise with a mean μ of 0 and a standard deviation σ that is inversely proportional to the similarity score: $\sigma = 1 - s(X_j^{\text{ext}}, X_M^{\text{ext}}) + 0.000001$. Higher similarity measure corresponds to a lower standard deviation, implying less noise is added to metric data.
2. Categorical Data: For categorical data, a noise level is drawn from $\mathcal{N}(0, \sigma^2)$, where again $\sigma = 1 - s(X_j^{\text{ext}}, X_M^{\text{ext}}) + 0.000001$. If the drawn value falls within a predefined interval around 0, the original value is retained, otherwise, a new value is drawn from the uniform distribution.

Finally, the synthetic dataset for the target country M is generated and contains the feature sets X_M^{global} (including HMD) and X_M^{local} as well as the exposure E_M for country M. The estimation of death counts, denoted as $\hat{D}_M$, is required. More details on the workflow of synthetic data generation is available in S5 Appendix.

Transfer model

Since the pretrained model excludes local factors like occupation class, which cannot be compared across countries, but may have significant impact on mortality, we calculate the specialized models with the local data on top. Each specialized model takes the output of the pretrained model from the first step and makes it more precise for that country. We find that incorporating local attributes during the latter phase of training offers optimal adaptability; this approach allows local nuances to be effectively integrated and, in cases where they are not applicable or transferable to the target country, they can be subsequently adjusted or mitigated. Initially, we utilize the global features of the synthetic dataset X_M^{global} to generate preliminary predictions $\hat{D}_M^{\text{pretrained}}$ using a pretrained model. Subsequently, we enhance these predictions by employing the specialized GBM models tailored for countries $j = 1, \ldots, K$. Through iterative boosting, the specialized model adjusts to the characteristics of the countries according to their similarity, thereby refining the mortality rate predictions. The final mortality rate predictions $\hat{D}_M$ are determined by combining the specialized predictions $\hat{D}_M^{\text{specialized}}$ and the pretrained predictions $\hat{D}_M^{\text{pretrained}}$ for all countries, as elaborated in the following Algorithm 1 and detailed out in S1 Appendix.

Agreement metrics

Using several metrics we evaluate the agreement of transferred mortality rates $\hat{\mu}_{\text{M}} = \hat{D}_M/E_M$ with the CMI mortality rates μ_{cmi}, as proxy for expected UK mortality. Specifically, we employed Spearman correlation, cosine similarity and R-squared with centered expected versus predicted mortality rates. These metrics are defined as follows:

1. Spearman correlation:

$$\rho = \frac{\text{cov}(\text{rank}(\hat{\mu}_{\text{M}}), \text{rank}(\mu_{\text{cmi}}))}{\sigma_{\text{rank}(\hat{\mu}_{\text{M}})}\sigma_{\text{rank}(\mu_{\text{cmi}})}}$$

2. Cosine similarity:

$$\text{c} = \frac{\hat{\mu}_{\text{M}} \cdot \mu_{\text{cmi}}}{\|\hat{\mu}_{\text{M}}\| \|\mu_{\text{cmi}}\|}$$

3. R-squared with centered actuals $\mu_{\text{cmi}}^{(i)}$ versus centered predicted vectors $\hat{\mu}_{\text{M}}^{(i)}$:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(\mu_{\text{cmi}}^{(i)} - \hat{\mu}_{\text{M}}^{(i)})^2}{\sum_{i=1}^{N}(\mu_{\text{cmi}}^{(i)} - \bar{\mu}_{\text{cmi}})^2}$$

The selection of Spearman correlation, cosine similarity, and centered R-squared is justified by their complementary insights into the evaluation of age- and gender-specific mortality predictions. Spearman correlation is robust for assessing rank-order relationships, making it ideal for capturing the alignment of predicted and actual mortality rankings, especially in the presence of non-linear trends and outliers. Cosine similarity focuses on the directional consistency of the predicted mortality distribution, ensuring that the shape and pattern of predictions align with CMI benchmarks, independent of magnitude differences. Centered R-squared evaluates the variance alignment between predictions and observed rates, emphasizing the model's goodness of fit to capture demographic-specific fluctuations. For instance,

Algorithm 1. Algorithmic representation of the transfer framework

1: Train the global GBM model $q(X^{\text{global}})$ on the pooled dataset, containing the datasets from countries $1,\ldots,K$. Here, X^{global} represents the features common across all countries.

2: For each country $j = 1,\ldots,K$ train a local GBM model $h_j(X_j)$ using country j's dataset, which includes both global features X^{global} and local features specific to country j. These models are initialised using the output predictions of the pre-trained global GBM model (as opposed to more conventional, i.e. random, initialisation).

3: For country M (= UK), calculate the similarity scores with each country $j = 1,\ldots,K$, based on external data with predefined similarity metric, which can include factors specific to life insurance, economic, and mortality.

4: For each country $j = 1,\ldots,K$, perform the following steps to create the synthetic dataset for country M (UK):

- Use the calculated similarity scores to proportionally resample exposures from each country j's dataset to contribute to country M's synthetic dataset. Ensure the total exposure for country M, E_M, is equal to the sum of resampled exposures E_{M_j} from each country j, which in total should amount to 100,000,000.
- Apply data augmentation by adding noise to the features to generate variability and improve the robustness of the model.
- Replace the population mortality variable in the dataset with that from country M, aligning the dataset with the mortality conditions of country M.
- Compile the resampled and augmented data to form the synthetic dataset X_M for country M. This dataset will be a row-block matrix where each block corresponds to data from a specific country j with different dimension, containing both global and local features. The first columns consist of global features to which the global model will be applied. Record the origin of each row to ensure that the specialized GBM model trained for that country can be subsequently applied.

5: Use the global model $q(X_M^{\text{global}})$ and the respective local models $h_j(X_M)$ to predict the expected value $E[D_M|X_M]$ for the synthetic dataset of country M:

$$E[D_M|X_M] = \sum_{j=1}^{K} \mu_{M_j} \cdot E_{M_j} = \sum_{j=1}^{K} q(X_{M_j}^{global}) \cdot h_j(X_{M_j}) \cdot E_{M_j} \tag{1}$$

where μ_{M_j} is the expected mortality rate for the synthetic data of country M derived from country j. The term $q(X_{M_j}^{\text{global}})$ is the global model's prediction using the global features of the synthetic dataset for country M derived from country j. The term $h_j(X_{M_j})$ represents the adjustment made by the local model of country j, applied to the portion of the synthetic dataset X_M that originated from country j. This ensures that the global model's predictions are fine-tuned to reflect the specific characteristics of country j that are as similar to country M, as determined by the similarity scores.

consider a scenario where the model accurately predicts mortality rates for males aged 30-50 but underestimates rates for older females (e.g., 70+). In this case, Spearman correlation would remain high if the ranking within age groups is preserved, even if predictions for older ages deviate in magnitude. Meanwhile, cosine similarity would decline due to a directional mismatch in the mortality profile for older females, reflecting a flatter or inconsistent trend compared to CMI rates. Lastly, the underprediction for older females would reduce the overall explained variance in the R-squared metric, highlighting that the model struggles with these demographic subgroups. Together, these metrics provide a fair assessment of the predictions, ensuring that both ranking, shape, and variance are considered, which is essential for accurately comparing predictions with CMI tables and understanding demographic trends.

Drift model evaluation

We propose a Drift model to evaluate the remaining disagreement by identifying and quantifying the drift drivers between target country's expected mortality and the mortality rates transferred from other countries to M.

We assume a Poisson distribution for mortality counts in country M, denoted as $D_M \sim \text{Poisson}(\mu_M \cdot E_M)$. Our analysis focuses on examining the discrepancy between the predicted mortality rate $\hat{\mu}_M$ and the actual rate μ_{cmi} across various features or feature categories. This

discrepancy, denoted as δ, serves as an indicator of the quality of transfer learning. We adopt the two-stage or residual model proposed by [44] to estimate δ:

$$D_{\mathrm{M}} \sim \mathrm{Poisson}(\underbrace{\delta \cdot \mu_{\mathrm{cmi}}}_{=\mu_{\mathrm{M}}} \cdot E_M). \tag{2}$$

A GLM is used with new exposure $D_{cmi} = \mu_{\mathrm{cmi}} \cdot E_M$, target $\hat{D}_{\mathrm{M}} = \hat{\mu}_{\mathrm{M}} \cdot E_M$ and model specification as follows [45]:

$$log(\delta) = \beta_0 + \beta_{age} \cdot x_1 + ... + \beta_{gender} \cdot x_p + \log(D_{cmi}) \tag{3}$$

In the Poisson case, [46] demonstrated that the method is mathematically equivalent to using the ratio $\frac{D_{\mathrm{M}}}{D_{\mathrm{cmi}}}$ as target and D_{cmi} as weights:

$$log(\frac{D_{\mathrm{M}}}{D_{\mathrm{cmi}}}) = \beta_0 + \beta_{age} \cdot x_1 + ... + \beta_{gender} \cdot x_p \tag{4}$$

The validation of our approach, presented in Results section, includes comprehensive evaluation, such as its application to the UK insurance population and drift analysis from CMI mortality tables.

Due to exclusive usage of publicly available anonymized data (CMI and HMD) and aggregated, anonymized insurance data for model pretraining, there was no direct interaction with human participants, and no personally identifiable information was accessed. The insurance company data used for pretraining was provided in an aggregated and anonymized form, with no possibility of tracing back to any individual policyholder. No UK-specific data from the insurance company was used. The UK-specific results were derived entirely from publicly available data and a synthetic dataset generated for this study, with no real UK life insurance data being used. Therefore, this study does not involve new data collection from human participants and participant consent was not applicable.

Results

Transfer learning application in the UK

The following section introduces the application of the transfer learning framework to the UK, where internal mortality data is unavailable. This analysis establishes the foundation for subsequent discussions and demonstrates a high level of agreement with expected outcomes.

The point of Fig 4 is to show the plausible transfer of knowledge from the countries to the UK, according to their similarity. It is clear that the degree of proximity is more pronounced in Europe, and therefore it makes more sense to resample from there than from the Latin American countries.

Fig 5 shows the predicted number of deaths for the UK based on the transfer model for age and gender. The remaining variables are not disclosed as they are considered to be insurer-specific and require confidential background information for proper interpretation. The categories with the highest exposure and claims are based on more original data, indicating greater reliability of the estimation and deserving of our focus.

Furthermore addressing the second research question, we aim to evaluate the transfer model's accuracy in matching the expected age-gender mortality rates using agreement metrics. CMI stands in for expected mortality rates in the UK's insured population, given the lack of access to internal portfolio data. Despite differences in datasets and modeling, we regard

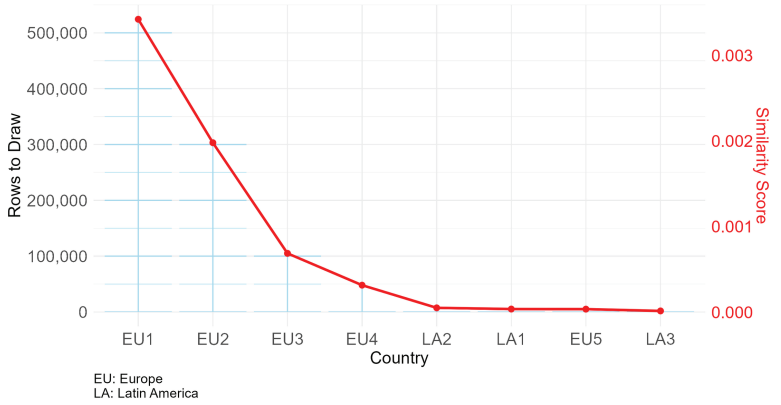


Fig 4. The composition of the exposure drawn from the countries for the synthetic dataset for the UK, proportional to the similarity score.

https://doi.org/10.1371/journal.pone.0313378.g004

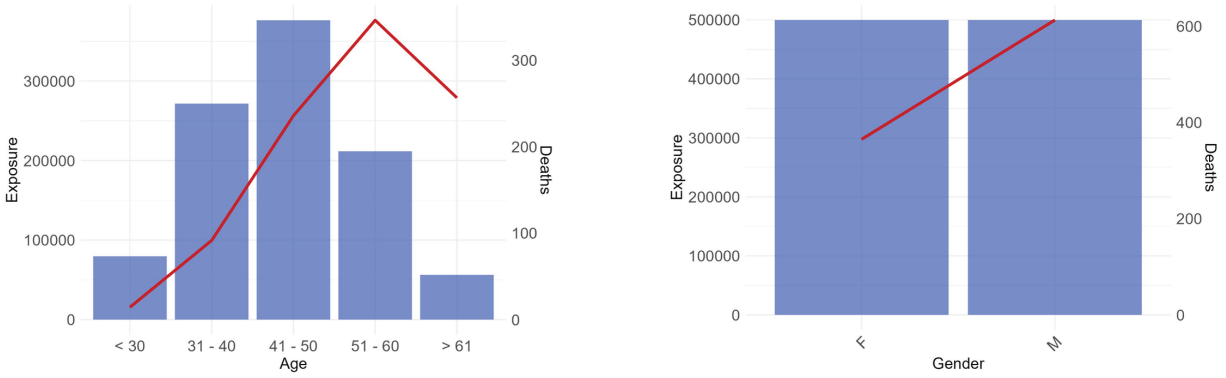


Fig 5. Exposure (bars) and predicted death counts (lines) by age and gender, derived from the synthetic-data-based transfer model. Age groups are defined retrospectively, and modeling is conducted using a metric scale. A. Age. B. Gender.

https://doi.org/10.1371/journal.pone.0313378.g005

CMI as a reliable proxy for UK policyholders' actual mortality rates. The analysis focused on transferring insights about the relative mortality impact from different features in the data as opposed to producing an accurate estimate of the overall rate of mortality. This decision was made in part because it is expected that data will be available in the receiving country to estimate the overall rate of mortality, either from publicly available resources, or more likely from internal data that better reflects the specifics of the cohort being considered. Therefore, for evaluation purposes, we use Spearman correlation, cosine similarity, and R-squared as agreement metrics. These metrics do not consider the agreement of the difference in average mortality, ensuring objectivity in our evaluation.

Table 3 provides these measures not only for the UK but also for 8 other countries in the pooled dataset, as the transfer model's predictive performance was also quantitatively examined for each of the 8 countries by pretraining the global model on the remaining seven. Given that the highest possible score is 1 for all metrics, we are within the highest acceptable range for the UK, as well as for the extended experiment.

Table 3. Evaluation metrics for different countries.

Country	Spearman Correlation	Cosine Similarity	R-squared
UK	0.9922	0.9878	0.9641
1	0.9221	0.9796	0.8641
2	0.8421	0.9253	0.8516
3	0.9711	0.9214	0.8912
4	0.9334	0.9658	0.8763
5	0.9242	0.9754	0.8773
6	0.8756	0.9564	0.8977
7	0.9360	0.8612	0.7822
8	0.9145	0.9700	0.8948

https://doi.org/10.1371/journal.pone.0313378.t003

While the table indicates a high level of precision in estimating the age-gender mortality using the transfer learning framework, the following section proposes using the Drift model to identify the cause of any remaining marginal discrepancies.

Fig 6 offers an initial insight into the disparities between the predictions of the transfer model and the CMI mortality rates, specifically examining age and gender. Despite an overall trend of underestimation in our estimates compared to CMI, our attention shifts to understanding the specific impacts of various features. Subsequently, we delve into the examination of age and gender as overlapping features present in both the predicted (transfer) and expected benchmark (CMI) mortality rates. To ensure monotonicity, it may be desirable to smooth the curves, i.e. to use them directly in pricing. We present our proposal for this in S3 Appendix, but in the main body we continue with the original version in order to remain faithful to the portfolio context and not to lose its specificity. Additionally, we offer the inclusion of confidence intervals through bootstrapping method as a validation technique, to provide a more detailed assessment of the uncertainty associated with the predictions. From Fig 7 it becomes evident that the confidence interval mostly contains the CMI, particularly

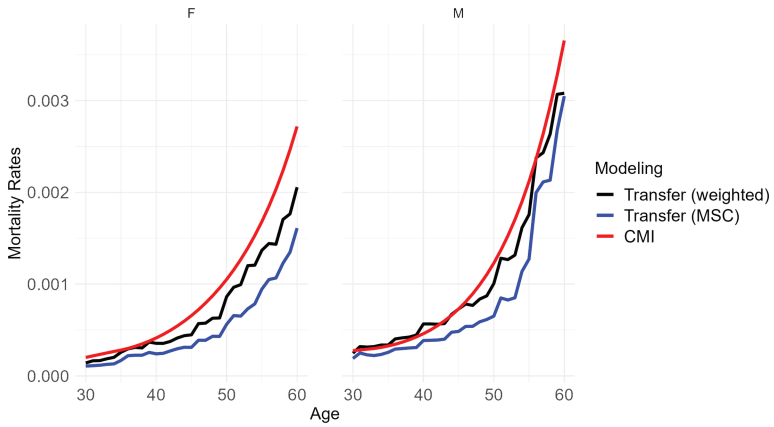


Fig 6. Comparison of UK mortality rates between Transfer Learning and CMI by age and gender. While transfer weighted by similarity score shows the above approach in black, the blue line shows the alternative of resampling only from the most similar country (MSC), which leads to a less accurate prediction.

https://doi.org/10.1371/journal.pone.0313378.g006

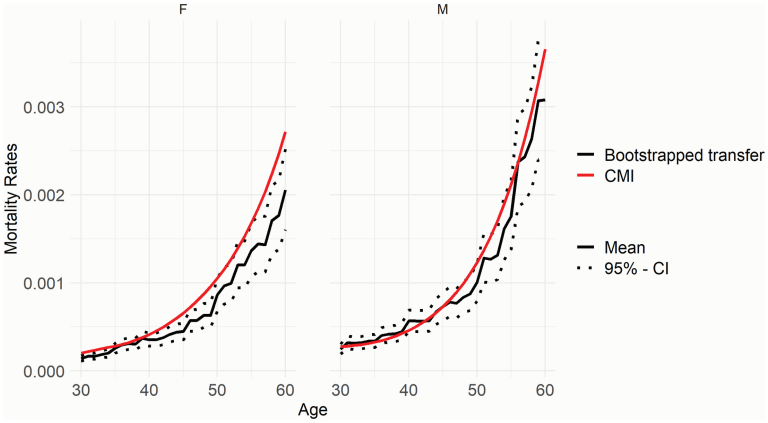


Fig 7. Bootstrap validation for 95% confidence interval of (weighted) transfer results.

https://doi.org/10.1371/journal.pone.0313378.g007

for males, solidifying the reliability of the results, especially given the reliance on synthetic datasets. The methodology details are documented in S4 Appendix.

Fig 8 illustrates the exponentiated coefficients of the Drift model, offering insights into the relationship between the two mortality tables by quantifying deviations from the average ratio. The red dashed line at approximately 0.5 represents the exponentiated intercept $exp(\hat{\beta}_0)$, indicating the average ratio across all features. An exponentiated effect of 1 for a specific feature implies no impact on the ratio, suggesting effective capture of pattern differences between the source and target countries for that feature.

The multiplicative effect of age in relation to the average ratio is approximately 1, indicating that age does not significantly influence the relationship between the transfer model and CMI. Although slight differences may exist in the age curve and average values, this suggests that the transfer learning framework effectively captured the shape variances between the other K (= 8) countries and the UK by age, resulting in a close replication of CMI. This

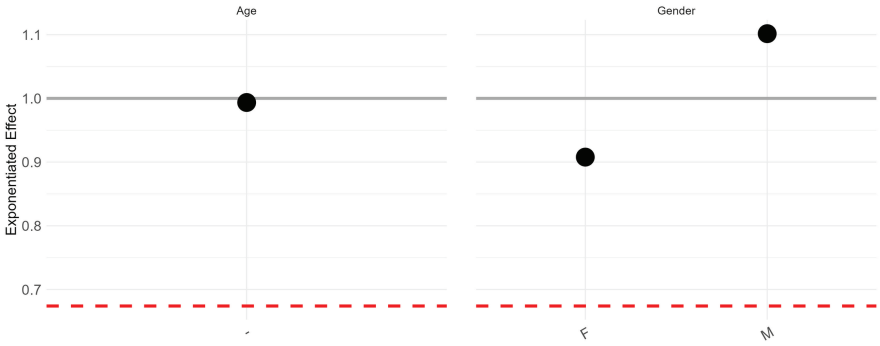


Fig 8. Exponentiated effects of age and gender on the ratio of transfer to CMI. The gray line represents the no-effect line, while the red dashed line is the exponentiated intercept.

https://doi.org/10.1371/journal.pone.0313378.g008

successful matching of age curves is a critical finding for insurance purposes, and lends confidence to subsequent analyses. Despite being from a different country, the methodology achieves a close match to the expected age curve, providing a strong basis for further analysis.

Regarding gender-specific mortality risks, while both the transferred results and CMI indicate higher mortality rates for males than females, the transferred estimations may show slight discrepancies: males are slightly overestimated and females underestimated compared to the average mortality risk. However, these deviations appear minor and likely stem from cohort distinctions between CMI and internal data, as well as cultural differences between the primary reference countries and the UK's insurance mortality data, possibly reflecting subtle cultural influences and evolving gender roles in different countries.

Building upon the strong alignment observed in the transfer learning process, the subsequent section investigates additional variables.

Improving baseline mortality through additional variables in the transfer model

The Drift model, which actually goes beyond age and gender, examines additional variables found in portfolio datasets but not included in the CMI. With the CMI serving as the insurer's base table, the exponentiated effects estimated by the Drift model for additional variables provide direct insight to insurers. This allows them to assess the potential impact of including these variables in the pricing model, and to determine possible loadings or discounts accordingly.

For example, considering Feature A with values A1, A2, A3, A4, A5, A6, absent from the CMI, Fig 9a shows that the predicted mortality rates increase from A1 to A6. Consequently, the Drift model's exponentiated effects reveal that policies falling under A1 have a 33% lower mortality ratio compared to the average, while those under A6 exhibit a 24% higher ratio, both ceteris paribus. Therefore, a UK insurer may include an extra risk factor in their pricing strategy due to the relative risk of A1 being approximately 54% (67/124) of A6. This justifies a 33% loading for A6 policyholders. It is suggested that selection effects would significantly impact the risk profile. The estimation of all other variables is presented in S3 Appendix.

In summary, the transfer learning framework effectively provides mortality risk predictions for the UK, leveraging a pretrained model from 8 other countries due to a lack of local mortality portfolio data, while refining the model using open-source UK total population mortality rates and data synthesized from the available countries accordingly to their similarity degree. While the model performs well with less culture-specific risk factors, discrepancies with CMI mortality tables highlight the need for evaluation using the Drift model. This is essential for comprehensive risk assessment and to inform pricing strategies, particularly in scenarios where data is not available.

Limitation of generalizability

Overall, the transfer model results provide notable advantages for generalizability, especially when new country data is entirely absent. It allows us to leverage existing models trained on data from other regions, thereby circumventing the need for extensive local data collection and reducing both time and resource requirements. By utilizing knowledge from a previously trained model, transfer learning can enhance performance in target countries that share similar characteristics with the source countries, effectively applying pretrained insights. However, the absence of local data presents unique challenges. Transfer learning is most effective when the source and target countries exhibit substantial similarity. As the disparity between these countries increases, the effectiveness of transfer learning diminishes. For example, applying

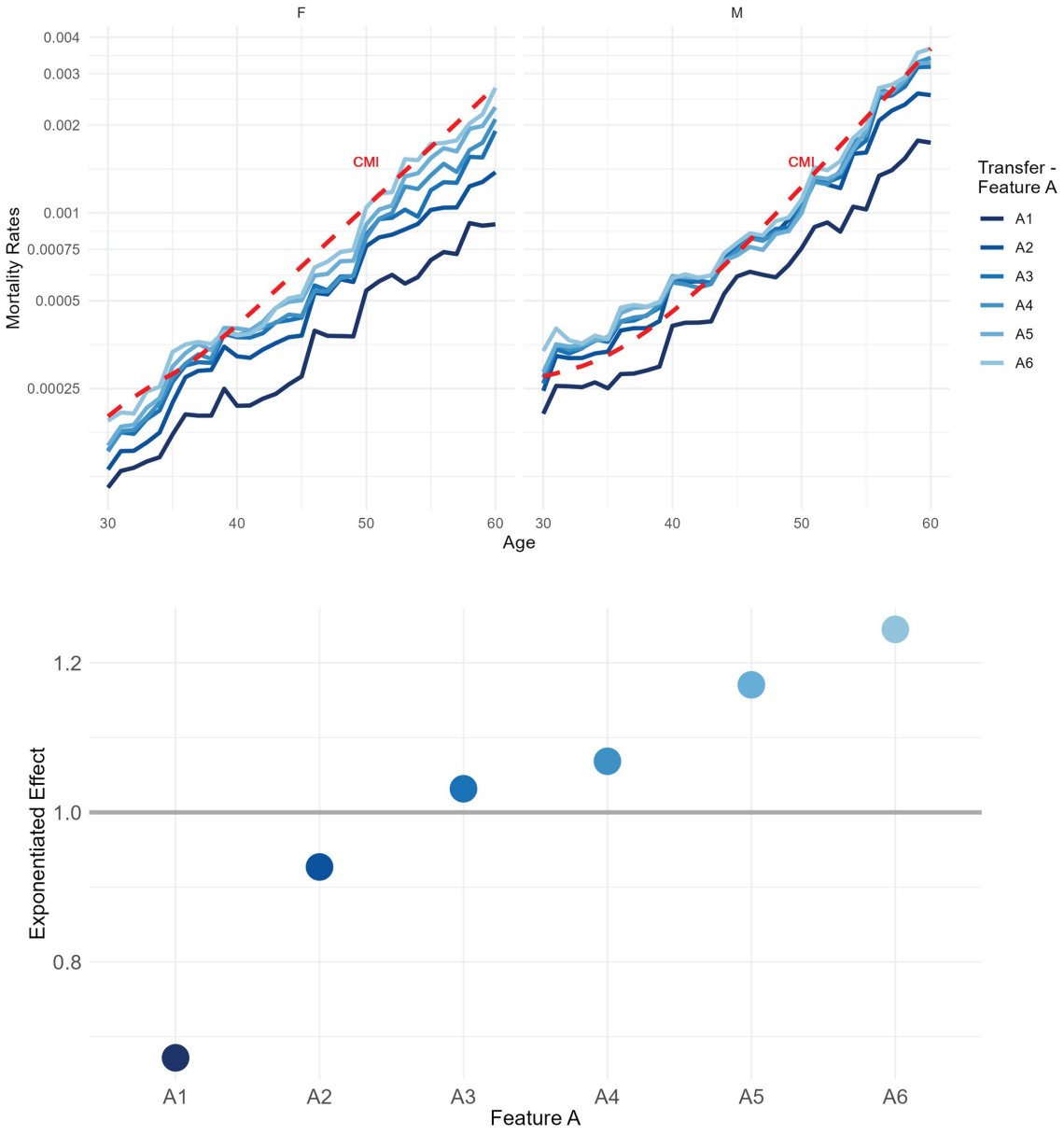


Fig 9. Feature A (with values A1-A6) evaluation as a risk factor for mortality. Transfer model results and evaluation of drift from CMI. (**A**) The mortality rates for the UK are displayed on a logarithmic scale, segmented by Feature A. Red line represents CMI mortality rates. (**B**) Exponentiated effects of Feature A on the ratio of transfer to CMI. The red line represents the exponentiated intercept, while the gray line represents the no-effect line.

https://doi.org/10.1371/journal.pone.0313378.g009

a model trained only on South American countries to predict outcomes in Asian countries may not be successful due to demographic, cultural, and economic differences. To address these challenges, we have implemented several mechanisms. The Country similarity index

considers external demographic, insurance, and mortality-specific characteristics, capturing the degree of similarity between countries. This index aids in selecting appropriate source countries, minimizing the risk of misaligned data transfer. Additionally, the Drift model helps analyze discrepancies between source and target countries, offering a tool to understand the limits of generalizability and the extent to which transfer learning can be applied. Bootstrapping confidence intervals provide an additional layer of validation, helping to understand potential biases and offering robust insights into model performance and reliability in regions lacking local data. In practical terms, while the transfer learning framework holds considerable promise, its generalizability in the absence of local data depends on the similarity between source and target countries. By incorporating mechanisms like the Country similarity index, Drift model, Bootstrapping confidence intervals we facilitate more informed and reliable applications in regions with differing cultural, demographic, or economic conditions, even when local data is completely missing.

Summary and outlook

This research presents a novel transfer learning framework designed to provide accurate mortality risk predictions for the UK, despite the complete absence of local mortality portfolio data. By leveraging pretrained and specialized models from eight other countries, along with UK population mortality rates obtained from open sources and synthesized data, we refine predictions for this data-scarce environment.

The framework establishes a solid foundation for mortality risk estimation and pricing, particularly benefiting small countries with insufficient data. Our predictive model shows strong agreement with the CMI mortality tables for age and gender, with only slight deviations detected via the Drift model. Expert validation further supports the inclusion of additional variables to enhance mortality risk estimation.

The approach offers several practical benefits, including strong predictive performance, reduced reliance on local data, and lower computational demands, making it efficient for multi-centre studies. It simplifies the development and deployment of ML models by eliminating the need for extensive training data in each new country. Our findings suggest that transfer learning is particularly effective for factors that are less influenced by cultural differences, although it may experience drift when capturing local specificities.

While the reliance on synthetic data helps overcome data scarcity, it may introduce uncertainties, particularly when source countries differ demographically or economically from the target country. The effectiveness of the Drift model also depends on the quality and similarity of external data used in the transfer learning process.

Future research could focus on addressing uncertainties in predictions by incorporating additional socio-economic and regional factors that may further improve mortality predictions. Expanding the framework to other regions and markets, especially those lacking sufficient local data, would provide valuable insights into its broader applicability. Testing the model in different settings could refine its use for life insurance product development in underserved demographic segments and emerging markets.

Supporting information

S1 Appendix. Methodology details of pretraining and specialization.
(PDF)

S2 Appendix. Lee-Carter model.
(PDF)

S3 Appendix. Additional results of the Drift model.
(PDF)

S4 Appendix. Bootstrap validation for confidence intervals.
(PDF)

S5 Appendix. Workflow for synthetic data generation.
(PDF)

Author contributions

Conceptualization: Asmik Nalmpatian, Christian Heumann.

Data curation: Asmik Nalmpatian, Levent Alkaya, William Jackson.

Formal analysis: Asmik Nalmpatian, Christian Heumann, Levent Alkaya.

Investigation: Asmik Nalmpatian, William Jackson.

Methodology: Asmik Nalmpatian, Christian Heumann, Levent Alkaya, William Jackson.

Project administration: Asmik Nalmpatian.

Resources: William Jackson.

Software: Asmik Nalmpatian, Levent Alkaya.

Supervision: Christian Heumann.

Validation: Asmik Nalmpatian, Christian Heumann, William Jackson.

Visualization: Asmik Nalmpatian.

Writing – original draft: Asmik Nalmpatian.

Writing – review & editing: Asmik Nalmpatian, Christian Heumann, Levent Alkaya, William Jackson.

References

1. Vincelli M. A machine learning approach to incorporating industry mortality table features into a company's insured mortality analysis. Soc Actuar Res Rep. 2019;2019(Sept):1–53.
2. Lim HB, Shyamalkumar ND. Incorporating industry stylized facts into mortality tables: Transfer learning with monotonicity constraints. 2024. Available from: https://papers.ssrn.com/abstract=3964181.
3. Yosinski J, Clune J, Bengio Y, Lipson H. How transferable are features in deep neural networks?. Advances in Neural Information Processing Systems 27 (NIPS 2014), vol. 27, 2014.
4. Tian Y, Feng Y. Transfer learning under high-dimensional generalized linear models. J Am Stat Assoc. 2022;118(544):1–14. https://doi.org/10.1080/01621459.2022.2071278
5. Gong JJ, Sundt TM, Rawn JD, Guttag JV. Instance weighting for patient-specific risk stratification models. In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2015. pp. 369–78.
6. Wiens J, Guttag J, Horvitz E. A study in transfer learning: leveraging data from multiple hospitals to enhance hospital-specific predictions. J Am Med Inform Assoc. 2014;21(4):699–706. https://doi.org/10.1136/amiajnl-2013-002162 PMID: 24481703.
7. Desautels T, Calvert J, Hoffman J, Mao Q, Jay M, Fletcher G, et al. Using transfer learning for improved mortality prediction in a data-scarce hospital setting. Biomed Inform Insights. 2017;9:1178222617712994. https://doi.org/10.1177/1178222617712994 PMID: 28638239

8. Lim HB, Shyamalkumar N. Incorporating industry stylized facts into mortality tables: transfer learning with monotonicity constraints. SSRN. 2021. Available from: https://ssrn.com/abstract=3964181.
9. Schelldorfer J, Wüthrich MV. Nesting classical actuarial models into neural networks. SSRN. 2019. Available from: https://ssrn.com/abstract=3320525.
10. Bauer DT, Ameringer CF. A framework for identifying similarities among countries to improve cross-national comparisons of health systems. Health Place. 2010;16(6):1129–35. https://doi.org/10.1016/j.healthplace.2010.07.004 PMID: 20675180.
11. Murray CJL, Frenk JD. Ranking 37th–measuring the performance of the US health care system. N Engl J Med. 2010;362(2):98–99. https://doi.org/10.1056/NEJMp0910064
12. Thorpe RJ Jr, Koster A, Kritchevsky SB, Newman AB, Harris T, Ayonayon HN, et al. Race, socioeconomic resources, and late-life mobility and decline: findings from the health, aging, and body composition study. J Gerontol A Biol Sci Med Sci. 2011;66(10):1114-23. https://doi.org/10.1093/gerona/glr102 PMID: 21743093.
13. Matus JC. A comparison of country's cultural dimensions and health outcomes. Healthcare (Basel). 2021;9(12):Article 12. Available from: https://doi.org/10.3390/healthcare9121234.
14. Hofstede G. National cultures revisited. Asia Pac J Manage 1984;2:22–28. https://doi.org/10.1007/BF01732507
15. McDowell I, Spasoff RA, Kristjansson B. On the classification of population health measurements. Am J Public Health. 2004;94(3):388-93. doi: 10.2105/ajph.94.3.388 PMID: 14998801.
16. Human Mortality Database (UK). Human Mortality Database. 2023. Available from: https://www.mortality.org/.
17. Oram E, Dash PB, Naik B, Nayak J, Vimal S, Nataraj SK. Light gradient boosting machine-based phishing webpage detection model using phisher website features of mimic URLs. Pattern Recognit Lett. 2021;152:100–106. https://doi.org/10.1016/j.patrec.2021.09.018
18. Singh K, Xie M. Bootstrap: a statistical method. Rutgers University, USA. 2008. pp. 1–14. Available from: https://statweb.rutgers.edu/mxie/rcpapers/bootstrap.pdf.
19. Continuous Mortality Investigation. Working Paper 154: Final "16" Series term assurance mortality and accelerated critical illness tables. 2021. Available from: https://www.actuaries.org.uk/learn-and-develop/continuous-mortality-investigation/cmi-working-papers/assurances/cmi-working-paper-154.
20. OECD. Insurance Indicators – Life insurance share. 2023. Available from: https://stats.oecd.org/index.aspx?queryid=25443.
21. OECD. Insurance Indicators – Density. 2023. Available from: https://stats.oecd.org/index.aspx?queryid=25445.
22. OECD. Insurance Indicators – Penetration. 2023. Available from: https://stats.oecd.org/index.aspx?queryid=25444.
23. OECD. Insurance Indicators – Total life gross premiums. 2023. Available from: https://stats.oecd.org/Index.aspx?DataSetCode=INSIND.
24. OECD. Insurance Indicators – Life retention ratio. 2023. Available from: https://stats.oecd.org/index.aspx?queryid=25441.
25. Global Residence Index. The Health Index. 2023. Available from: https://globalresidenceindex.com/hnwi-index/health-index/.
26. The World Bank. Indicator Data – Retirement Pension. 2023. Available from: https://wbl.worldbank.org/en/data/exploretopics/getting-a-job#Retirement
27. The World Bank. World Health Organization's Global Health Workforce Statistics - Physicians. 2023. Available from: https://data.worldbank.org/indicator/SH.MED.PHYS.ZS?most_recent_value_desc=true.
28. The World Bank. The Global Health Observatory – Hospital beds. 2023. Available from: https://www.who.int/data/gho/data/indicators/indicator-details/GHO/hospital-beds-(per-10-000-population).
29. ChartsBin.com. Basic health services by country. 2023. Available from: http://chartsbin.com/view/41517.
30. The World Bank. World Health Organization Global Health Expenditure database – Current health expenditure. 2023. Available from: https://data.worldbank.org/indicator/SH.XPD.CHEX.PC.CD.
31. The World Bank. The Program in Global Surgery and Social Change – Risk of impoverishing expenditure for surgical care. 2023. Available from: https://data.worldbank.org/indicator/SH.XPD.CHEX.PC.CD.
32. Human Mortality Database. HMD - United Kingdom Total Population. 2023. Available from: https://www.mortality.org/Country/Country?cntr=GBR_NP/.

33. Li S, Cai TT, Li H. Transfer learning for high-dimensional linear regression: prediction, estimation and minimax optimality. J R Stat Soc Ser B Stat Methodol. 2022;84(1):149–73. https://doi.org/10.1111/rssb.12479
34. Li S, Cai T, Duan R. Targeting underrepresented populations in precision medicine: a federated transfer learning approach. Ann Appl Stat. 2023;17(4):2970–92. https://doi.org/10.1214/23-AOAS1747
35. Xu K, Bastani H. Learning across bandits in high dimension via robust statistics. arXiv, preprint, arXiv:2112.14233. 2021.
36. Perlibakas V. Distance measures for PCA-based face recognition. Pattern Recognit Lett. 2004;25(6):711–24. https://doi.org/10.1016/j.patrec.2004.01.011
37. Reps JM, Rijnbeek PR, Ryan PB. Identifying the DEAD: development and validation of a patient-level model to predict death status in population-level claims data. Drug Saf. 2019;42(11):1377–86. https://doi.org/10.1007/s40264-019-00827-0 PMID: 31054141.
38. Carmona R, Delarue F, et al. Probabilistic theory of mean field games with applications I–II. Springer; 2018.
39. Gunay EE, Kula U. A two-stage stochastic rule-based model to determine pre-assembly buffer content. J Ind Eng Int. 2018;14:655–63. https://doi.org/10.1007/s40092-017-0252-4
40. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial networks. Commun ACM. 2020;63(11):139–44. https://doi.org/10.1145/3422622
41. Bonabeau E. Agent-based modeling: Methods and techniques for simulating human systems. Proc Natl Acad Sci U S A. 2002;99:7280–7. https://doi.org/10.1073/pnas.082080899
42. Simard PY, Steinkraus D, Platt JC, et al. Best practices for convolutional neural networks applied to visual document analysis. In: Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings, Edinburgh, UK, 2003, pp. 958–63. https://doi.org/10.1109/ICDAR.2003.1227801.
43. Bishop CM, Nasrabadi NM. Pattern recognition and machine learning. Springer; 2006.
44. Levantesi S, Pizzorusso V. Application of machine learning to mortality modeling and forecasting. Risks. 2019;7(1):26. https://doi.org/10.3390/risks7010026
45. Fahrmeir L, Kneib T, Lang S, Marx B. Regression models. Springer; 2013.
46. Yan J, Guszcza J, Flynn M, Wu CS. Applications of the offset in property-casualty predictive modeling. Casualty Actuarial Soc e-forum. 2009;1(1):366–85.

Part V.

IPF Simulation for Segmented Mortality

10. Mortality simulations for insured and general populations

Contributing article

Nalmpatian, A., and Heumann, C. (2025). Mortality simulations for insured and general populations. *arXiv:2502.08814v2.*

Code and data repository

https://github.com/asmiknalmpatian/Simulation-of-segmented-mortality-tables?tab=readme-ov-file

Copyright information

Author contributions

The idea to simulate segmented mortality datasets using Iterative Proportional Fitting (IPF) to support modeling for under-documented populations was proposed and developed by Asmik Nalmpatian. She designed the simulation framework, implemented the IPF algorithm, and generated and evaluated multiple synthetic mortality datasets. Christian Heumann contributed to the validation strategy and was actively involved in the methodological refinement. All statistical analyses, visualizations, and manuscript drafting were carried out by Asmik Nalmpatian. Christian Heumann provided supervision throughout and collaborated closely on the interpretation and review of the final manuscript.

Supplementary material

- Interactive Shiny app dashboard: https://advancedmortalitymodeling.shinyapps.io/simulate_mortality_tables_v1/

Mortality simulations for insured and general populations

Asmik Nalmpatian[a], Christian Heumann[a]

[a]*Department of Statistics, LMU Munich, Germany*

Abstract

This study presents a framework for high-resolution mortality simulations tailored to insured and general populations. Due to the scarcity of detailed demographic-specific mortality data, we leverage Iterative Proportional Fitting (IPF) and Monte Carlo simulations to generate refined mortality tables that incorporate age, gender, smoker status, and regional distributions. This methodology enhances public health planning and actuarial analysis by providing enriched datasets for improved life expectancy projections and insurance product development.

Keywords: mortality, simulation, actuarial science, smoker status, insured population, statistical modeling

1. Statement of need

Detailed and disaggregated mortality simulations are critical for understanding variations in mortality risk across different demographic groups. However, acquiring high-quality, granular mortality datasets is challenging due to privacy restrictions, proprietary control over insurance data, and legal barriers to data sharing. This lack of detailed data affects public health policy, risk assessment, and insurance calculations.

Current efforts, while valuable, often suffer from limited scope, resolution, or are confined to specific demographics. For instance, methodologies for estimating mortality rates from narrow age windows (Goldstein et al., 2023), small-area mortality estimation (Denecke et al., 2023), and COVID-related predictions (Duchemin et al., 2022) demonstrate the utility of such

approaches but also underscore the inadequacy of existing data for high-resolution research. Further studies have shown the potential of granularity for improving mortality modeling but also highlight the challenges associated with data standardization and accessibility (RKI, 2014; El Emam et al., 2011; Nusselder and Mackenbach, 1997).

For insured populations, precise mortality estimates are essential for setting fair premiums, evaluating longevity risk, and designing insurance products that accurately reflect demographic differences. In the absence of comprehensive datasets, actuaries and researchers must rely on aggregated data, leading to potential biases in mortality estimates.

This study introduces a simulation-based framework that overcomes these limitations by generating synthetic but statistically accurate mortality datasets. By enriching mortality tables with demographic covariates, we enable more precise analysis of mortality trends, supporting both public health initiatives and actuarial applications.

2. Notation

In this section, we provide a summary of the notation and symbols used throughout the paper for clarity and ease of reference. Our analyses are based on multi-dimensional demographic cells (e.g., combinations of age, gender, smoker status, region, etc.), which are often indexed using multiple subscripts. To simplify later model specification, we introduce a unified indexing scheme that maps each multi-dimensional demographic subgroup to a single index.

- x_{ijk}: Count (e.g., population or deaths) in the demographic subgroup defined by the combination of dimensions i, j, and k. For example, i might index age groups, j gender, and k smoker status.
- $x_{ij\cdot}$: Marginal count obtained by summing over the third dimension (e.g., smoker status), i.e., $x_{ij\cdot} = \sum_k x_{ijk}$.
- π_{ijk}: Joint probability (e.g., population share) associated with subgroup (i, j, k).

To facilitate model estimation, we collapse and re-index the multidimensional demographic structure into a single flat index i, where each value of

i corresponds to a unique combination of categorical levels across all dimensions (e.g., a 40-year-old female smoker in Bavaria). This one-dimensional indexing refers to demographic subgroups — not individual persons — and simplifies notation in subsequent modeling steps such as regression:

- D_i: Observed number of deaths in the insured population for demographic subgroup i.
- D_i^P: Observed number of deaths in the general population for demographic subgroup i.
- E_i: Exposure (e.g., population size or person-years) for subgroup i.
- μ_i: Mortality rate for subgroup i in the insured population, to be estimated.
- $\hat{\mu}_i$: Estimated mortality rate for subgroup i in the insured population.
- $f_1(\text{age}_i)$: Smooth function capturing the non-linear effect of age on mortality.
- $f_2(D_i^P)$: Smooth function capturing the relationship between deaths in the general population and mortality in the insured population.
- $\text{gender}_i \times \text{smoker}_i$: Interaction term indicating combined effect of gender and smoking status in the model.

Throughout the paper, we use the term *demographic subgroup* to refer to a unique combination of variables such as age, gender, region, and smoker status. When referring to the index i, we mean a specific demographic subgroup (not an individual), and in the context of modeling, we treat each subgroup as one observation unit.

3. Methodology

To address the challenge of generating high-resolution mortality data, our methodological framework proceeds in three key stages. It combines demographic inference, synthetic data generation, and advanced statistical modeling to create reliable and granular mortality estimates for both insured and general populations:

1. We start by estimating mortality rates using available marginal distributions of demographic variables such as age, gender, and smoker sta-

tus. Due to limitations in fully observed data, we incorporate known constraints via marginals to approximate mortality across subgroups.

2. Using Iterative Proportional Fitting (IPF), we derive joint distributions over the population structure and associated mortality patterns that are consistent with the known marginals. These joint distributions serve as the basis for generating new data via Monte Carlo simulation, where death counts are sampled from Poisson distributions according to the inferred demographic composition.

3. The simulated datasets are then used to estimate mortality rates with greater granularity. Specifically, we apply Generalized Additive Models (GAMs) with Poisson assumptions and demographic covariates to account for non-linear effects and interactions, enabling flexible and robust predictions even in sparse data settings. This modeling step enables us to infer insured population mortality rates from general population data, particularly in countries where insured-specific data is limited or unavailable.

3.1. Iterative Proportional Fitting

IPF is a widely used deterministic method for adjusting contingency tables to match known marginal totals and has been a cornerstone in statistical analysis since its introduction (Deming and Stephan, 1940). It iteratively refines initial estimates to ensure consistency across multiple demographic dimensions while preserving the structure of the observed data. Renowned for its efficiency and robustness, IPF calculates non-integer weights that reflect how representative each individual is within each zone, effectively reweighting the data to align with known marginal totals. This method is particularly advantageous in scenarios requiring the estimation of internal cells of a matrix based on these marginals, as it maximizes entropy by exploring the number of configurations that could yield the same marginal counts (Cleave et al., 1995).

The IPF process involves iteratively adjusting an input matrix to ensure that its internal cells align with given marginal totals, which typically represent known values across an entire population. For example, in voter migration analysis, the input matrix might represent voter preferences across

different election years, with known marginal totals indicating actual vote distributions. Each iteration of IPF refines the matrix by alternately adjusting row and column totals to match the respective marginal distributions, using Maximum Likelihood estimation to update internal cell values. However, convergence is not always guaranteed, particularly when zero entries are present, necessitating practical constraints such as iteration limits or tolerance thresholds for deviations (Pukelsheim, 2014).

In our context, IPF is employed to calculate multi-dimensional distributions essential for population simulations. Given that mortality data comprises populations and deaths within each subgroup, our objective is to determine the joint distribution for each additional variable. For instance, knowing the age and state population distributions, we aim to compute the joint distribution across age and state categories. Consider a multiway table in N dimensions, each representing a sociodemographic variable. For illustrative purposes, assume $N = 3$. The multiway table π_{ijk} contains unknown components, subject to constraints defined by marginal distributions $\{x_{ij\cdot}, x_{i\cdot k}, x_{\cdot jk}\}$. The constraints ensure that the sum of observations in each category matches the known marginals and the total number of observations, n. The IPF process begins with an initial estimate $\pi^{(0)}_{ijk}$ and proceeds through iterations to adjust the table according to the given marginals. The algorithm can be extended to higher dimensions, facilitating the synthesis of population data at varying resolutions. For instance, when considering three demographic variables, one iteration of the IPF process can be represented as follows:

$$\pi^{(1)}_{ijk} = \frac{1}{n} \frac{x_{ij\cdot} \pi^{(0)}_{ijk}}{\pi^{(0)}_{ij\cdot}} \tag{1}$$

$$\pi^{(2)}_{ijk} = \frac{1}{n} \frac{x_{i\cdot k} \pi^{(1)}_{ijk}}{\pi^{(1)}_{i\cdot k}} \tag{2}$$

$$\pi^{(3)}_{ijk} = \frac{1}{n} \frac{x_{\cdot jk} \pi^{(2)}_{ijk}}{\pi^{(2)}_{\cdot jk}} \tag{3}$$

Each equation represents an update step where the estimated cell probability π_{ijk} is iteratively adjusted to match the given marginals. Specifically,

equation (1) adjusts the initial estimate $\pi_{ijk}^{(0)}$ to align with the marginal totals $x_{ij\cdot}$, ensuring consistency along the first dimension. Equation (2) further refines π_{ijk} using the marginal totals $x_{i\cdot k}$ from the second dimension. Equation (3) completes the iteration by incorporating $x_{\cdot jk}$, ensuring alignment with the third dimension.

This iterative process continues until convergence, ensuring that the synthesized dataset accurately represents the given marginal distributions across all dimensions (Agresti, 2012).

Incorporating additional variables, such as smoker status, into mortality risk assessments requires accounting for distinct mortality risks while keeping all other characteristics constant. By applying known hazard ratios for different categories, we can refine mortality tables to reflect these differences accurately. Specifically, we first estimate total deaths using age-gender-specific mortality rates for a hypothetical population of 100.000. Then, using the given hazard ratios, we allocate these deaths proportionally across smoker and non-smoker groups of the same total size. This approach ensures that the original age-gender mortality risks are preserved within each subgroup while maintaining the intended hazard ratio structure.

We implemented our methodology using the `mipfp` R package. For multi-dimensional interactions (e.g., age-gender, gender-smoker), there are two possible approaches:

1. **Separate IPF runs:** One option is to run IPF separately for different subgroups (e.g., separately for males and females) while ensuring that each subgroup aligns with the corresponding one-dimensional marginal distributions (e.g., for age, state, and smoker status).

2. **Incorporating cross-tabulated constraints:** Alternatively, the `mipfp` package allows for directly incorporating interactions by using cross-tabulated marginal distributions (e.g., age-gender bivariate marginals). This approach provides a more compact implementation, reducing the degrees of freedom for the algorithm and enabling faster convergence without compromising accuracy.

The advantage of including cross-tabulated constraints is that it ensures dependencies between variables are explicitly modeled, which becomes in-

creasingly relevant as the number of interaction dimensions grows. This results in a more efficient and scalable implementation, particularly when dealing with complex dependencies among demographic variables.

In summary, IPF serves as a foundational method for population and death synthesis, enabling the creation of detailed and accurate demographic distributions necessary for high-resolution mortality data simulations.

3.2. Monte Carlo Simulation

When analytical solutions are unavailable, Monte Carlo simulations provide a solid alternative by approximating these expectations through the simulation of random processes. Using predefined probability distributions, we generate synthetic mortality scenarios that allow for variability assessment. By averaging the simulated values, we obtain estimates that often closely approximate the true expectations. This approach leverages the principle that sample averages are frequently reliable estimators of their corresponding population expectations (Robert and Casella, 2004):

$$\bar{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \to \theta = \mathbb{E}[X]$$

This convergence is underpinned by the assumption that the data are independent and identically distributed (iid) from a distribution with finite variance. The Central Limit Theorem (CLT) provides the convergence in distribution of the sample mean to a normal distribution:

$$\sqrt{n}(\bar{\theta}_n - \theta') \xrightarrow{d} \mathcal{N}(0, \sigma^2)$$

where $\sigma^2 = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$ represents the variance of the underlying distribution. This theorem is instrumental in constructing approximate confidence intervals for the Monte Carlo error, providing a measure of the reliability of our estimates.

Thus, Monte Carlo simulations are employed in this study to generate repeated samples from Poisson distributions, which are used to model count data such as yearly deaths given population size as exposure. This probabilistic approach allows us to quantify the variability and uncertainty of mortality projections. For a Poisson distribution, the variance is equal to

its expected value, which we utilize to assess the dispersion of our mortality estimates. This framework is essential for ensuring that simulated distributions align with empirical observations. While the mean mortality rate remains unchanged, Monte Carlo provides insights into variance, skewness, and extreme outcomes, helping to better understand the probability of rare but significant deviations (tail risks).

3.3. Generalized Additive Models

GAMs offer a flexible approach for estimating mortality rates in insured populations by leveraging population-level mortality data and incorporating demographic variables such as age, gender and smoker status. The model assumes that the observed number of insured deaths (D_i) follows a Poisson distribution, a common choice for modeling count data in mortality studies.

The GAM framework is specified with Poisson distributional assumption and log-link. The use of Poisson regression ensures non-negativity of predicted counts and facilitates interpretability through the log-link function. Incorporating smooth terms enhances the model's ability to capture these patterns while avoiding overfitting. The Poisson framework and GAM methodology are well-established in demographic and actuarial research. Studies such as McCullagh and Nelder (1989) and Haberman and Renshaw (1996) highlight the use of generalized linear models, including Poisson regression, for mortality analysis. Additionally, Currie et al. (2004) demonstrate the advantages of smoothing techniques for estimating mortality rates in sparse data settings. The inclusion of the offset term, $\log(E_i)$, ensures that the model predicts mortality rates rather than raw death counts, enabling meaningful comparisons across demographic groups with varying levels of exposure.

$$D_i \sim \text{Poisson}(\mu_i \cdot E_i), \tag{4}$$

Thus, the proposed model for expected insured mortality rates $\hat{\mu}_i$ is as follows:

$$\log(\hat{\mu}_i) = f_1(\text{age}_i) + f_2(D_i^P) \cdot \text{gender}_i \times \text{smoker}_i + \log(E_i) \tag{5}$$

To ensure reliable estimates in countries where insured mortality rates are unavailable, we train the model on data from the most similar country where insured rates exist. We assume that the ratio between insured and general population mortality rates remains constant across comparable demographic variables between source and target country. If this assumption is difficult to justify, existing research on country similarity scores, based on insurance and mortality characteristics, can provide guidance (Nalmpatian et al., 2024). These scores help identify the most analogous countries for model training and adjustment, thereby improving the robustness of mortality rate predictions.

Overall, the proposed model provides a robust framework for predicting insured mortality rates by leveraging population-level data and demographic segmentation. Its foundation in Poisson regression and the incorporation of GAM smooth terms make it particularly well-suited for handling the complexities of mortality data.

4. Application

To demonstrate the applicability of our methodology in generating granular mortality data for both insured and general population, we explore three distinct scenarios for Germany, Italy, and Switzerland. Mortality data typically consists of exposure (i.e., population) and death counts, and the IPF method can be applied to both.

Scenario 1 focuses on enhancing demographic precision while assuming uniform mortality rates across states. Scenarios 2 and 3 incorporate an additional mortality risk factor (smoker status) with distinct hazard rates, while Scenario 3 further extends the methodology to general population data by incorporating an insured population adjustment.

The application begins by selecting relevant demographic variables and loading distributional assumptions from available general population data (Table 1), under the assumption that similar patterns apply to insured populations. If specific insured population data is available, it can be directly incorporated to improve accuracy.

Table 1: Overview of data sources for marginal distributions by country

	Germany	Italy	Switzerland
Population and deaths by age and gender	HMD (2023)	HMD (2023)	HMD (2023)
Population by smoker and gender	Zeiher et al. (2017)	Semyonov et al. (2012)	Gmel et al. (2017)
Population by state	Destatis (2025)	ISTAT (2025)	BFS (2025)
Hazard rates smokers vs. non-smokers	–	Menotti et al. (2014)	McEvoy et al. (2012)
Base mortality rates (general population)	HMD (2023)	HMD (2023)	HMD (2023)
Base mortality rates (insured population)	DAV (2022)	ANIA (2014)	–

4.1. Scenario 1: Enhancing population granularity using base insurance mortality tables

We begin with a base mortality table segmented by age, gender, and smoker status for the insured population in Germany. The objective is to improve demographic precision by incorporating state-level variations while assuming uniform mortality rates across states. Using marginal demographic distributions (age-gender, smoker-gender, and state) along with age-gender-smoker-specific DAV insurance rates, we disaggregate mortality data to the state level using IPF and generate Monte Carlo simulations. This approach enhances granularity without introducing additional mortality risk differentiation and is extendable to other demographic variables. This scenario exemplifies a minimal input data case, demonstrating what can be achieved when only marginal population distributions of an additional variable are available. It highlights the capability of IPF to enhance segmentation by adding one extra demographic variable (state), even in the absence of direct state-specific mortality data. Although we do not possess state-specific mortality rates, death counts still vary across states because the mortality rates are applied to state-segmented population distributions, reflecting differentiated demographic patterns. Simultaneously, Monte Carlo simulations assist in quantifying uncertainty in mortality rates by generating confidence intervals that incorporate population segmentation effects. This is particularly crucial for small states, where mortality estimates can be highly uncertain. The Poisson distribution assumes that the variance equals the mean death count, resulting in different variances for each state.

4.2. Scenario 2: Accounting for distinct mortality risks in addition to population granularity

Unlike Scenario 1, this scenario introduces an additional dimension of mortality risk differentiation while refining demographic segmentation. Starting with a base mortality table segmented by age and gender for the insured population in Italy, we extend mortality data to include smoker status and state-level variations. We assume that smokers and non-smokers exhibit distinct hazard ratios, requiring separate mortality rate estimates for each group. This enables a more realistic and differentiated mortality structure while preserving demographic precision. In summary, while Scenario 1 uses IPF to refine population segmentation with fixed mortality rates, in Scenario 2, we extend this by disaggregating death counts while keeping the population constant, thereby refining mortality rates segmentation. Of course, if state-specific mortality data were available, it could be directly incorporated. However, the goal of Scenario 1 is to illustrate how demographic refinements alone (without additional mortality data) already add value.

4.3. Scenario 3: Extending granular mortality data to the general population

This scenario builds upon Scenario 2 but begins with a base mortality table for the general population instead of the insured population. The objective is to generate an age-gender-smoker-state mortality table for the entire population, not just insured individuals. Additionally, assuming proportional relationships between insured and general populations in both the target (Switzerland) and source (Germany) countries, we employ a GAM with Poisson regression and an offset to infer mortality estimates from the general to the insured population. This approach demonstrates how base population mortality rates can be adjusted to reflect insured-specific risk characteristics.

Beyond pure simulated mortality data, we provide visual analyses to facilitate comparisons between simulated, population, and insured mortality rates across multiple countries. These visualizations offer an intuitive means of evaluating the plausibility and consistency of the simulated rates. Furthermore, the application includes a comprehensive suite of validation tests

to ensure data integrity and accuracy in rate calculations. These tests verify the consistency of demographic proportions and hazard ratios, reinforcing the reliability of the simulated datasets and derived insights.

5. Results

This section details the outcomes of our study, focusing on the disaggregation of mortality data using IPF and Monte Carlo simulations across various countries. The results are accessible for review and download via an interactive Shiny app dashboard, which includes a 95% confidence interval. The app, along with the code and datasets, is freely available on GitHub.

For Germany, we disaggregated the population data using known marginal distributions from open sources, assuming that the insurance population mirrors the general population. The distributions for gender-smoker, state, and age-gender are presented in Tables 2, 3, and 4, respectively.

Table 2: Smoker-gender population distribution

Smoker	**Gender**	
	Female	Male
Yes	20.8	27.0
No	79.2	73.0

Using these distributions, we applied IPF to obtain the joint age-gender-smoker-state distribution. Table 5 shows a portion of the resulting distribution.

Assuming a population size of 1 million, we utilized the derived distribution to estimate expected deaths by applying it to the base mortality table. This involved drawing samples and calculating expected mortality figures, which were then used as inputs for Monte Carlo simulations. Through these simulations, we established 95% confidence intervals by identifying the 2.5th and 97.5th percentiles of the simulated mortality rates. Figure 1 provides a detailed visualization of the final mortality rates for Germany, categorized by state, gender, and smoker status. The figure reveals that smaller states exhibit wider confidence intervals, indicating greater variability and uncertainty in mortality estimates due to their smaller population sizes. Smokers

Table 3: State population distribution

State	Population
Baden-Württemberg	13.4
Bayern	15.9
Berlin	4.47
Brandenburg	3.05
Bremen	0.817
Hamburg	2.26
Hessen	7.58
Mecklenburg-Vorpommern	1.92
Niedersachsen	9.64
Nordrhein-Westfalen	21.5
Rheinland-Pfalz	4.93
Saarland	1.17
Sachsen	4.83
Sachsen-Anhalt	2.58
Schleswig-Holstein	3.50
Thüringen	2.51

Table 4: Age-gender population distribution

Age	Gender	
	Female	Male
20	1.439913	1.5818712
21	1.507098	1.6599224
22	1.503754	1.6463640
23	1.483638	1.6237836
24	1.515971	1.6515359
25	1.573950	1.7035385
26	1.609090	1.7303510
27	1.671727	1.7916157
28	1.823225	1.9605025
29	1.809394	1.9270780
30	1.846709	1.9747674
31	1.812673	1.9276493
32	1.790005	1.8826642
33	1.739514	1.8217951
...	...	...

demonstrate higher mortality rates compared to non-smokers across all de-

Table 5: Result after IPF: Age-gender-state-smoker population distribution

Age	Gender	State	Smoker	Population
20	M	Baden-Württemberg	Yes	0.02852311
21	M	Baden-Württemberg	Yes	0.02993047
22	M	Baden-Württemberg	Yes	0.02968600
23	M	Baden-Württemberg	Yes	0.02927884
24	M	Baden-Württemberg	Yes	0.02977925
25	M	Baden-Württemberg	Yes	0.03071692
26	M	Baden-Württemberg	Yes	0.03120039
27	M	Baden-Württemberg	Yes	0.03230506
28	M	Baden-Württemberg	Yes	0.03535030
29	M	Baden-Württemberg	Yes	0.03474762
30	M	Baden-Württemberg	Yes	0.03560752
31	M	Baden-Württemberg	Yes	0.03475792
32	M	Baden-Württemberg	Yes	0.03394678
33	M	Baden-Württemberg	Yes	0.03284924
...	...	...	...	...

mographic groups, highlighting the essential impact of smoking on mortality. Additionally, males consistently show higher mortality rates than females, underscoring gender as a critical factor in mortality risk assessment. These observed trends are consistent across all states, reflecting our model's ability to account for the distribution of the population across different states. While we assume that the mortality rates themselves are consistent across states, the model adjusts for the proportions of the population within each state. This means that the model effectively captures demographic patterns in mortality by considering how populations are distributed across states. The consistency in trends highlights the ability of our methodology in applying these demographic distributions accurately.

Aggregating over all states, Figure 2 shows that simulated mortality rates align with the base table. For smokers, insurance mortality rates exceed population rates, whereas non-smokers show the opposite trend.

For Italy, since the original base table lacked smoker distinction, we first disaggregated the base mortality table using IPF, starting with age-gender specific mortality data (Table 6) from the ANIA insurance population. We applied a hazard ratio of 1.4 to distinguish between smokers and non-smokers,

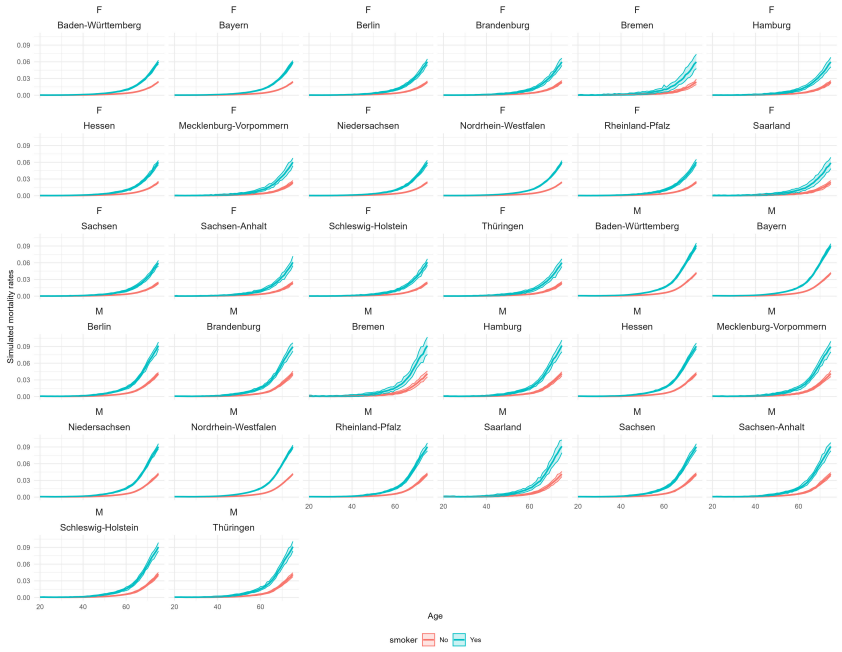


Figure 1: Simulated mortality rates for Germany.

based on the marginal mortality risks (0.014 vs. 0.010). While this ratio was applied uniformly across all subgroups in our primary scenario, the methodology allows for hazard ratios to be specified in a more granular way—varying across age-gender combinations or even higher-dimensional interactions if such detailed information is available.

The resulting age-gender-smoker mortality rates are shown in Table 7 and Figure 3. The curves maintain their shape but shift upwards for smokers and downwards for non-smokers, according to the predefined hazard ratio.

Figure 4 demonstrates that, unlike Germany, Italy's population mortality rates for both smokers and non-smokers are generally lower.

For Switzerland, the base table lacked smoker distinction and was derived from the general population. Disaggregation into smoker and non-smoker

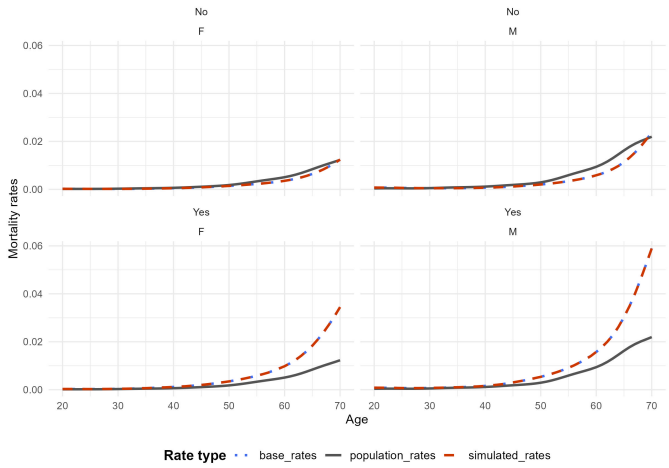


Figure 2: Aggregated base (insurance), simulated and population mortality rates for Germany.

Table 6: Age-gender mortality rates for insurance population in Italy

Age	Gender	Rates
20	M	0.000532
21	M	0.000526
22	M	0.000518
23	M	0.000508
24	M	0.000492
25	M	0.000506
26	M	0.000528
27	M	0.000572
28	M	0.000634
29	M	0.000705
...	...	...

categories resulted in distinct mortality curves. Assuming the insured-to-general population ratio mirrors that of Germany, we predicted Swiss population trends, as shown in Figure 5. This assumption validates the consistency of our methodology across different national contexts.

Overall, the results demonstrate the effectiveness of our methodology in

Table 7: Resulting age-gender-smoker mortality rates for insurance population in Italy

Age	Gender	Smoker	Rates
20	M	Yes	0.000621
20	M	No	0.000444
21	M	Yes	0.000614
21	M	No	0.000438
22	M	Yes	0.000604
22	M	No	0.000431
23	M	Yes	0.000593
23	M	No	0.000424
24	M	Yes	0.000574
24	M	No	0.000410

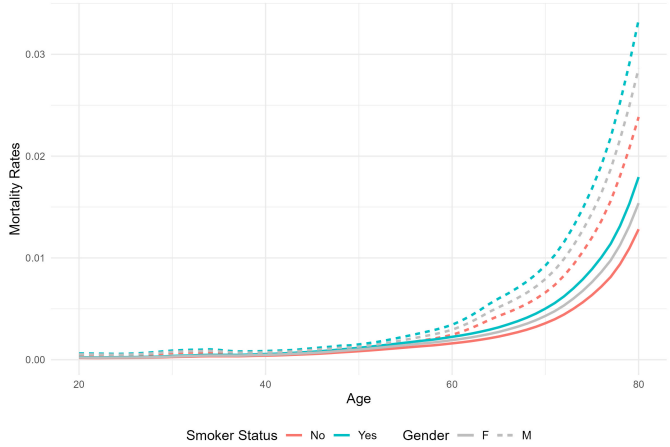


Figure 3: Disaggregated base mortality table in Italy with IPF.

disaggregating and analyzing mortality data across different countries, providing valuable insights into population-specific mortality trends.

6. Limitations

Our framework lays a strong foundation for mortality simulations in both insured and general populations, howver there are several limitations that

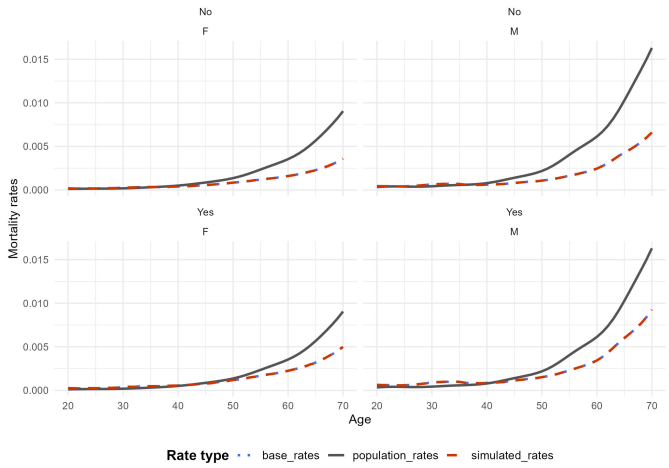


Figure 4: Aggregated base (insurance), simulated and population mortality rates for Italy

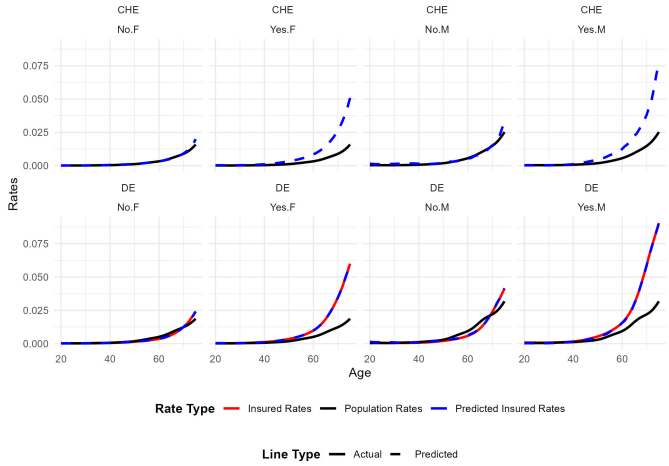


Figure 5: Inferring insurance mortality for Switzerland based on Germany

present opportunities for future research:

A key limitation in Scenarios 1 and 2 is the potential for selection bias when general population marginals are used in the absence of insured-specific data. Our current approach allows for insured-specific marginals to be inputted when available, which would directly incorporate these differences into the model. However, when such data is unavailable, we use general population marginals as a reasonable approximation. This method acknowledges that some selection effects, such as smoker prevalence, may not be fully captured. An alternative approach could involve adjusting the IPF method to explicitly model selection effects, though this would still require assumptions about the insured distribution if direct data were unavailable. To address the limitations of using general population data, Scenario 3 employs a GAM with Poisson regression. This approach adjusts insured mortality estimates based on observed demographic differences, helping to account for systematic differences between insured and general mortality patterns beyond simple demographic matching. This adjustment highlights the need for more sophisticated modeling techniques when insured-specific marginals are not available. Future research could integrate additional data sources, like coverage amounts or policy duration, to better model selection effects.

The current model's effectiveness is contingent on the availability and granularity of demographic data. While the methodology allows for extensions to additional demographic variables, the primary challenge remains obtaining sufficiently granular data to support these extensions. For Germany for example, we disaggregate population by state and apply uniform mortality rates, assuming that differences in mortality stem from demographic composition rather than state-specific factors. This simplification overlooks regional disparities in healthcare, environment, or socioeconomic conditions due to the absence of state-level mortality data. Future research could focus on expanding data sources and improving data collection methods.

While Monte Carlo simulations help quantify uncertainty, our approach assumes independent mortality realizations across subgroups. In reality, mortality risks may be correlated across demographic groups, influenced by shared socioeconomic factors. Future work could explore these dependencies to offer more elaborated risk assessments.

Our framework is adaptable to various countries, yet its accuracy hinges on data availability. We have incorporated data from Germany, Italy, and Switzerland, but the quality and granularity of inputs differ across regions. Further validation with additional datasets would be beneficial to assess the approach's generalizability to other markets.

7. Summary

In this study, we addressed the challenge of simulating detailed mortality data for both insured and general populations. By integrating multi-dimensional distributional constraints, we employed IPF, enabling the handling of complex demographic interactions and the application of Monte Carlo simulations. Our approach leverages the `mipfp` R package, facilitating efficient and scalable modeling of population distributions while maintaining accuracy.

We disaggregated mortality data, including both population and death counts, for Germany, Italy, and Switzerland, taking into account demographic distributions like age, gender, smoker status, and state, along with their interactions. Our findings show that the simulated mortality rates closely match the base tables when aggregated at a higher level. They also provide significant insights into demographic impacts on mortality at a more granular level, generating synthetic insured and general populations while preserving realistic distributional assumptions.

As a prototype, this study presents a robust, privacy-compliant methodology that advances mortality research and actuarial science. Each scenario can be further extended to include more countries, additional variables, or more complex dimensional interactions. The tools and datasets developed are accessible through an open-source interactive dashboard, promoting transparency and further research opportunities. Additionally, the code is available for reproducibility and potential extensions. For an overview of insurance mortality tables from other countries, please refer to the OECD (2023) publication.

References

Agresti, A. (2012). *Categorical data analysis*. John Wiley & Sons.

ANIA (2014). Le basi demografiche per rendite vitalizie a1900-2020 e a62. Accessed: 2025-02-05.

BFS (2025). Swiss federal statistical office (bfs). population statistics. Accessed: 2025-02-05.

Cleave, M. et al. (1995). Entropy maximization and bayesian analysis in statistical theory. *Journal of Statistical Planning and Inference*, 47(2):123–137.

Currie, I. D., Durban, M., and Eilers, P. H. (2004). Smoothing and forecasting mortality rates. *Statistical Modelling*, 4(4):279–298.

DAV (2022). Deutschen aktuarvereinigung. herleitung der dav-sterbetafel 2008— lebensversicherung. Raucher- und Nichtrauchersterbetafeln für Lebensversicherungen mit Todesfallcharakter.

Deming, W. E. and Stephan, F. F. (1940). *On a least squares adjustment of a sampled frequency table when the expected marginal totals are known*. Springer.

Denecke, E., Grigoriev, P., and Rau, R. (2023). Evaluation of small-area estimation methods for mortality schedules.

Destatis (2025). Statistisches bundesamt (destatis). population by länder. Accessed: 2025-02-05.

Duchemin, L., Veber, P., and Boussau, B. (2022). Bayesian investigation of sars-cov-2-related mortality in france. *Peer Community Journal*, 2.

El Emam, K., Jonker, E., Arbuckle, L., and Malin, B. (2011). A systematic review of re-identification attacks on health data. *PloS one*, 6(12):e28071.

Gmel, H., Kuendig, H., Notari, L., and Gmel, C. (2017). Suchtmonitoring schweiz: Konsum von alkohol, tabak und illegalen drogen in der schweiz im jahr 2016. Technical report, Sucht Schweiz.

Goldstein, J., Osborne, M., Atherwood, S., et al. (2023). Mortality modeling of partially observed cohorts using administrative death records. *Population Research and Policy Review*, 42(36).

Haberman, S. and Renshaw, A. E. (1996). Generalized linear models and actuarial science. *The Statistician*, 45(4):407–436.

HMD (2023). Human Mortality Database, University of California, Berkeley (USA) and Max Planck Institute for Demographic Research (Germany). https://www.mortality.org.

ISTAT (2025). Italian national institute of statistics (istat). resident population. Accessed: 2025-02-05.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. CRC Press.

McEvoy, J., Blaha, M., Rivera, J., Budoff, M., Khan, A., Shaw, L., Berman, D., Raggi, P., Min, J., Rumberger, J., Callister, T., Blumenthal, R., and Nasir, K. (2012). Mortality rates in smokers and nonsmokers in the presence or absence of coronary artery calcification. *JACC Cardiovascular Imaging*, 5(10):1037–1045. Erratum in: JACC Cardiovasc Imaging. 2013 Jun;6(6):747.

Menotti, A., Puddu, P., Lanti, M., Maiani, G., Catasta, G., and Alberti Fidanza, A. (2014). Lifestyle habits and mortality from all and specific causes of death: 40-year follow-up in the italian rural areas of the seven countries study. *The Journal of Nutrition, Health & Aging*, 18(3):314–321.

Nalmpatian, A., Heumann, C., Alkaya, L., and Jackson, W. (2024). Transfer learning for mortality risk: A case study on the united kingdom.

Nusselder, W. J. and Mackenbach, J. P. (1997). Rectangularization of the survival curve in the netherlands: An analysis of underlying causes of death. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 52(3):S145–S154.

OECD (2023). *Mortality and the Provision of Retirement Income*. OECD Publishing, Paris.

Pukelsheim, F. (2014). Biproportional scaling of matrices and the iterative proportional fitting procedure. *Annals of Operations Research*, 215(1):269–283.

RKI (2014). Robert koch institute. mortality and life expectancy. https://www.rki.de/EN/Content/Health_Monitoring/Health_Reporting/GBEDownloadsK/2014_2_mortality_life_expectancy.pdf?__blob=publicationFile.

Robert, C. P. and Casella, G. (2004). *Monte Carlo statistical methods*. Springer.

Semyonov, L., Iarocci, G., Boccia, A., and La Torre, G. (2012). Socioeconomic differences in tobacco smoking in italy: is there an interaction between variables? *ScientificWorldJournal*, 2012:286472.

Zeiher, J., Kuntz, B., and Lange, C. (2017). Smoking among adults in germany. *J Health Monit*, 2(2):57–63.

Contributing Publications

Nalmpatian, A., Heumann, C., and Pilz, S. (2024). Forecasting Mortality Trends: Advanced Techniques and the Impact of COVID-19. *Stats.*

Nalmpatian, A., Heumann, C., Alkaya, L., and Jackson, W. (2026). Local and global mortality experience: A novel hierarchical model for regional mortality risk. *PLoS One* (in press).

Nalmpatian, A., Heumann, C., Alkaya, L., and Jackson, W. (2025). Transfer learning for mortality risk: A case study on the United Kingdom. *PLoS One*, 20(4): e0313378.

Nalmpatian, A., and Heumann, C. (2025). Mortality simulations for insured and general populations. *arXiv:2502.08814v2.*

References

Agresti, A. (2012). *Categorical data analysis*. John Wiley & Sons. 10, 32

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2003). *Hierarchical modeling and analysis for spatial data*. Chapman and Hall/CRC. 19

Barigou, K., Loisel, S., and Salhi, Y. (2020). Parsimonious predictive mortality modeling by regularization and cross-validation with and without covid-type effect. *Risks*, 9(1): 5. 14

Beard, R. E. (1959). Appendix: Note on some mathematical mortality models. In G.E.W. Wolstenholme and M. O'Connor, editors, *Ciba Foundation Symposium - The Lifespan of Animals (Colloquia on Ageing)*, pages 302–311. John Wiley & Sons, Ltd. 6

Beard, R. E. (1971). Some aspects of theories of mortality, cause of death analysis, forecasting and stochastic processes. *Biological aspects of demography*, 999: 57–68. 6

Beckman, R. J., Baggerly, K. A., and McKay, M. D. (1996). Creating synthetic baseline populations. *Transportation Research Part A: Policy and Practice*, 30(6): 415–429. 10

Booth, H., Maindonald, J., and Smith, L. (2002). Applying lee-carter under conditions of variable mortality decline. *Population Studies*, 56(3): 325–336. 4

Brillinger, D. R. (1986). The natural variability of vital rates and associated statistics. *Biometrics*, 42(4): 693–734. 5

Brouhns, N., Denuit, M., and Vermunt, J. K. (2002). A poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics and Economics*, 31(3): 373–393. 4

Cairns, A. J., Blake, D., and Dowd, K. (2006). A two-factor model for stochastic mortality with parameter uncertainty. *Journal of Risk and Insurance*, 73(4): 687–718. 7

Cairns, A. J., Blake, D., Dowd, K., and Coughlan, G. D. (2011a). Mortality density forecasts: An analysis of six stochastic mortality models. *Insurance: Mathematics and Economics*, 48(3): 355–367. 8, 10

Cairns, A. J., Blake, D., Dowd, K., Coughlan, G. D., and Khalaf-Allah, M. (2011b). Bayesian stochastic mortality modelling for two populations. *ASTIN Bulletin: The Journal of the IAA*, 41(1): 29–59. 4, 13

Cameron, A. C. and Trivedi, P. K. (2013). *Regression analysis of count data*. 53. Cambridge university press. 20

Carstensen, B. (2007). Age–period–cohort models for the lexis diagram. *Statistics in medicine*, 26(15): 3018–3045. 15

Chen, Y. and Khaliq, A. Q. (2022). Comparative study of mortality rate prediction using data-driven recurrent neural networks and the lee–carter model. *Big Data and Cognitive Computing*, 6(4): 134. 9

Clements, M. S., Armstrong, B. K., and Moolgavkar, S. H. (2005). Lung cancer rate predictions using generalized additive models. *Biostatistics*, 6(4): 576–589. 13, 15

Congdon, P. (2007). *Bayesian statistical modelling*. John Wiley & Sons. 19

Cover, T. and Hart, P. (1967). Nearest neighbor pattern classification. *IEEE transactions on information theory*, 13(1): 21–27. 27

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2): 187–202. 20

DAV. (2022). Deutschen aktuarvereinigung. herleitung der dav-sterbetafel 2008— lebensversicherung. Raucher- und Nichtrauchersterbetafeln für Lebensversicherungen mit Todesfallcharakter. 33

Delwarde, A., Denuit, M., and Eilers, P. H. (2006). Smoothing the lee-carter and poisson log-bilinear models for mortality forecasting: A penalized likelihood approach. *Statistical Modelling*, 6(1): 29–48. 4

Denecke, E., Grigoriev, P., and Rau, R. (2023). Evaluation of small-area estimation methods for mortality schedules. 31

Deprez, P., Shevchenko, P. V., and Wüthrich, M. V. (2017). Machine learning techniques for mortality modeling. *European Actuarial Journal*, 7(2): 337–352. 8

Dowd, K., Cairns, A. J., and Blake, D. (2020). Cbdx: a workhorse mortality model from the cairns–blake–dowd family. *Annals of Actuarial Science*, 14(2): 445–460. 10

Duchemin, L., Veber, P., and Boussau, B. (2022). Bayesian investigation of sars-cov-2-related mortality in france. *Peer Community Journal*, 2. 31

Engel, K. and Ziegler, S. (2020). Pandora's box. *A report on the human zoonotic disease risk in Southeast Asia with a focus on wildlife markets. Report, WWF*. 4

Fahrmeir, L., Kneib, T., Lang, S., and Marx, B. (2013). *Regression: Models, Methods and Applications*. Springer Science & Business Media. 28

Goldstein, J., Osborne, M., Atherwood, S., et al. (2023). Mortality modeling of partially observed cohorts using administrative death records. *Population Research and Policy Review*, 42(36). 31

Gompertz, B. (1825). On the nature of the function expressive of the law of human mortality. *Philosophical Transactions of the Royal Society of London*, 115: 513–585. 6

Graunt, J. (1662). *Natural and Political Observations Made upon the Bills of Mortality*. Royal Society. 6

Halley, E. (1693). An estimate of the degrees of the mortality of mankind. *Philosophical Transactions of the Royal Society of London*, 17: 596–610. 6

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Additive Models*, volume 43. CRC press. 13

Heligman, L. and Pollard, J. H. (1980). The age pattern of mortality. *Journal of the Institute of Actuaries*, 107(1): 49–80. 6

HMD. (2024). Human mortality database. *University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany)*. Accessed: 2024-06-15. 4, 15, 26

Hoem, J. M. (1983). The reticent trio: some little-known early discoveries in life insurance mathematics by lhf oppermann, tn thiele and jp gram. *International Statistical Review/Revue Internationale de Statistique*, pages 213–221. 6

Hollmann, F. W., Mulder, T. J., and Kallan, J. E. (2000). *Methodology and Assumptions for the Population Projections of the United States: 1999-2100*. Population Projections Branch, Population Division, US Census Bureau 4

Human Mortality Database. (2024). Short-term mortality fluctuations (stmf) data series. `https://www.mortality.org/`. Accessed 2024-06-15. 10

Hyndman, R. J., Booth, H., and Yasmeen, F. (2013). Coherent mortality forecasting: the product-ratio method with functional time series models. *Demography*, 50: 261–283. 4, 8

(IASB), I. A. S. B. (2025). Ifrs 17: Insurance contracts. Official accounting standard for insurance contracts, including guidance on mortality assumptions. 11

of Insurance Supervisors (IAIS), I. A. (2024). Global insurance market report 2024. IAIS Global Monitoring Exercise. Contains the latest evaluation of longevity and mortality risk transfer activity. 11

Ioannidis, J. P. (2021). Over-and under-estimation of covid-19 deaths. *European journal of epidemiology*, 36(6): 581–588. 13

Jarner, S. F. and Kryger, E. S. (2011). Modelling adult mortality in small populations: The saint model. *ASTIN Bulletin*, 41(2): 377–418. 8

Johnson, C. O., DeCleene, N. K., Blacker, B. F., Cunningham, M. W., Aravkin, A., Dieleman, J. L., Gakidou, E., Naghavi, M., Ogunniyi, M. O., Zheng, P., et al. (2023). State-level cardiovascular mortality rates among hispanic, non-hispanic black, and non-hispanic white populations, 1990 to 2019. *JAMA cardiology*, 8(5): 429–442. 31

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., and Liu, T.-Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30. 20

Kleinow, T. (2015). A common age effect model for the mortality of multiple populations. *Insurance: Mathematics and Economics*, 63: 147–152. 13

Laplace, P. S. (1825). *Essai philosophique sur les probabilités*. Bachelier. iii, 4

Lee, R. and Miller, T. (2000). Assessing the performance of the lee-carter approach to modeling and forecasting mortality. In *Annual Meeting of the Population Association of America, Los Angeles*. 7

Lee, R. D. and Carter, L. R. (1992). Modeling and forecasting us mortality. *Journal of the American statistical association*, 87(419): 659–671. 4, 7

Levantesi, S. and Pizzorusso, V. (2019). Application of machine learning to mortality modeling and forecasting. *Risks*, 7(1): 26. 8, 27

Li, N. and Lee, R. D. (2005). Coherent mortality forecasts for a group of populations: An extension of the lee-carter method. *Demography*, 42(3): 575–594. 4, 8, 10, 13, 15

Lim, H. B. and Shyamalkumar, N. (2021). Incorporating industry stylized facts into mortality tables: Transfer learning with monotonicity constraints. *Available at SSRN 3964181*. 25

Lomax, N. and Norman, P. (2016). Estimating population attribute values in a table:"get me started in" iterative proportional fitting. *The Professional Geographer*, 68(3): 451–461. 10

Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30. 23

Makeham, W. (1867). On the law of mortality. *Journal of the Institute of Actuaries*, 13: 325–358. 6

Meier, D. and Wuthrich, M. V. (2020). Convolutional neural network case studies:(1) anomalies in mortality rates (2) image recognition. *Available at SSRN 3656210*. 4

de Moivre, A. (1725). *Annuities upon Lives*. William Pearson. 6

Murray, C. J., Laakso, T., Shibuya, K., Hill, K., and Lopez, A. D. (2007). Can we achieve millennium development goal 4? new analysis of country trends and forecasts of under-5 mortality to 2015. *The lancet*, 370(9592): 1040–1054. 20

Nigri, A., Levantesi, S., Marino, M., Scognamiglio, S., and Perla, F. (2019). A deep learning integrated lee–carter model. *Risks*, 7(1): 33. 9

Nusselder, W. J. and Mackenbach, J. P. (1997). Rectangularization of the survival curve in the netherlands: An analysis of underlying causes of death. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 52(3): S145–S154. 31

Oppermann, L. (1870). On the graduation of life tables, with special application to the rate of mortality in infancy and childhood. *The Insurance Record Minutes from a meeting in the Institute of Actuaries*, 42. 6

Pan, S. J. and Yang, Q. (2009). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10): 1345–1359. 10

Perla, F., Richman, R., Scognamiglio, S., and Wüthrich, M. V. (2021). Time-series forecasting of mortality rates using deep learning. *Scandinavian Actuarial Journal*, 2021(7): 572–598. 4

Perla, F., Richman, R., Scognamiglio, S., and Wüthrich, M. V. (2024). Accurate and explainable mortality forecasting with the localglmnet. *Scandinavian Actuarial Journal*, 2024(7): 739–761. 4

Plat, R. (2009). On stochastic mortality modeling. *Insurance: Mathematics and Economics*, 45(3): 393–404. 8

Postema, J. T. and van Es, R. (2022). Improved mortality rate forecasting using machine learning and open data. White paper, Milliman. https://www.milliman.com/en/insight/Improved-mortality-rate-using-ML-and-open-data. 9

Python Software Foundation. (2025). *The Python Language Reference*. Python Software Foundation, Wilmington, DE, USA. 3

R Core Team. (2025). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. 3

Raftery, A. E., Lalic, N., and Gerland, P. (2014). Joint probabilistic projection of female and male life expectancy. *Demographic research*, 30: 795. 8

Rajendran, S. (2024). The impact of ai on mortality forecasting: Interpretability and fairness. https://www.soa.org/4a5e82/globalassets/assets/files/resources/research-report/2024/impact-ai-mortality/2024-impact-ai-mort-essay-rajendran.pdf. 9

Renshaw, A. E. and Haberman, S. (2003). Lee-carter mortality forecasting: A parallel generalized linear modelling approach for england and wales mortality projections. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 52(1): 119–137. 5

Renshaw, A. E. and Haberman, S. (2006). A cohort-based extension to the lee–carter model for mortality reduction factors. *Insurance: Mathematics and Economics*, 38(3): 556–570. 7, 10

Renshaw, A. E. and Haberman, S. (2008). *Modelling mortality with actuarial applications*. Cambridge University Press. 5

Richman, R. and Wüthrich, M. V. (2021). A neural network extension of the lee–carter model to multiple populations. *Annals of Actuarial Science*, 15(2): 346–366. 4

Richman, R. and Wüthrich, M. V. (2023). Localglmnet: interpretable deep learning for tabular data. *Scandinavian Actuarial Journal*, 2023(1): 71–95. 4, 13

RKI. (2014). Robert koch institute. mortality and life expectancy. https://www.rki.de/EN/Content/Health_Monitoring/Health_Reporting/GBEDownloadsK/2014_2_mortality_life_expectancy.pdf?__blob=publicationFile. 31

Robben, J., Antonio, K., and Devriendt, S. (2022). Assessing the impact of the covid-19 shock on a stochastic multi-population mortality model. *Risks*, 10(2): 26. 10, 14

Roux, A. D. (2002). A glossary for multilevel analysis. *Journal of Epidemiology & Community Health*, 56(8): 588–594. 19

Sánchez-Romero, M., Lee, R., and Fürnkranz-Prskawetz, A. (2019). Redistributive effects of different pension structures when longevity varies by socioeconomic status in a general equilibrium setting. 19

Schnürch, S., Kleinow, T., Korn, R., and Wagner, A. (2022). The impact of mortality shocks on modelling and insurance valuation as exemplified by covid-19. *Annals of Actuarial Science*, 16(3): 498–526. 14

Shi, Y., Ke, G., Soukhavong, D., Lamb, J., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.-Y., Titov, N., and Cortes, D. (2025). *lightgbm: Light Gradient Boosting Machine*. R package version 4.6.0.99. 20

Shiferaw, Y. A. (2021). Regime shifts in the covid-19 case fatality rate dynamics: A markov-switching autoregressive model analysis. *Chaos, Solitons & Fractals: X*, 6: 100059. 10

Siler, W. (1983). Parameters of mortality in human populations with widely varying life spans. *Statistics in medicine*, 23: 451–473. 6

Smith, E., Ismail, Z., et al. (2021). Mortality risk models for persons with dementia: a systematic review. *Journal of Alzheimer's Disease*, 80(1): 103–111. 31

Thatcher, A. R., Kannisto, V., and Vaupel, J. W. (1998). The force of mortality at ages 80 to 120. *Odense University Press*. 6

Thiele, T. N. (1871). On a mathematical formula to express the rate of mortality throughout the whole of life, tested by a series of observations made use of by the danish life insurance company of 1871. *Journal of the Institute of Actuaries*, 16: 313–329. 6

United Nations. (2022). World population prospects 2024. 11

Van Buuren, S. and Van Buuren, S. (2018). *Flexible imputation of missing data*, volume 10. CRC press Boca Raton, FL. 19

Vincelli, M. (2019). A machine learning approach to incorporating industry mortality table features into a company's insured mortality analysis. *Soc Actuar Res Rep*, pages 1–53. 25

Wang, Y. (2024). Ai and mortality: Data-driven health risk forecasting using big data. *Society of Actuaries Research Brief*. 9

Weigert, M., Bauer, A., Gernert, J., Karl, M., Nalmpatian, A., Küchenhoff, H., and Schmude, J. (2021). Semiparametric APC analysis of destination choice patterns: Using generalized additive models to quantify the impact of age, period, and cohort on travel distances. *Tourism Economics*. 14, 15

Weigert, M., Bauer, A., Gernert, J., Karl, M., Nalmpatian, A., Kuechenhoff, H., and Schmude, J. (2022). Semiparametric apc analysis of destination choice patterns: Using generalized additive models to quantify the impact of age, period, and cohort on travel distances. *Tourism Economics*, 28(5): 1377–1400. 14

Weiss, K., Khoshgoftaar, T. M., and Wang, D. (2016). A survey of transfer learning. *Journal of Big data*, 3(1): 9. 10

Woolhouse, M. and Gaunt, E. (2007). Ecological origins of novel human pathogens. *Critical reviews in microbiology*, 33(4): 231–242. 4

World Health Organization. (2024). Ageing and health. WHO Fact Sheet. "By 2050, the world's population of people aged 60 years and older will double (2.1 billion).". 11

Yakovyna, V., Shakhovska, N., and Szpakowska, A. (2024). A novel hybrid supervised and unsupervised hierarchical ensemble for covid-19 cases and mortality prediction. *Scientific Reports*, 14(1): 9782. 19

References

Yan, J., Guszcza, J., Flynn, M., and Wu, C.-S. P. (2009). Applications of the offset in property-casualty predictive modeling. In *Casualty actuarial society e-forum*, volume 1, pages 366–385. Casualty Actuarial Society Arlington. 28

Eidesstattliche Versicherung (Affidavit)

(Siehe Promotionsordnung vom 12. Juli 2011, § 8 Abs. 2 Pkt. 5)

Hiermit erkläre ich an Eides statt, dass die Dissertation von mir selbstständig, ohne unerlaubte Beihilfe angefertigt ist.

München, den 12.11.2025 Asmik Nalmpatian