DEMOGRAPHIC RESEARCH
A peer-reviewed, open-access journal of population sciences

DEMOGRAPHIC RESEARCH

VOLUME 51, ARTICLE 33, PAGES 1017–1058
PUBLISHED 25 OCTOBER 2024
https://www.demographic-research.org/Volumes/Vol51/33
DOI: 10.4054/DemRes.2024.51.33

Research Article

A Bayesian model for age at death with cohort effects

Matteo Dimai

Marek Brabec

Contents

Demographic Research: Volume 51, Article 33
Research Article

A Bayesian model for age at death with cohort effects

Matteo Dimai[1]

Marek Brabec[2]

Abstract

BACKGROUND
Ongoing mortality trends affect the distribution of age at death, typically described by parametric models. Cohort effects can markedly perturb the distribution and reduce the fit of such models, and this needs to be specifically taken into account.

OBJECTIVE
This study examines the integration of cohort effects in a three-component parametric model for the age-at-death distribution, applying it to data with significant cohort effects.

METHODS
We employed a mixture model with a half-normal and two skew-normal components, adapted to a Bayesian framework to include multiplicative cohort effects. The model was applied to data from five Italian regions, with cohort effects estimated for the 1915–1925 cohorts.

RESULTS
Incorporating cohort effects significantly improved the model's fit. A notable finding of the comprehensive model is the shift in Italy from premature to middle-age mortality components over time. Our results also demonstrate the tendency for mortality structures to spatially homogenize over time in Italy.

CONCLUSIONS
The study underscores the importance of including cohort effects in mortality models in order to provide a more detailed picture of mortality trends.

CONTRIBUTION
This work introduces a novel application of a Bayesian mixture model with cohort effects, offering enhanced tools for demographic analysis and new insights into the

[1] Department of Economics, Business, Mathematics and Statistics, University of Trieste, Trieste (TS), Italy. Email: matteo.dimai@phd.units.it.
[2] Institute of Computer Science, The Czech Academy of Sciences, Praha, Czech Republic. Email: mbrabec@cs.cas.cz.

evolution of mortality components in Italy. This approach is general but fully formalized and hence it can be readily used for demographic studies in other regions as well.

1. Introduction

Life expectancy has seen remarkable global increases in the past century, resulting in substantial shifts in the distribution of age at death. Everyone dies and everyone dies once: hence the distribution of ages at death can be seen as a probability distribution, which has historically been proven to be a popular choice for explaining and forecasting mortality. Traditionally, demographers have employed a three-component framework to describe this distribution, comprising infant, premature or young adult, and adult or old-age mortality, leading to a mixture model. However, the age-at-death curve often departs from the smooth patterns expected in a mixture of probability distributions. Notably, cohort effects – persistently elevated or reduced mortality rates experienced by specific birth cohorts throughout their lifetimes – create disruptions that a traditional smooth component mixture model cannot fit. Empirical evidence, such as that observed in the rapid mortality improvements among cohorts born in England and Wales between 1925 and 1945 (Willets 2004), underpins the impact of cohort effects.

These perturbations manifest as clusters of cohorts with distinct mortality rates, introducing discontinuities or even multiple peaks in the age-at-death distribution. Despite being localized, these irregularities can easily complicate parameter estimation and hinder the model accuracy of the global smooth mixture approximation.

This paper contributes to structured mortality modeling in two significant ways. First, it leverages a recently developed three-component parametric model for ages at death (Zanotto, Canudas-Romo, and Mazzuco 2021) and situates it fully within a Bayesian framework, an inherently suitable choice for modeling a probability distribution. Second, driven by a strong empirical feature present in the data, the model is extended to encompass cohort effects, enhancing model fit and producing convincing information about the existence of specific cohort effects. We present the application of this model to several Italian regions over the period 1974–2022.

The rest of this paper is structured as follows. Section 2 provides a review of relevant literature, section 3 introduces the Bayesian three-component mixture model incorporating cohort effects, section 4 presents the main results of the application, and section 5 provides a conclusion and discussion.

2. Literature review

There is considerable variety in the approaches to mortality modeling and forecasting. Models differ in what they are trying to model; for instance, death rates (Lee and Carter 1992), life expectancy (Raftery et al. 2013), or odds ratios (Heligman and Pollard 1980). The classical approach, pioneered by De Moivre in the 18th century and Gompertz in the 19th (1825), is to model a summary of mortality that lends itself to a suitable parametric form, with parameters that can be explained in terms of known demographic phenomena. On the other end of the scale are models that do not require specific functional forms, i.e., the original Lee and Carter (1992) model. There is an unresolved tension between model parsimony and fit, with intermediate approaches, like the procedure advocated by Hunt and Blake (2014) that models mortality with a multiplicity of simple parametric forms. Forecasting techniques can then be applied to the components of the model that vary in time, be they distribution parameters or time terms.

More profound differences exist in the mechanism that drives the future evolution of mortality, with extrapolation, explanation, and expert approaches (Stoeldraijer et al. 2013). Finally, there are differences in the statistical tools used. For example, Lee and Carter (1992) propose likelihood-based bilinear models and Czado, Delwarde, and Denuit (2005) Bayesian bilinear models, while more recently Scognamiglio (2022) and Marino, Levantesi, and Nigri (2023) have suggested neural networks.

A popular choice, with a long history, is to describe mortality parametrically through the distribution of deaths by age, which can be viewed as a probability distribution and can be decomposed into multiple components. Lexis (1879) identified three components to mortality – infant, premature, and adult – in what is arguably the most famous but not the only approach (Barnett 1958; Pearson 1897). In a recent paper, Zanotto, Canudas-Romo, and Mazzuco (2021) modeled the same three components identified by Lexis (1879) using a mixture model with a half-normal and two skew-normal distributions (Azzalini 1985). This model forms the basis of our study and is reviewed in more detail in section 3.2.

Bayesian mortality models usually do not model the whole distribution of ages at death. Girosi and King (2008) and Alexander, Zagheni, and Barbieri (2017) model mortality rates, Raftery et al. (2013) model life expectancy. Yet, considering the age at death as a random variable, the problem of building a stochastic mortality model becomes quite naturally a hierarchical Bayesian model which is easy to generalize in many ways, i.e., with a multipopulation mortality model or by describing the evolution of the parameters in time and/or space as stochastic processes.

Cohort effects – that is, some birth cohorts experiencing higher or lower mortality throughout their lives – are a known feature of demographic modeling. Renshaw and Haberman (2006) introduced the first extension of the Lee–Carter model to cohort effects,

those of English cohorts born between 1925 and 1945 (Willets 2004). The Cairns–Blake–Dowd (CBD) model (Cairns, Blake, and Dowd 2006) has a cohort effect extension as well (Cairns et al. 2009) and the model by Plat (2009) also incorporates cohort effects. The inclusion of cohort effects in a Bayesian setting is more uncommon: Fung, Peters, and Shevchenko (2019) provide an example where a Bayesian state–space approach incorporates them in the model and Hunt and Blake (2021a) offer a Bayesian perspective on the modeling of cohort effects within the context of Age–Period–Cohort (APC) models reviewed in Hunt and Blake (2021b). To the authors' knowledge, this study is the first where cohort effects are included in a Bayesian mixture model of the age at death.

3. Model and data

3.1 Data

The data analyzed throughout this paper consist of the regional life tables published by the Italian statistical institute (ISTAT) for the years 1974–2022. Five regions are analyzed: Lombardy, Lazio, Sicily, Sardinia, and Friuli Venezia Giulia (FVG). The five regions comprise about 39.5% of the Italian population and represent the full range of differences found across Italian regions, with both small and large regions, northern and southern regions, and varying degrees of urbanization, as outlined in Table 1. For brevity, only male mortality was considered, but the model can be applied to female or unisex mortality without modification.

Table 1: Key information about the Italian regions studied

Region	NUTS 1 classification	Population at 1.1.2022	Male life expectancy at birth in 1974	Male life expectancy at birth in 2022	Improvement in life expectancy 1974–2022
Friuli Venezia Giulia	North-East	1.194.647	67.847	80.652	12.805
Lazio	Center	5.714.882	70.602	80.791	10.189
Lombardy	North-West	9.943.004	68.071	81.137	13.066
Sardinia	Islands	1.587.413	71.115	79.579	8.464
Sicily	Islands	4.833.329	71.073	79.441	8.368

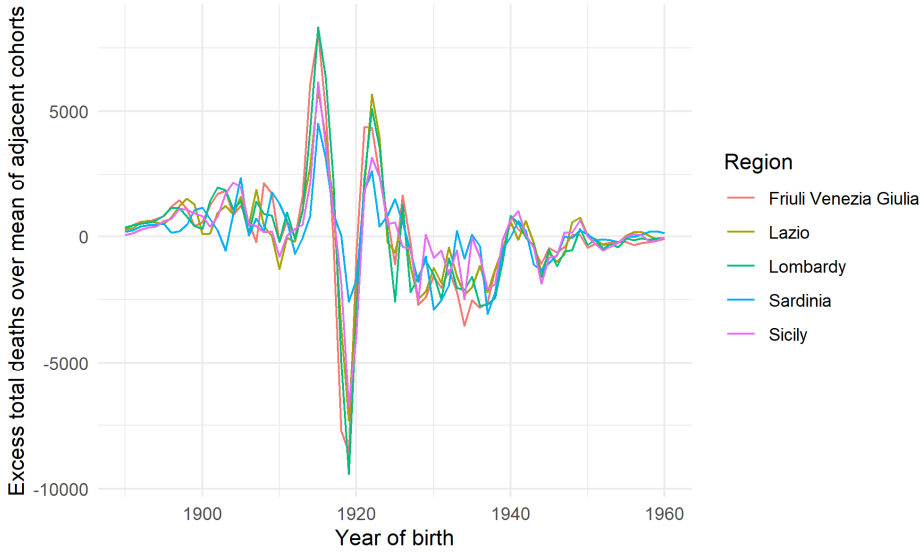
Source: ISTAT

Regional differences within Italy are significant, with southern regions lagging behind in GDP per capita, and Bozzo, Levantesi, and Menzietti (2021) have linked GDP levels to mortality differentials. The differences in life expectancy therefore reflect the differences in development. Demographically, there is a long-standing pattern of migration from the south to the north that was strongest in the 1960s and 1970s, but is

still present today.[3] Additionally, the southern regions used to have the highest fertility in the country, though fertility differentials have greatly reduced since the turn of the millennium. Notably, Sardinia went from having the highest fertility among all regions in the 1950s to the lowest fertility of all in 2022, with a Total Fertility Ratio of 0.95.

The presence of cohort effects and their approximate magnitudes were estimated empirically by calculating the excess deaths for each cohort over all the available life tables compared to the mean of the adjacent cohorts. This preliminary analysis reveals whether mortality patterns are strongly perturbed by cohort effects and suggests which cohorts might be most affected, informing the decision on which ones to include in the model. The results are presented in Figure 1. The data exhibit cohort effects that vary in strength across regions and over time: Lombardy and FVG show the strongest cohort effects, while they are much weaker in Sardinia. The effects are largest, and most similar in magnitude, for the years between approximately 1915 and 1925. Therefore, the impact of cohort effects on measures like life expectancy is largest in years when these cohorts are close to the model age at death – the 1990s and the first part of the 2000s.

Figure 1: Excess life table deaths by cohort and region over the average of the two adjacent cohorts, cohorts 1890–1960, selected Italian regions



[3] Recent data on interregional migration is available on ISTAT's website at https://demo.istat.it/tavole/?t=apr4. Demo.istat.it is the source for all data in this section.

3.2 Model structure

Mortality modeling describes the varying patterns of mortality across different ages x and years t. The indicators modeled can vary, from (age-standardized) life table deaths by age D_x to mortality rates μ_x to death probabilities q_x. This study models life table deaths D_x. That is, we assume that a fictitious population of size l_0 at age 0 is exposed to a series of mortality rates and for each year t we model the number of deaths at each completed age $x = 0, \dots, N$. The model is limited to deaths occurring up to a maximum age $N = 100$, for a total of $D_t = \sum_{t=0}^{100} D_{x,t}$ deaths varying from year to year. The mortality rates used to obtain the life table deaths are the actual death rates for all ages of a real population in a given calendar year, and hence refer to different cohorts.

Since death only happens once, the age at death can be seen as a probability distribution and modeled accordingly. Causes of death vary significantly with age, leading to the usual identification of three components that describe mortality at various ages, each with its separate evolution in time:

1) Infant mortality: high at birth, then rapidly declining. Causes of death due to infant mortality are typically birth defects, pediatric diseases, complications during childbirth, infectious diseases in less developed countries, etc.
2) Premature mortality: mortality due to causes that generally affect adults but are not related to old age: accidental mortality, drug abuse etc.
3) Adult or old-age mortality: mortality due to the effect of bodily decay.

Usually, these components are modeled with continuous distributions, while ages at death are recorded as whole years. Given a vector of parameters θ_t and $P(x;\theta_t)$ representing the probabilities of dying at a given age in year t, life table deaths $D_t \sim Multinomial(\theta_t)$ and the subsequent likelihood are:

$$L(\theta_t; D_{x,t}) = \prod_{x=0}^{N} P(x;\theta_t)^{D_{x,t}}.$$

The probability $P(x;\theta_t)$ can therefore be calculated as:

$$\int_x^{x+1} f(u;\theta_t)du,$$

with $f(x;\theta_t)$ being the underlying continuous mixture of the three aforementioned components.

The focus of the analysis is mortality in a subset of the Italian regions. Italy is a country with one of the highest life expectancies in the world; hence, the adult mortality component is dominant, while infant and premature mortality are less important. The present analysis is based on the eight-parameter model introduced by Zanotto, Canudas-Romo, and Mazzuco (2021) with the frequentist approach, with θ_t being a vector of parameters and $f(x;\theta_t)$ the continuous distribution of the age at death:

$$\theta_t = \{\eta_t, \beta_t, \mu_{m,t}, \sigma_{m,t}, \gamma_{m,t}, \mu_{M,t}, \sigma_{M,t}, \gamma_{M,t}\},$$

$$f(x;\theta_t) = \eta_t \cdot f_I(x) + (1-\eta_t)\left(\beta_t f_m\left(x;\mu_{m,t},\sigma_{m,t},\gamma_{m,t}\right) + (1-\beta_t) f_M\left(x;\mu_{M,t},\sigma_{M,t},\gamma_{M,t}\right)\right).$$

Of the three components, f_I represents infant mortality and is a half-normal with parameters 0 and 1, with the mixing parameter η_t acting as a measure of intensity of infant mortality. We set the scale parameter to 1 to avoid identification issues. Therefore, its density function is:

$$f_I(x) = \frac{\sqrt{2}}{\sqrt{\pi}} exp\left(-\frac{x^2}{2}\right) \quad x > 0.$$

f_m and f_M represent premature and adult mortality respectively, and are both skew-normals (Azzalini 1985). The skew-normal is a class of probability distributions which includes the normal distribution as a special case and has the following density:

$$f(x;\xi,\omega,\lambda) = \frac{2}{\omega}\phi\left(\frac{x-\xi}{\omega}\right)\Phi\left(\lambda\frac{x-\xi}{\omega}\right),$$

where $\phi(\cdot)$ is the standard normal probability distribution function and $\Phi(\cdot)$ is the standard normal cumulative distribution function. The parameters ξ, ω, and λ represent location, scale, and skewness, respectively.

The distributions chosen for the mixture components allow a large variety of age-at-death curves to be fit with relatively few parameters. Infant mortality has a strong maximum close to birth and a steep decrease afterwards; hence, the one-parameter half-normal is parsimonious and offers a good fit. The skew-normal, on the other hand, allows for a more flexible fit for ages where the mortality curve is more complex, such as that from early adulthood to middle age.

As in Zanotto, Canudas-Romo, and Mazzuco (2021), a slightly different parametrization is used in the model, the centered parametrization, with parameters μ (mean), σ (standard error), and γ (skewness), compared to the direct parametrization ξ (location), ω (scale), and λ (shape). The centered parametrization resolves some issues

with the likelihood in the frequentist approach, namely an inflection in the profile likelihood when the skewness parameter is equal to zero, and maps nicely into the mean, standard error, and skewness index of the distribution, allowing for an easier interpretation. The centered parameters (μ, σ, γ) can be transformed into the direct parameters (ξ, ω, λ) with the following equations:

$$c = sgn(\gamma)\left(\frac{2\gamma}{4-\pi}\right)^{\frac{1}{3}}$$

$$\mu_z = \frac{c}{\sqrt{1+c^2}}$$

$$\xi = \mu - \frac{\sigma\mu_z}{\sqrt{1-\mu_z^2}}$$

$$\omega = \frac{\sigma}{\sqrt{1-\mu_z^2}}$$

$$\lambda = \frac{\mu_z\sqrt{\frac{\pi}{2}}}{\sqrt{1-\frac{\pi\mu_z^2}{2}}}$$

Our aim is to produce a descriptive model that minimizes bias; therefore we do not model temporal dynamics of the parameters through stochastic processes. This deliberate choice to stratify the model by calendar year and region helps to reduce bias, a major concern in standard frequentist approaches. By treating each year and region independently, we avoid potential issues deriving from incorrect model assumptions that could arise from stitching different strata together.

However, we acknowledge that this approach comes with the trade-off of increased variability in parameter estimates due to the lack of temporal regularization. In Bayesian modeling, borrowing strength across time periods can indeed lead to better estimates by smoothing out the variations, particularly in smaller populations where actual death rates can exhibit high variability.

Despite this, it is crucial to first establish a robust benchmark model that is free of bias introduced by temporal smoothing. This baseline allows us to later introduce temporal dynamics in a controlled manner and compare the results against the unbiased estimates. As discussed in Section 5, our future work will involve developing a model that incorporates temporal and spatial dynamics, with the current model serving as a necessary foundation. This step-wise approach ensures that any biases introduced by temporal smoothing in future models can be accurately assessed and mitigated by comparison with the benchmark established here.

For both flexibility and computational reasons, we formulate the model in a Bayesian way, assuming a set of priors on the parameters (see Appendix for details). The posterior simulations are then obtained by Hamiltonian Monte Carlo (Homan and Gelman 2014). Stan is used to obtain estimates of the posterior distributions. The model by Zanotto, Canudas-Romo, and Mazzuco (2021) is ported to Stan with minor adaptations. The two mixing parameters η and β are reparametrized with a three-dimensional mixing parameter ζ, distributed as a Dirichlet variable, with no loss of generality. Additionally, the dispersion parameter for infant mortality is parametrized with σ_I, to be estimated instead of setting $\sigma_I = 1$, as preliminary results suggested that 1 was not a value providing adequate fit and that σ_I could be estimated without significant identification issues. The main contribution of the model is the way it handles cohort effects, as described in section 3.2.1, and, to a lesser degree, the simplifications that speed up the estimation phase with a negligible impact on parameter estimates.

In order to speed up the estimation phase, some approximations are used. Calculating actual probabilities θ_t of death at a given age from the three-component model slows down estimation considerably. To address this, we investigated methods to expedite model estimation while maintaining the accuracy of the estimates. We modeled deaths $D_{x,t}$ at age x as a continuous variable, assuming that deaths at age x have happened at age $x + 0.5$ (0.1 at age 0), approximating $P(x;\theta_t)$ with $f(x + 0.5;\theta_t)$. This approximation led to a marked increase in estimation speed, with minimal impact on the accuracy of the parameter estimates. The model is restricted to ages between 0 and 100, while the mortality components are not; therefore normalization of the premature and adult mortality components is necessary. Each of them is normalized by dividing it by the probability of assuming a value between 0 and 100. That is, $f_m(x;\theta_t)$ is actually replaced with $f_{m*}(x;\theta_t) = \frac{f_m(x+0.5;\theta_t)}{F_m(100;\theta_t)-F_m(0;\theta_t)}$, with F being the cumulative distribution function of f, and the same applies to f_M and f_{M*}. The infant mortality component is defined for ages 0 and above and is essentially zero well before the maximum age; hence no normalization is necessary.

3.3 Cohort effects

The Italian data exhibit cohort effects: that is, some cohorts exhibit a higher or lower mortality throughout their lives (Carfora, Cutillo, and Orlando 2017). Empirically, comparing the death of a cohort with the deaths of the adjacent cohorts and calculating excess (or missing) deaths against the two adjacent cohorts, the cohorts most affected by these effects are those born during World War I and in the years shortly afterwards. The full model that reflects these effects, as well as the temporal evolution of the parameters, is as follows:

$$P(x;\theta_t,t,\alpha) = \frac{\alpha_{t-x} f(x;\theta_t)}{\sum_{x=0}^{100} \alpha_{t-x} f(x;\theta_t)},$$

with x denoting age, t the year of the life table, $t-x$ the cohort (year of birth), α_{t-x} the multiplier that represents the cohort effect for cohort $t-x$,

$$f(x;\theta_t) = \zeta_{1,t} \cdot f_I\left(x;\sigma_{I,t}\right) + \zeta_{2,t} \cdot f_{m*}\left(x;\mu_{m,t},\sigma_{m,t},\gamma_{m,t}\right) + \zeta_{3,t} \cdot f_{M*}\left(x;\mu_{M,t},\sigma_{M,t},\gamma_{M,t}\right)$$

being the mixture adapted from Zanotto, Canudas-Romo, and Mazzuco (2021) and where $\zeta \sim Dirichlet(3)$ is a mixing parameter with 2 degrees of freedom, f_I is a half-normal with mean 0 and variance σ_I, and f_m and f_M are skew-normals. $P(x;\theta,t,\alpha)$ is therefore the probability of an individual born in the year $t-x$ dying at age x in year t given their cohort effect α_{t-x} and the probability of dying in year t θ_t. The cohorts for which α_{t-x} can deviate from 1 are chosen by the researcher, with α_{t-x} set to 1 for the other cohorts.

The mixture model is formulated for a single calendar year t; hence to fit 1974–2022 we stratify the estimation on the year. The cohort terms α_{t-x} are therefore the only common terms across multiple years. Theoretically, each α_{t-x} term could be any non-negative number, although in practice it is reasonable to expect them to be close to 1. In order to regularize estimates while allowing the α_{t-x} parameters wide flexibility, the distributions of the α_{t-x} parameters depend on a common parameter α_μ, further discussed in the Appendix.

Modeling cohort effects as multipliers that differ by cohort, but are constant for each cohort at all ages, is a typical choice in Age–Period–Cohort (APC) models, a large family recently reviewed by Hunt and Blake (2021b). The main issue in those models is collinearity, since cohorts (years of birth) are equal to periods (years) minus age. As discussed by Hunt and Blake (2020), in APC models this means that linear trends in the parameters cannot be uniquely split between cohort, age, and period effects. Mixture models are not part of the APC class of models, even though there are some similarities. The mixture models the probability of dying at a given age for a person of age 0, as do some APC models, and with a logarithmic transformation the multiplicative cohort term becomes an additive term, just as in APC models. However, there are no proper age–period terms, since the underlying mixture does not lend itself to being represented as a sum of bilinear age–period terms. Consequently, the relationships between model parameters in mixture models are more complex, and collinearity does not result in quite the same issues as in APC models. However, it is conceivable that trends that affect both age and period can also conflict with the estimation of cohort effect multipliers in mixture models.

In both our mixture model and the APC models there is an assumption of independence between cohort effects and age–period effects: an assumption which is, as

stated by Hunt and Blake (2020), both practical and parsimonious. In order to select the cohort effects to model, the models were run without cohort effects (that is, with all α_{t-x} set to 1), the presence of cohort effects was established through residual graphs, as advocated in Hunt and Blake (2014), and an empirical estimate was produced, as shown in Section 3.1, to select the cohorts with the most prominent effects.

4. Results

The following results were obtained by sampling with 4 chains, 10,000 iterations per chain with a burn-in of 5,000 iterations, for a total of 20,000 sampling iterations, unless otherwise specified. Cohort effects were estimated for the 1915–1925 cohorts, where they are most prominent.

4.1 Parameter estimates

The parameter estimates for the three-component mixture model are shown in Table 2. The proportion of deaths due to infant mortality, ζ_1, declines by an order of magnitude in all regions over the period. On the other hand, premature mortality undergoes a transformation: in 1974 it is a proper premature mortality component, with average age at death μ_m ranging between 18.03 (Lombardia) and 34.24 (Sardinia), while in 2022 it is more of a middle-age mortality component, ranging between 55.07 (Sardinia) and 74.29 (Lombardy). Tables with the 10th and 90th percentiles of the parameter estimates are in the Appendix.

The adult mortality component evolves as expected, with an increase in average age at death μ_M over the period (shift) and a decrease in variability σ_M (compression). The component is strongly negatively skewed and in all regions the skewness parameter γ_M is more extreme than in 1974.

Table 2: Median parameter estimates by region: median, 1974 and 2022 estimates

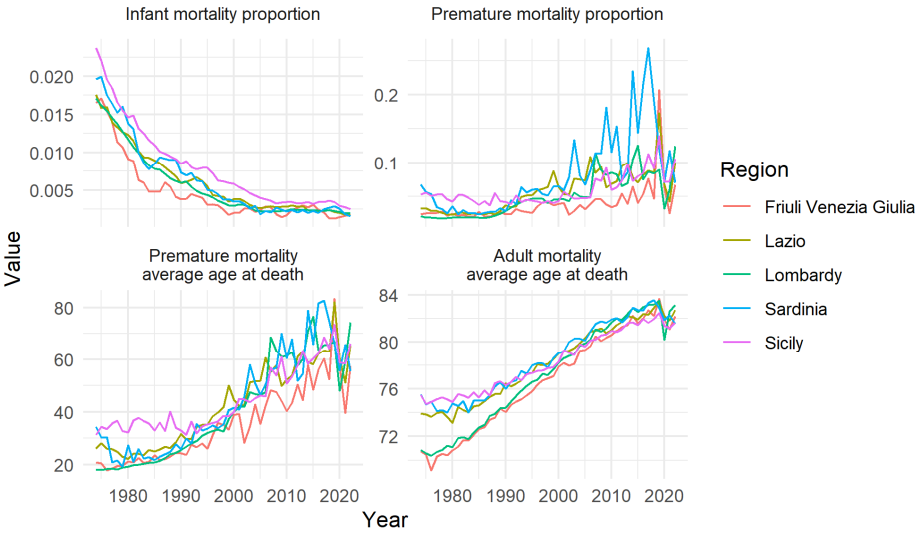
Region	Estimate	ζ_1	ζ_2	ζ_3	σ_I	μ_m	σ_m	γ_m	μ_M	σ_M	γ_M
Friuli Venezia Giulia	1974	0.017	0.02	0.96	0.54	20.87	11.56	0.02	70.74	13.74	–0.75
	2022	0.002	0.07	0.93	0.63	57.62	21.88	0.03	82.21	10.13	–0.80
	Median	0.003	0.03	0.96	0.63	33.22	15.06	0.04	76.88	11.90	–0.73
Lazio	1974	0.018	0.03	0.95	0.54	26.05	18.26	0.09	73.94	11.93	–0.66
	2022	0.002	0.10	0.90	0.64	66.18	22.95	0.02	82.71	9.82	–0.78
	Median	0.004	0.05	0.94	0.58	41.83	18.58	0.05	78.11	11.09	–0.68
Lombardy	1974	0.017	0.02	0.96	0.56	18.10	12.19	0.03	70.88	12.73	–0.72
	2022	0.002	0.12	0.87	0.65	74.41	25.54	0.02	83.14	9.26	–0.78
	Median	0.003	0.04	0.95	0.62	33.34	13.39	0.02	77.28	11.47	–0.68
Sardinia	1974	0.020	0.07	0.91	0.60	34.31	27.83	0.11	75.54	12.36	–0.73
	2022	0.002	0.07	0.93	0.63	55.44	21.46	0.04	81.55	10.77	–0.81
	Median	0.004	0.06	0.93	0.60	35.46	18.40	0.03	78.17	11.71	–0.72
Sicily	1974	0.024	0.05	0.92	0.57	31.35	31.14	0.11	75.50	11.96	–0.71
	2022	0.003	0.11	0.89	0.59	66.33	24.24	0.02	81.64	9.84	–0.77
	Median	0.006	0.05	0.94	0.58	38.52	22.12	0.07	77.78	11.13	–0.71

Figure 2 shows the temporal evolution of a few selected parameters, namely infant mortality proportion ζ_1, premature mortality proportion ζ_2, premature mortality average age at death μ_m, and adult mortality average age at death μ_M. All the major demographic trends are clearly visible. Infant mortality declines from around 2% of all deaths to about 0.2%. Adult mortality average age at death μ_M steadily increases, with a sharp drop in 2020 due to the COVID-19 pandemic, especially apparent in Lombardy, with a subsequent partial recovery. Also, μ_M estimates for the different regions converge.

But the most interesting result is the evolution of premature mortality. Its importance increases over the years, with substantial variability in ζ_2 in all regions and most notably in Sardinia. On the other hand, the average age at death μ_m also increases steadily, with values commonly over 50 years and rather 'wiggly' estimates in the later years. Therefore, we argue that premature mortality transitions from describing the mortality hump between 20 and 40 years of age to a 'middle-age' mortality component. Zanotto, Canudas-Romo, and Mazzuco (2021) report similar results for France, Sweden, and East Germany.

Throughout the years the mortality hump does not disappear: an increase in mortality between ages 10–15 and 20–24 is still clearly visible from the life tables. However, its importance as a fraction of all deaths diminishes to the point where the premature mortality component better fits a larger span of ages, as evidenced by the corresponding increase in the variability parameter σ_m in Friuli Venezia Giulia, Lazio, and Lombardy, the three regions where μ_M was under 30 years in 1974.

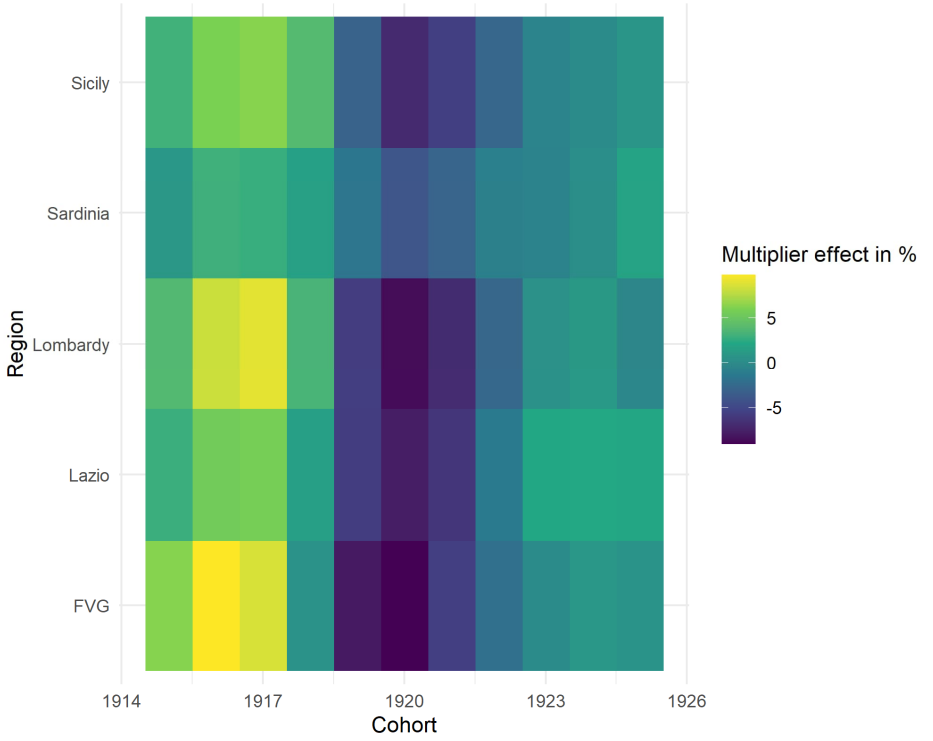
Figure 2: Parameter estimates by calendar year and region: infant mortality proportion, premature mortality proportion, premature mortality average age at death, adult mortality average age at death, 1974–2022



4.2 Cohort effects

Figure 3 shows the magnitude of the estimated cohort effect by region and cohort. Cohorts born in 1916 and 1917 experienced higher mortality throughout their lives, while the cohorts 1919–1921, and especially the 1920 cohort, had substantially lower mortality. The magnitude of the effects varies by region, with the strongest cohort effects estimated for Friuli Venezia Giulia and Lombardy, while cohort effects in Sardinia are the weakest.

Figure 3: Multiplicative cohort effect by region and cohort, 1915–1925 cohorts



The cohort effects are tabulated in Table 3. Eighty percent uncertainty intervals have in general a width between 0.007 and 0.008 that is rather consistent across regions and birth cohorts.

Table 3: Cohort effects by region and cohort: 10th, 50th (median), and 90th percentile

Region	Perc.	1915	1916	1917	1918	1919	1920	1921	1922	1923	1924	1925
Friuli Venezia Giulia	10	1.059	1.094	1.082	1.002	0.917	0.906	0.940	0.975	0.997	1.006	1.003
	50	1.063	1.098	1.086	1.005	0.920	0.909	0.943	0.979	1.000	1.010	1.006
	90	1.067	1.102	1.089	1.009	0.924	0.913	0.947	0.982	1.004	1.013	1.010
Lazio	10	1.025	1.051	1.054	1.012	0.939	0.919	0.934	0.983	1.018	1.019	1.017
	50	1.029	1.055	1.057	1.016	0.942	0.922	0.938	0.987	1.022	1.023	1.021
	90	1.032	1.059	1.061	1.020	0.946	0.926	0.941	0.990	1.026	1.026	1.025
Lombardy	10	1.034	1.078	1.086	1.030	0.940	0.910	0.928	0.970	1.001	1.007	0.993
	50	1.038	1.082	1.090	1.034	0.943	0.914	0.931	0.973	1.005	1.011	0.997
	90	1.042	1.086	1.094	1.038	0.946	0.917	0.934	0.977	1.008	1.015	1.000
Sardinia	10	1.006	1.026	1.024	1.013	0.980	0.956	0.967	0.987	0.990	0.998	1.014
	50	1.010	1.030	1.028	1.017	0.984	0.960	0.971	0.991	0.993	1.002	1.018
	90	1.014	1.034	1.032	1.021	0.987	0.964	0.975	0.994	0.997	1.006	1.022
Sicily	10	1.027	1.056	1.059	1.035	0.966	0.926	0.940	0.968	0.991	0.996	1.004
	50	1.031	1.060	1.063	1.039	0.969	0.930	0.944	0.972	0.995	1.000	1.008
	90	1.035	1.064	1.067	1.043	0.973	0.933	0.947	0.976	0.998	1.004	1.012

4.3 Goodness of fit

The inclusion of cohort effects in the model allows us to fit, in a satisfactory way, complex, even multimodal, distributions of age at death, as Figure 4 can attest. Less populous regions like Sardinia and Friuli Venezia Giulia exhibit substantial variation in age at death, which makes modeling especially challenging. The absence of a pronounced cohort effect in Sardinia is not due to the distribution of ages at death being regular, but due to the inconsistent variation around the main three-component curve.

The improvement due to the inclusion of cohort effects is also noticeable with a traditional metric such as the Mean Squared Error (MSE), as shown in Figure 5. The improvement is especially noticeable for the years when the 1915–1925 cohorts are close to the modal age at death and the impact of the cohort effects is the greatest, while it is reduced essentially to zero in the last years, when the affected cohorts have mostly died off.

Figure 4: Actual (black) and fitted (blue) life table deaths, with 95% confidence interval (red), by age and region, selected years

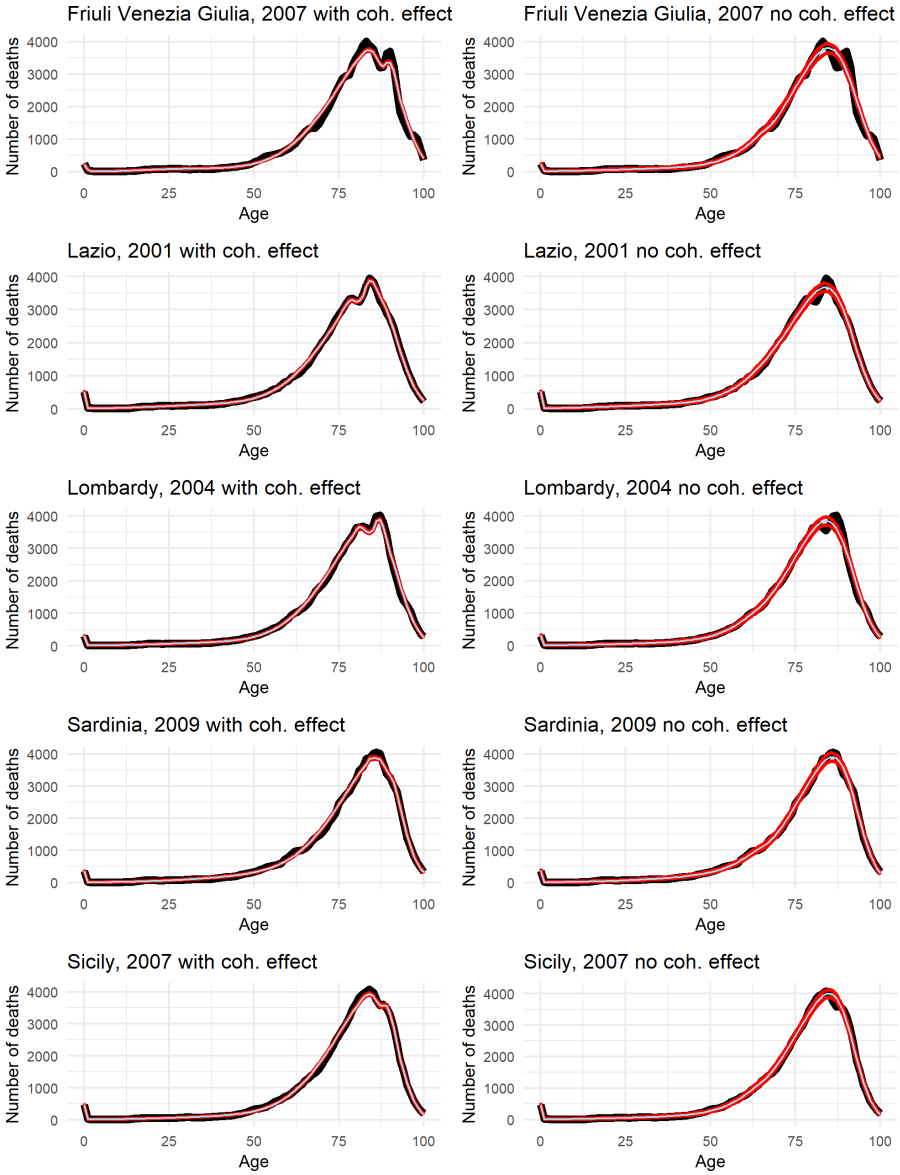
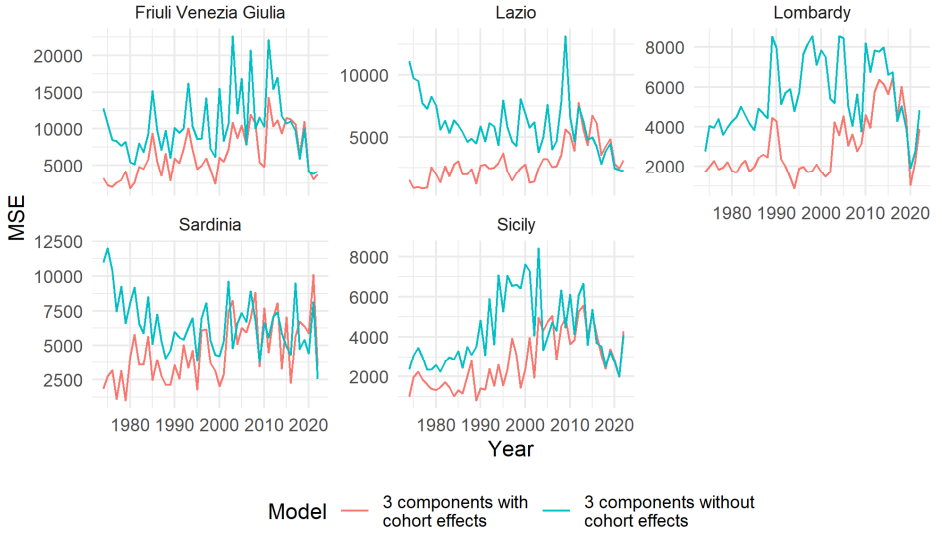
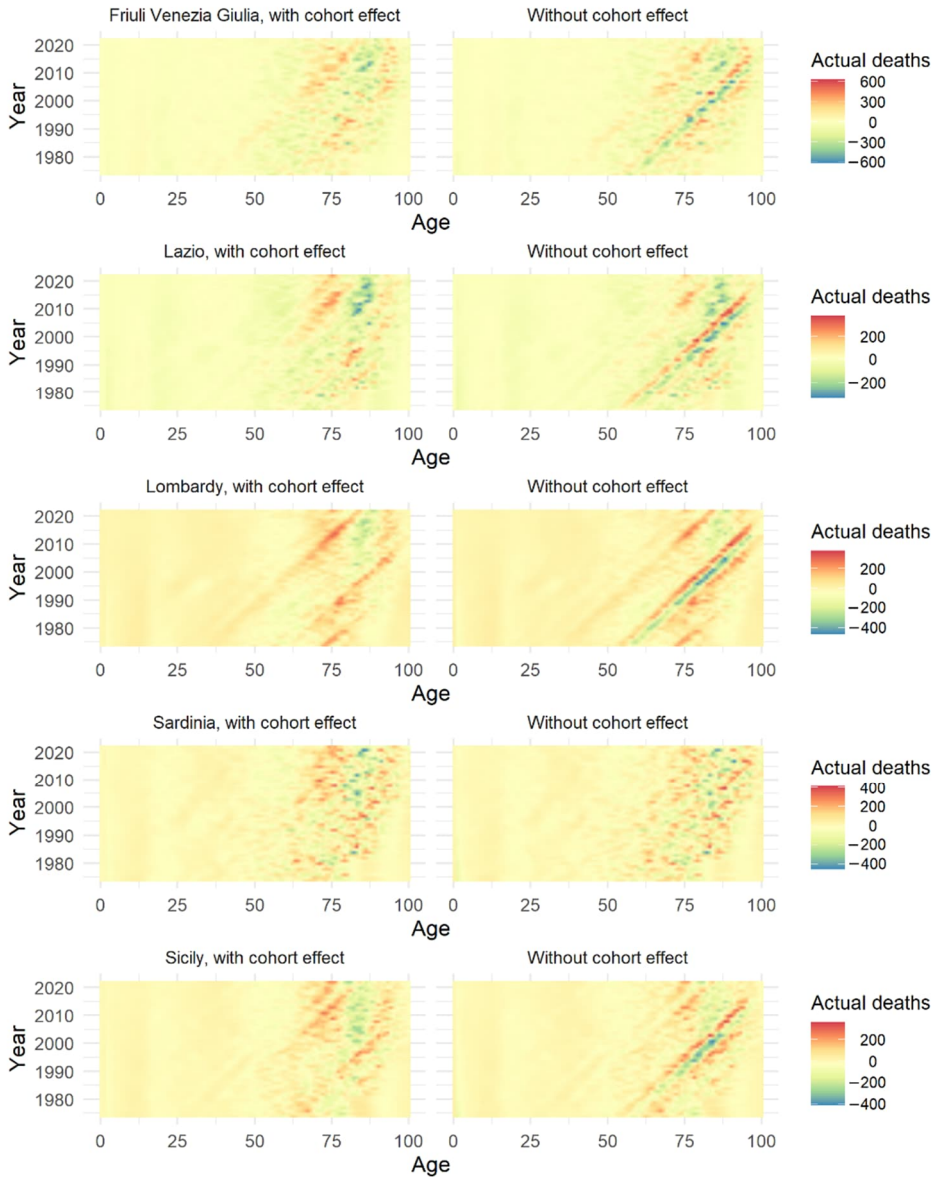


Figure 5: Mean squared error by year and model



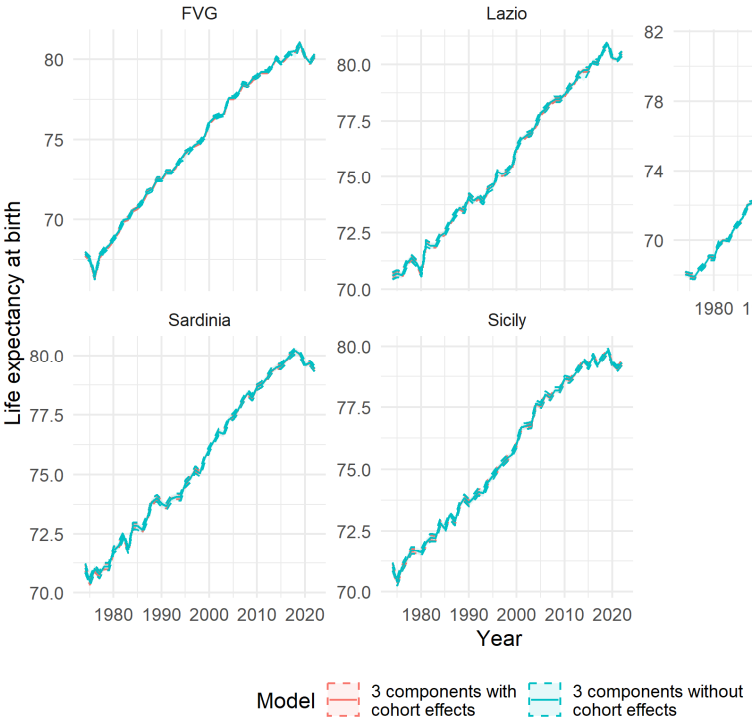
The better fit of the model with cohort effects is also evident when looking at the residuals, as shown in Figure 6. The effects of the 1915–1925 cohort are distinctly visible in the graphs on the right, which do not incorporate cohort terms, whereas in the graphs on the left, where these terms are included, their presence is minimized. Cohorts not included in the plots show smaller effects, with weaker diagonal lines visible (especially for Lazio, Lombardy, and Sicily). As the colors in the diagonal lines show, the residuals in the affected cohorts follow the distribution of the age at death, with the largest residuals around the modal age at death, supporting the choice to model cohort effects as multipliers. Taken all together, the presence of the cohort effects in Italian regional data is firmly established.

Figure 6: Model residuals by age and year, model with cohort effects (left) vs. model without cohort effects (right)



While the impact of the introduction of cohort effects on residuals is significant, summaries like life expectancy at birth are less affected. As Figure 7 shows, the estimated life expectancy at birth is rather similar both with and without cohort effects: however, it should be noted that the prediction intervals are always tighter when cohort effects are included, sometimes substantially so, like for Lazio in 2013 and FVG in 2018.

Figure 7: Estimated life expectancy at birth by year, model with and without cohort effects, median and 95% prediction intervals, 1974–2022



The inclusion of cohort effects trades better fit for increased model complexity. An appropriate way to assess whether the trade-off is worth it is to compare models using a predictive information criterion. We use the Widely Applicable Information Criterion (Watanabe 2010), where the log pointwise predictive density is counterbalanced by its variance to penalize overly complex models. As indicated by Table 4, the model's WAIC

score is lower when cohort effects are incorporated, demonstrating that the augmented model complexity is outweighed by the improvements in model fit.

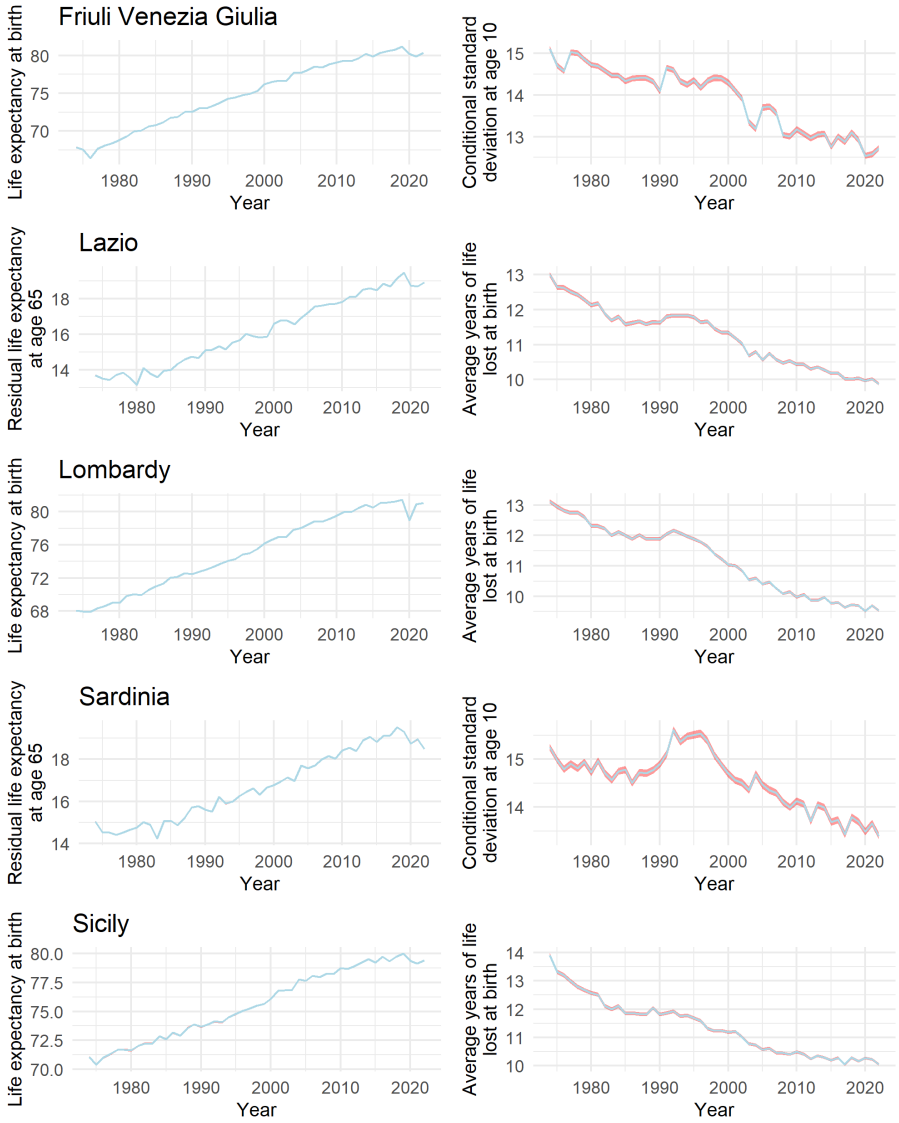
Table 4: WAIC and its standard error by region, 3-component mixture model with and without cohort effects

Region	WAIC, cohort effects	WAIC s.e., cohort effects	WAIC, no cohort effects	WAIC s.e., no cohort effects
Friuli Venezia Giulia	38,503,454	141,392.8	38,515,221	142,631.2
Lazio	38,180,492	98,147.1	38,196,890	99,749.5
Lombardy	38,226,800	138,742.1	38,240,171	140,094.8
Sardinia	38,687,094	109,528.1	38,705,649	109,569.8
Sicily	38,173,785	88,089.7	38,195,678	90,342.2

4.4 Derived quantities

The Bayesian approach enables straightforward computation and visualization of the distribution of various indicators over time, including rather complicated nonlinear characteristics. Following Debón et al. (2017) and Bohk-Ewald, Ebeling, and Rau (2017), we select a set of indicators in order to characterize both average lifespan and lifespan disparity, namely life expectancy at birth e_0, residual life expectancy at age 65 e_{65}, conditional standard deviation at age 10 S_{10} (Edwards and Tuljapurkar 2005), and average number of years lost at birth $e_0^\dagger$ (Vaupel and Canudas-Romo 2003). The results are presented in Figure 8: dispersion measures appear to have a more variable distribution than life expectancy measures.

Figure 8: **Distribution of various indicators by year and region, selected pairs, median (blue) and 90% uncertainty intervals (red): life expectancy at birth, residual life expectancy at age 65, conditional standard deviation at age 10, average years of life lost at birth, 1974–2022**



4.5 Graphic representation of mortality convergence or spatial homogenization

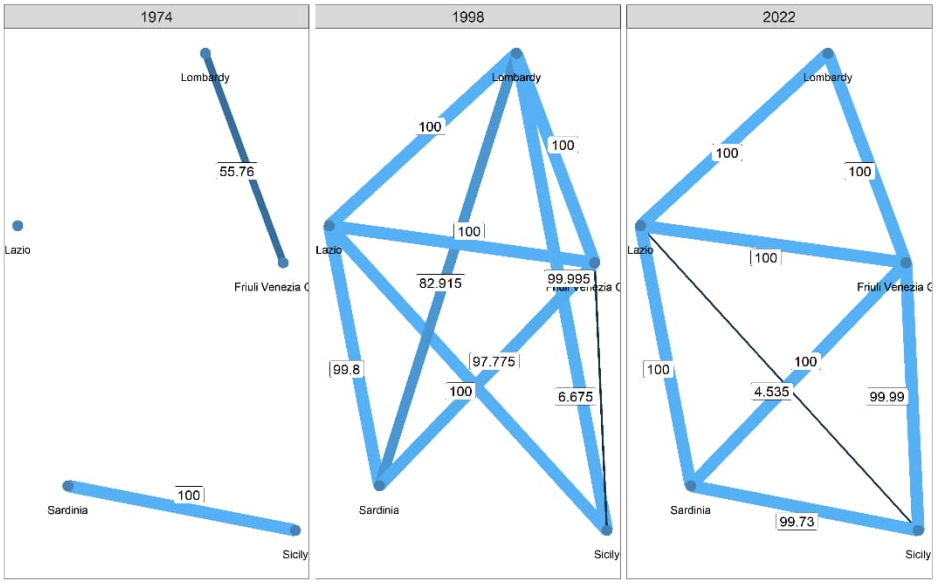
Regions in a country share many common characteristics and, as shown in Section 4.1, parameter estimates of the different regions converge over time. Considering the age at death as a random variable allows us to compare the distribution across regions in various years with the Kullback-Leibler divergence, which is as follows, given two discrete probability distributions P and Q:

$$D_{KL}(P \parallel Q) = \sum_{x \in X} P(x) log\left(\frac{P(x)}{Q(x)}\right)$$

The Bayesian approach employed allows us to calculate the average of the two Kullback-Leibler divergences ($D_{KL}(P \parallel Q)$ and $D_{KL}(Q \parallel P)$) for each of the 20,000 simulations. In this way, for each pair of regions we obtain the distribution of the similarity between their mortality for each year. We can then represent regions in a given year as a network, with the strength of the edge between two regions being equal to the proportion of simulations closer to each other than a set threshold. This set of relationships can be represented as a weighted graph, as shown in Figure 9. The threshold was set to 0.009 in order to emphasize the differences between the three graphs.

The convergence of mortality between regions is particularly strong in the first half of the years observed, with the average degree of the graph nodes increasing from 0.8 in 1974 to 3.6 in 1998. In 1974 the relationships are expressed across strictly geographical lines: the only pairs with any simulation below the threshold are the two northern regions, FVG and Lombardy, and the two southern regions, Sicily and Sardinia. The latter graph edge is conspicuously absent in 1998, only to return in 2022, when Lombardy stops being connected with Sicily and Sardinia and the average degree drops to 3.2.

Figure 9: Weighted network graph, with each edge representing the percentage of simulations with average Kullback–Leibler distance lower than the 0.009 threshold, years 1974, 1998, and 2022



5. Discussion

A common goal in mortality modeling is to simplify summaries of mortality, in this case life table deaths, with a few notable figures. The above analysis shows that the three-component parametric model introduced by Zanotto, Canudas-Romo, and Mazzuco (2021) can be successfully applied in a Bayesian setting, showing a satisfactory fit. The three components and the parameters in general have a clear demographic interpretation and are easy to understand. The ability to meaningfully estimate a three-component model in a setting where two components are small but not negligible is a significant methodological challenge and one that, as long as mortality continues to improve, will be more relevant in the future.

Italian regional mortality data is significantly perturbed by cohort effects. The effects are present for the cohorts born from World War I up to World War II, but are most prominent for the cohorts born in 1915–1925. The magnitude of the effects varies between regions, highlighting the need for a solution that is flexible enough to fit regions

with strong effects like Lombardy and Friuli Venezia Giulia and regions with little to no effects like Sardinia. Our model formulation is satisfactory in this regard and substantially improves the model fit, as confirmed both graphically and in terms of WAIC in Section 4.3. While our approach represents, to our knowledge, the first implementation of cohort effects in a Bayesian mixture model for age at death, further work is needed to enhance the model's capability to estimate all cohort effects up to 1945 comprehensively and completely address the issue of cohort effects for all relevant cohorts, not just the most affected ones. For that future work, cohorts will need to be modeled together in order to reduce the risk of overparametrization.

During the period considered, the premature mortality component, which was clearly identifiable in the 1970s, becomes progressively closer to adult mortality, with an increase of its average age and an overall reduction in skewness. This is consistent with some of Zanotto, Canudas-Romo, and Mazzuco's results (2021), with the premature mortality component becoming a sort of 'middle-age mortality' component. Skewness has to be constrained to non-negative values to avoid this component confounding itself with adult mortality, impacting model convergence. In general, the choice of initial values for the parameters has non-negligible effects on posterior estimates; therefore plausible values need to be provided in order to guarantee convergence.

In terms of mortality indicators, the well-known shift of mortality to later ages and its compression are both confirmed. Life expectancy both at birth (e_0) and at 65 years (e_{65}) trends up throughout the period, with a strong negative effect of the Covid-19 pandemic, while dispersion indicators trend down, save for a brief period in the mid-1990s. In this regard, the advantage of the Bayesian approach is that it allows an easy quantification of the distribution of these quantities: the estimates of dispersion indicators appear substantially more uncertain than life expectancy.

The lack of relationships between model parameters from different years, with the exception of cohort effects, is both a deliberate choice and a limitation. Stratification on the year, i.e., estimating the parameters for each year separately, reduces bias at the expense of increasing the variability in parameter estimates. The model in the present study is the first block towards building a forecasting model with temporal and spatial dynamics, and should be viewed as a benchmark model that minimizes bias.

We do acknowledge that without any temporal regularization MCMC chains are not well behaved in all years, with rising variability in estimates of premature mortality parameters. Premature mortality parameters start behaving differently around the year 2000. Why is that? The estimates point to an increase in the relative importance of middle-age mortality, and much-discussed themes like lifespan inequality spring to mind. But in order to explore this change in behavior it is important to have a clear picture of the temporal evolution of the parameters. At the turn of the century the cohorts most affected by cohort effects approach the modal age at death, and therefore cohort effects

have the maximum impact in absolute terms. It is therefore important to use a foundational model, such as the one presented, to remove the cohort effects in order to have a clearer picture of the temporal dynamics before modeling them. Modeling the parameters as a temporal process would regularize the estimates and possibly improve effective sample size. This is a milestone to be achieved in the future, a model that will have to be compared to the one in the present study to ascertain its bias. It is conceivable that a model with temporal dynamics will have estimates with lower variability and achieve good fit in most years, with severe bias in years where the temporal dynamics do not fit as well – 2020 and the disparate impact of the pandemic on Italian regions springs to mind. In order to assess that bias, a model like the present one, stratified on each year, is needed.

Our model uses mildly informative priors, and some restrictions on parameters and initial values in order to achieve convergent chains and stable estimates. The restrictions are imposed in order to induce the model components to represent the known demographic phenomena that they are meant to represent, not other specific features of the age-at-death curve. However, the need for these restrictions suggests that especially in the 21st century, when the chain convergence is more difficult, the age-at-death curve is becoming more complex and may need a more complex model to be represented adequately. This is a promising path for future research.

Spatial and temporal dynamics are not the only possible development. Location parameters like μ_m and μ_M could be regressed on an external variable such as regional GDP, building on Bozzo, Levantesi, and Menzietti (2021), who find a relationship between the level of GDP by region and the k_t time index in a Lee–Carter model, or on additional external variables as in Dimai (2024).

References

Alexander, M., Zagheni, E., and Barbieri, M. (2017). A flexible Bayesian model for estimating subnational mortality. *Demography* 54(6): 2025–2041. doi:10.1007/s13524-017-0618-7.

Aliverti, E., Mazzuco, S., and Scarpa, B. (2022). Dynamic modelling of mortality via mixtures of skewed distribution functions. *Journal of the Royal Statistical Society Series A* 185(3): 1030–1048. doi:10.1111/rssa.12808.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics.* 12(2): 171–178.

Barnett, H. (1958). Experiments in mortality graduation and projection using a modification of Thiele's formula. *Journal of the Institute of Actuaries* 84(2): 212–229. doi:10.1017/S0020268100037550.

Bohk-Ewald, C., Ebeling, M., and Rau, R. (2017). Lifespan disparity as an additional indicator for evaluating mortality forecasts. *Demography* 54(4): 1559–1577. doi:10.1007/s13524-017-0584-0.

Bozzo, G., Levantesi, S., and Menzietti, M. (2021). Longevity risk and economic growth in sub-populations: Evidence from Italy. *Decisions in Economics and Finance* 44(1): 101–115. doi:10.1007/s10203-020-00275-x.

Cairns, A.J.G., Blake, D., and Dowd, K. (2006). A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration. *Journal of Risk and Insurance* 73(4): 687–718. doi:10.1111/j.1539-6975.2006.00195.x.

Cairns, A.J.G., Blake, D., Dowd, K., Coughlan, G.D., Epstein, D., Ong, A., and Balevich, I. (2009). A quantitative comparison of stochastic mortality models using data from England and Wales and the United States. *North American Actuarial Journal* 13(1): 1–35. doi:10.1080/10920277.2009.10597538.

Carfora, M.F., Cutillo, L., and Orlando, A. (2017). A quantitative comparison of stochastic mortality models on Italian population data. *Computational Statistics and Data Analysis* 112: 198–214. doi:10.1016/j.csda.2017.03.012.

Czado, C., Delwarde, A., and Denuit, M. (2005). Bayesian Poisson log-bilinear mortality projections. *Insurance: Mathematics and Economics* 36(3): 260–284. doi:10.1016/j.insmatheco.2005.01.001.

Debón, A., Chaves, L., Haberman, S., and Villa, F. (2017). Characterization of between-group inequality of longevity in European Union countries. *Insurance: Mathematics and Economics* 75: 151–165. doi:10.1016/j.insmatheco.2017.05.005.

Dimai, M. (2024). Modeling and forecasting mortality with economic, environmental and lifestyle variables. *Decisions in Economics and Finance*. doi:10.1007/s10203-024-00434-4.

Edwards, R.D. and Tuljapurkar, S. (2005). Inequality in life spans and a new perspective on mortality convergence across industrialized countries. *Population and Development Review* 31(4): 645–674. doi:10.1111/j.1728-4457.2005.00092.x.

Fung, M.C., Peters, G.W., and Shevchenko, P.V. (2019). Cohort effects in mortality modelling: A Bayesian state-space approach. *Annals of Actuarial Science* 13(1): 109–144. doi:10.1017/S1748499518000131.

Gelman, A., Jakulin, A., Pittau, M.G., and Su, Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics* 2(4): 1360–1383. doi:10.1214/08-AOAS191.

Girosi, F. and King, G. (2008). *Demographic forecasting*. Princeton, NJ: Princeton University Press.

Gompertz, B. (1825). XXIV. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. In a letter to Francis Baily, Esq. FRS &c. *Philosophical Transactions of the Royal Society of London* (115): 513–583. https://doi.org/10.1098/rstl.1825.0026.

Heligman, L. and Pollard, J.H. (1980). The age pattern of mortality. *Journal of the Institute of Actuaries* 107(1): 49–80. doi:10.1017/S0020268100040257.

Homan, M.D. and Gelman, A. (2014). The No-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research* 15(1): 1593–1623. doi:10.5555/2627435.2638586.

Hunt, A. and Blake, D. (2014). A General Procedure for Constructing Mortality Models. *North American Actuarial Journal* 18(1):116–138. doi:10.1080/10920277.2013.852963.

Hunt, A. and Blake, D. (2020). Identifiability in age/period/cohort mortality models. *Annals of Actuarial Science* 14(2): 500–536. doi:10.1017/S1748499520000123.

Hunt, A. and Blake, D. (2021a). A Bayesian approach to modeling and projecting cohort effects. *North American Actuarial Journal* 25(sup1): S235–S254. doi:10.1080/10920277.2019.1649157.

Hunt, A. and Blake, D. (2021b). On the structure and classification of mortality models. *North American Actuarial Journal* 25(sup1): S215–S234. doi:10.1080/10920277.2019.1649156.

Lee, R.D. and Carter, L.R. (1992). Modeling and forecasting US mortality. *Journal of the American Statistical Association* 87(419): 659–671. doi:10.1080/01621459.1992.10475265.

Lexis, W.H.R.A. (1879). Sur la durée normale de la vie humaine et sur la théorie de la stabilité des rapports statistiques. *Annales de Démographie Internationale* 2: 447–460.

Marino, M., Levantesi, S., and Nigri, A. (2023). A neural approach to improve the Lee–Carter mortality density forecasts. *North American Actuarial Journal* 27(1): 148–165. doi:10.1080/10920277.2022.2050260.

Pearson, K. (1897*). Chances of death*. London: Edward Arnold.

Pewsey, A. (2000). Problems of inference for Azzalini's skewnormal distribution. *Journal of Applied Statistics* 27(7): 859–870. doi:10.1080/02664760050120542.

Plat, R. (2009). On stochastic mortality modeling. *Insurance: Mathematics and Economics* 45(3): 393–404. doi:10.1016/j.insmatheco.2009.08.006.

Raftery, A.E., Chunn, J.L., Gerland, P., and Ševčíková, H. (2013). Bayesian probabilistic projections of life expectancy for all countries. *Demography* 50(3): 777–801. doi:10.1007/s13524-012-0193-x.

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. doi:10.1016/j.insmatheco.2005.12.001.

Scognamiglio, S. (2022). Calibrating the Lee–Carter and the Poisson Lee–Carter models via neural networks. *ASTIN Bulletin* 52(2): 519–561. doi:10.1017/asb.2022.5.

Stoeldraijer, L., Duin, C. van, Wissen, L. van, and Janssen, F. (2013). Impact of different mortality forecasting methods and explicit assumptions on projected future life expectancy: The case of the Netherlands. *Demographic Research* 29(13): 323–354. doi:10.4054/DemRes.2013.29.13.

Vaupel, J.W. and Canudas-Romo, V. (2003). Decomposing change in life expectancy: A bouquet of formulas in honor of Nathan Keyfitz's 90th birthday. *Demography* 40(2): 201–216. doi:10.1353/dem.2003.0018.

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research* 11(116): 3571–3594.

Willets, R.C. (2004). The cohort effect: Insights and explanations. *British Actuarial Journal* 10(4): 833–877. doi:10.1017/S1357321700002762.

Zanotto, L., Canudas-Romo, V., and Mazzuco, S. (2021). A mixture-function mortality model: Illustration of the evolution of premature mortality. *European Journal of Population* 37(1): 1–27. doi:10.1007/s10680-019-09552-x.

Appendix

Prior distributions

The prior distributions are as follows:

- $log(\alpha_{t-x}) \sim U(log(\alpha_\mu) - 0.7, log(\alpha_\mu) + 0.4)$;
- $\alpha_\mu \sim Beta(1.5,1.5) + 0.5$;
- $\zeta_t \sim Dir(1,1,8)$;
- $\sigma_{i,t} \sim U(0,1)$;
- $\mu_{m,t} \sim N(30,5)$;
- $\sigma_{m,t} \sim Inv.\Gamma(0.001,0.001)$;
- $\gamma_{m,t} \sim N(0,1)$, restricted to $(0, +0.99527)$ (Aliverti, Mazzuco, and Scarpa 2022; Pewsey 2000);
- $\mu_{M,t} \sim N(80,10)$;
- $\sigma_{M,t} \sim Inv.\Gamma(0.001,0.001)$;
- $\gamma_{M,t} \sim N(0,1)$ restricted to $(-0.99527, +0.99527)$ (Pewsey 2000).

In selecting the priors for our model, we aimed to achieve a balance between several objectives: ensuring a realistic representation of the age-at-death distribution, maintaining clear separation of the three mortality components to reduce the chance of label switching, and minimizing influence on the posterior distributions.

The issue we faced with cohort effect multipliers was the desire to have as little information as possible embedded in the prior distributions, while keeping estimates within a reasonable range. A flat prior on the logarithmic scale was chosen, as the $[-0.7; 0.4]$ interval on the log scale corresponds approximately to the $[0.5; 1.5]$ interval on the real scale, with the probability of death after applying cohort effects potentially ranging between 50% and 150% of the probability without cohort effects – a very wide range. While we deemed a range of 100 percentage points to be sufficiently wide, we did not want to put hard limits on the maximum size of a cohort effect. Hence, we shrunk the estimates of the α_{t-x} parameters towards a parameter α_μ, thus theoretically allowing for larger cohort effects as long as the range of cohort effects remains under 100 percentage points. On the other hand, cohort effects describe the impact of cohort on mortality net of age and period effects and a common restriction in APC models is for them to have a zero mean on the log scale (Hunt and Blake 2020); hence we expect them to have a prior mean of 0 on the log scale as well (1 on the original scale). This consideration led us to choose a weakly informative distribution, $Beta(1.5,1.5)$, over the uniform.

For the ζ_t parameter, the prior was chosen in order to obtain, on average, a distribution of 10% for both infant and premature mortality, and 80% for adult mortality. This distribution ensures that the adult mortality component accounts for the majority of deaths, while still permitting the infant and premature components a substantial range of plausible values, in line with the empirical share of deaths at ages 0–2 and under 40 years. In our tests the specific choice of the prior of the ζ_t parameter did not affect estimates significantly as long as the adult mortality component represented the majority of deaths.

The mean parameters of the two skew-normal distributions, $\mu_{m,t}$ and $\mu_{M,t}$, are assigned Gaussian priors. We decided not to choose a more heavy-tailed distribution for the priors in order to keep the components distinct and with separate demographic interpretations. A central issue is that the mortality hump is a rather small part of the age-at-death curve, although one that retains demographic significance and whose inclusion is meaningful. Even though to a certain degree all deaths before the modal age at death in old age can be seen as premature, a premature mortality component with a mean age at death close to the mean of old age mortality can no longer be considered a proper premature mortality component. The 95% credibility intervals for the prior means are slightly tighter than $[20; 40]$ and $[60; 100]$, ensuring a wide range of prior parameters, but also minimizing the risk of overlap. The variance parameters $\sigma_{m,t}$ and $\sigma_{M,t}$ have very weakly informative distributions. Lastly, the skewness parameters $\gamma_{m,t}$ and $\gamma_{M,t}$ have truncated Gaussian priors, rather than uniform priors. This choice is intended to mildly shrink the estimates towards zero, thereby regularizing them, since we found that uniform priors led to unstable estimates in our testing. The restriction to $(-0.99527, +0.99527)$ is due to unstable estimates with values of the skewness parameter being close to ± 1, as discussed by Pewsey (2000), and its practical effect is to regularize estimates akin to the use of weakly informative priors advocated by Gelman et al. (2008).

Parameter restrictions and initial values

The model presents several challenges that require careful handling to ensure meaningful estimation in Stan. To address these challenges, we impose specific restrictions.

First, we restrict the α_{t-x} cohort effects. Their geometric mean is set to 1, expressed as $\frac{\prod_{c=1}^{C} \alpha_c}{C} = 1$. This is equivalent to the common constraint in APC models that the cohort parameters should be centered around zero, as discussed in Hunt and Blake (2020). Additionally, we constrain $\gamma_{m,t}$ and $\gamma_{M,t}$ to fall within the range of $(-0.99527, 0.99527)$. This constraint is based on insights from Pewsey (2000), who note inference issues with more extreme values, since the skew-normal distribution becomes increasingly degenerate as the skewness parameter γ approaches 1. We also enforce non-negativity

for $\gamma_{m,t}$, extending Aliverti, Mazzuco, and Scarpa's model (2022) where the premature mortality component is restricted to a normal distribution.

A common challenge with Bayesian mixture models is label switching, where MCMC can misidentify and swap mixture components. In our case, this risk arises with the two skew-normal components, premature and adult mortality. To mitigate this risk, we carefully select prior distributions, in line with demographic expertise, and initial values. In fact, label switching is the only issue to which the choice of prior distributions is sensitive: this has led us to choose normals as prior distributions for $\mu_{m,t}$ and $\mu_{M,t}$ with reasonable parameters for premature and adult mortality, respectively. While the choice of priors ensures sufficient separation, as a further precaution and as a safeguard against accidental label switching during MCMC sampling, we enforced $\mu_{m,t} \leq \mu_{M,t}$ by declaring the two parameters as an ordered vector: this restriction can be safely dropped if the components are clearly separated.

Throughout our analysis, we observed that overly broad prior distributions and incorrect initial parameters cause the model components to fail to consistently describe the same features of the age-at-death curve, leading to poor chain mixing, multimodality in posterior distributions, and low effective numbers of draws (n_{eff}).

Of particular concern are the starting values for the share of adult mortality ζ_3, μ_m, σ_m, and σ_M, as well as the difference in starting values $\mu_M - \mu_m$. Starting values for these parameters must be such that the model components adequately describe the demographic phenomena that we wish to model. Our experience suggests that good chain mixing can be consistently achieved with the following ranges for these initial values:

- $\zeta_3 >= 0.5$ (adult mortality accounts for the majority of the deaths)
- $20 < \mu_m < 40$ (mean age at premature death is at young ages)
- $75 < \mu_M < 85$ (mean age at adult death is close or slightly lower than modal age at death)
- $5 < \sigma_m < 12$ (95% of premature mortality probability lies in a 20 to 40-year interval)
- $5 < \sigma_M < 12$ (95% of adult mortality probability lies in a 20 to 40-year interval)
- $0.15 < \gamma_m < 0.5$ (moderate positive skewness of adult mortality)
- $-0.9 < \gamma_M < -0.15$ (adult mortality is negatively skewed)

While the ranges may appear arbitrary, they reflect reasonable expectations of what the model components are expected to represent. Adult mortality accounts for the overwhelming majority of deaths, hence $\zeta_3 >= 0.5$. As for $5 < \sigma_m < 12$ and $5 < \sigma_M < 12$, a small variance parameter implies that the component only governs death over a narrow range of ages, which leads to the mixing parameter being reduced accordingly and the other parameter compensating, thus trying to model both components and failing.

On the other hand, a larger σ, for example 20, means that the component is trying to fit almost the whole age-at-death curve and therefore resembles more of a background mortality component, more similar to a Gompertz–Makeham setting.

The skewness settings have been chosen along similar lines. The adult mortality component is known to be negatively skewed and a wide range such as $-0.9 < \gamma_M < -0.15$ ensures that the component adequately models adult mortality without leaving out too much of middle-age mortality (if $\gamma_M > -0.15$) and without degenerating to an excessively skewed distribution (if $\gamma_M < -0.9$). On the other hand, premature mortality is less skewed than adult mortality, with posterior parameters often relatively close to zero. Initial parameters indicating moderate positive skewness allow the posterior to accomodate a wider range of values while still retaining salience. Initial skewness values closer to zero lead to both components, one symmetrical and one asymmetrical, modeling adult mortality. Higher positive skewness, on the other hand, can result in the premature mortality component becoming excessively skewed and degenerate.

Finally, the mean of premature mortality is set to values within its 95% prior credibility interval and the mean of old age mortality is set to values in the vicinity of the mode.

Actual generation of the initial values was performed according to the following distributions. Other distributions that meet the criteria outlined above may be equally appropriate.

- $\zeta_{1,t} \sim \frac{U(0.05,0.15)}{\sum_{i=1}^{3}\zeta_i}$;
- $\zeta_{2,t} \sim \frac{U(0.2,0.3)}{\sum_{i=1}^{3}\zeta_i}$;
- $\zeta_{3,t} \sim \frac{U(0.6,0.8)}{\sum_{i=1}^{3}\zeta_i}$;
- $\sigma_I \sim U(0.8,1)$;
- $\mu_{m,t} \sim U(20,35)$;
- $\sigma_{m,t} \sim U(5,12)$;
- $\gamma_{m,t} \sim U(0.15,0.5)$;
- $\mu_{M,t} \sim U(75,85)$;
- $\sigma_{M,t} \sim U(5,12)$;
- $\gamma_{M,t} \sim U(-0.9,-0.15)$;
- $log(\alpha_{t-x}) \sim U(-0.22,0.18)$.

Overall, the rationale behind the parameter restrictions and choice of initial values is to fit the distribution of deaths from a starting curve where each parameter retains an unambiguous and meaningful demographic interpretation.

Convergence of chains

Convergence of the MCMC chains is assessed through the Gelman–Rubin statistic $\hat{R}$, which quantifies the mixing of the chains and how well they achieve a stationary distribution, and the effective sample size n_{eff}, which measures the equivalent number of independent and identically distributed samples that would provide the same quantity of information in the MCMC sample.

As shown in Table A-1, $\hat{R}$ is below 1.1 in all 49 years for all regions and parameters except for σ_I in Lazio and Sardinia and ζ_1 in Sardinia. ζ_1 is usually very small, below 1%, but in some years after 2000, due to low infant mortality and a slightly larger mortality hump, one of the chains temporarily diverges towards a higher proportion of infant mortality than warranted and also a higher variability. This issue may be resolved by substituting the infant mortality component with a point probability at age 0, but since infant mortality is usually well identified, we decided to keep the half-normal in order to improve fit.

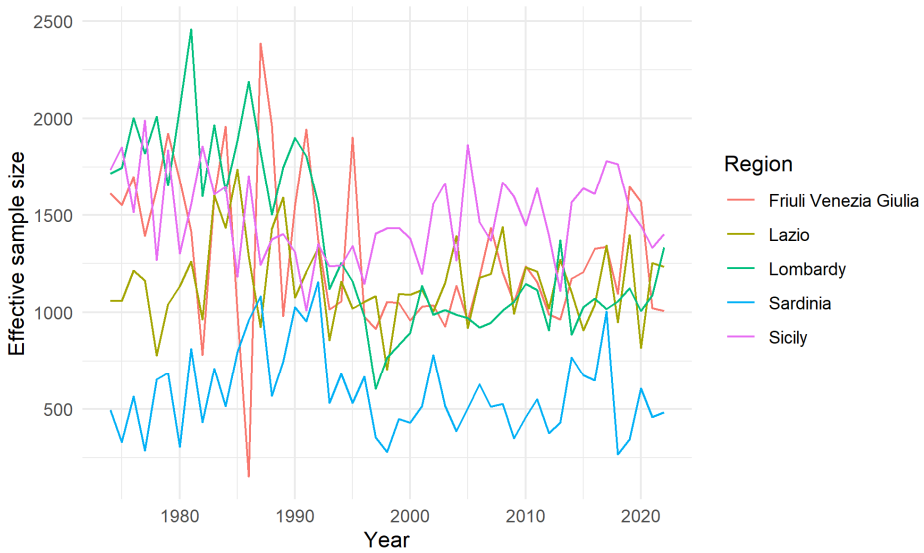
Table A-1: Number of years (out of 49), by region and parameter, where $\hat{R} < 1.1$

Parameter	Friuli Venezia Giulia	Lazio	Lombardy	Sardinia	Sicily
ζ_1	49	49	49	48	49
ζ_2	49	49	49	49	49
σ_I	49	48	49	48	49
μ_m	49	49	49	49	49
σ_m	49	49	49	49	49
γ_m	49	49	49	49	49
ξ_m	49	49	49	49	49
μ_M	49	49	49	49	49
σ_M	49	49	49	49	49
γ_M	49	49	49	49	49
ξ_M	49	49	49	49	49
α_{1915}	49	49	49	49	49
α_{1916}	49	49	49	49	49
α_{1917}	49	49	49	49	49
α_{1918}	49	49	49	49	49
α_{1919}	49	49	49	49	49
α_{1920}	49	49	49	49	49
α_{1921}	49	49	49	49	49
α_{1922}	49	49	49	49	49
α_{1923}	49	49	49	49	49
α_{1924}	49	49	49	49	49
α_{1925}	49	49	49	49	49

In Figure A-1 the effective sample size n_{eff} is shown for all years and regions for ζ_2, the proportion of deaths attributed to premature mortality, which is the parameter most likely to diverge and be less stable over time. Average n_{eff} across all regions and years

is 1,171.5. Minimum effective sample size for ζ_2 is 151.0 (FVG, 1986), with a value lower than 500 in 19 cases (7.7%). For more stable parameters the effective sample size is substantially higher: mean age at death of adult mortality μ_M has an average n_{eff} of 2,075.0 and a minimum of 413.8 (Sardinia, 1975). In both cases, effective sample size is lower after 1990. Each cohort effect has a single estimate for all years and the effective sample size is much higher: for the 1920 cohort the average across all regions is 19,029.0.

Figure A-1: Effective sample size for the proportion of deaths due to premature mortality, by region and year



Sensitivity to number of iterations

The model for Lombardy was run with a lower number of iterations per chain in order to assess whether the results were sensitive to the length of the chains. The results are presented for 4,000 iterations per chain with a burn-in of 2,000 iterations, and 2,000 iterations per chain with a burn-in of 1,000 iterations. The differences are minimal and are presented in Figures A-2 and A-3 for the year 2004, one of the years when the differences are most noticeable.

Figure A-2: Premature mortality average age at death, density, year 2004, 1,000 iterations (blue) and 2,000 iterations (red)

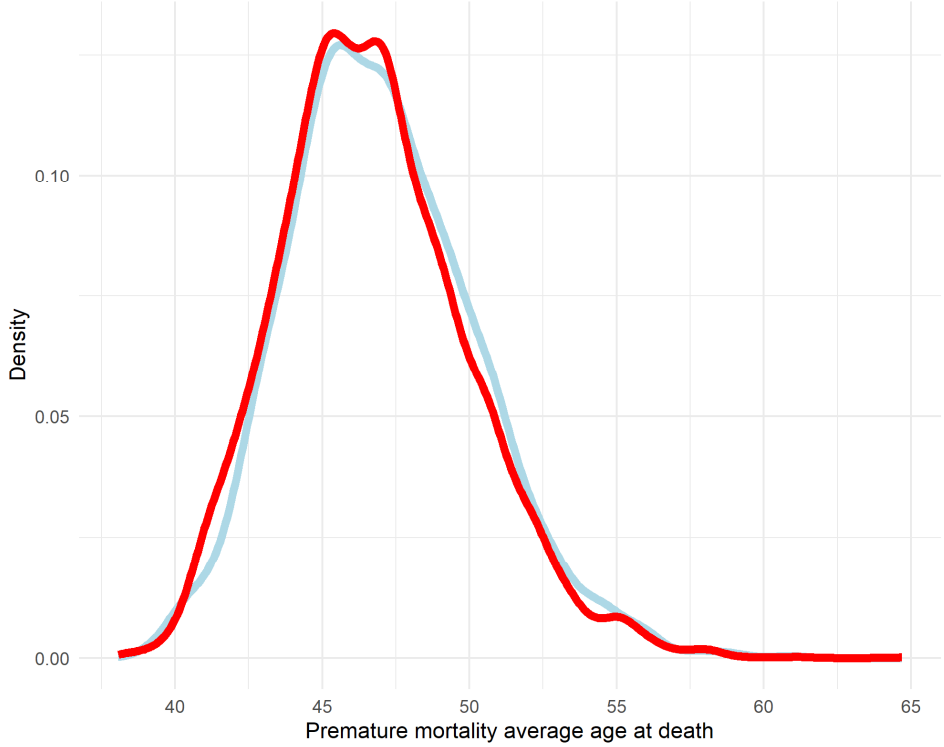
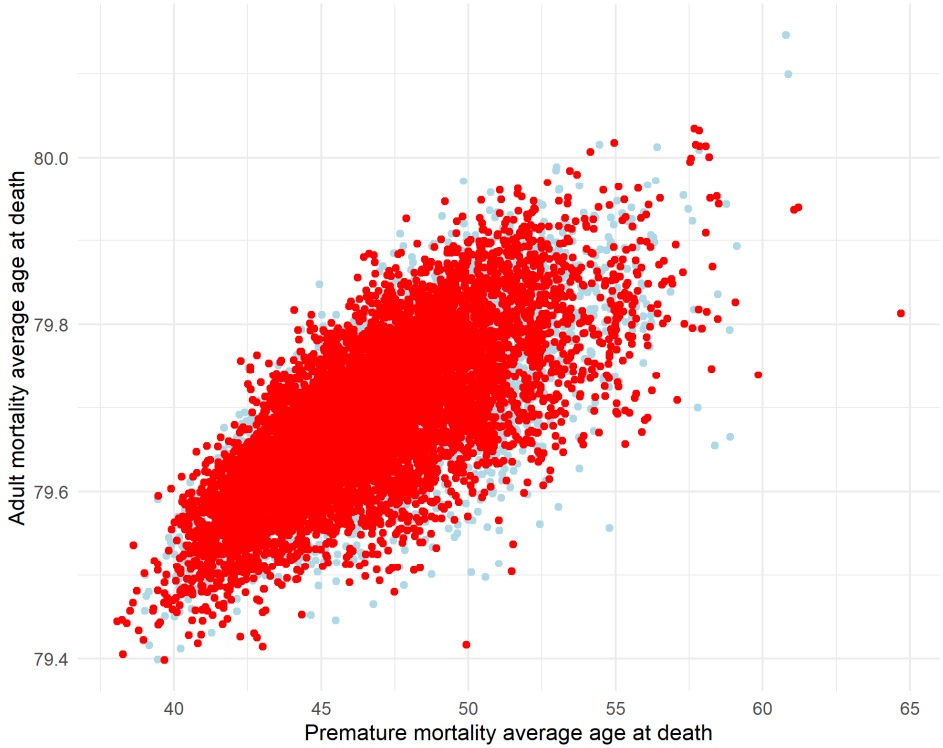


Figure A-3: Premature mortality and adult mortality average ages at death, year 2004, 1,000 iterations (blue) and 2,000 iterations (red)

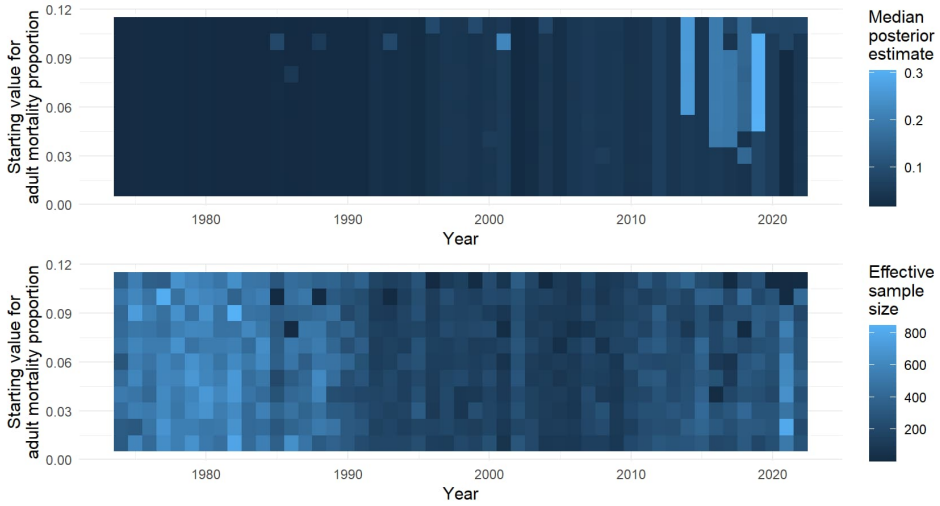


There is no noticeable difference in the forecasts of the number of deaths by age, which are therefore not shown.

Sensitivity to starting values

As previously mentioned, the starting values for the model should be chosen carefully in order for the chains to initialize and converge. An example from a model estimated without cohort effects on single years is shown in Figure A-4: the plots show how the proportion of deaths attributable to premature mortality rises over time, but with convergence issues (n_{eff} as low as 2) and implausibly high estimates more common with higher starting values of ζ_2.

Figure A-4: Median posterior estimate and effective sample size by year for premature variability proportion, model without cohort effects



Model parameters

The model parameters for age x, year t, and cohort $t-x$, with their respective prior distributions, are as follows:

Mixing parameter:

- $\zeta_t \sim Dir(1,1,8)$, with $\zeta_{1,t}$ representing the proportion of infant mortality deaths, $\zeta_{2,t}$ the proportion of premature mortality deaths, and $\zeta_{3,t}$ the proportion of adult mortality deaths;

Cohort effects parameters:

- $\alpha_\mu \sim Beta(1.5,1.5) + 0.5$;
- $log(\alpha_{t-x}) \sim U\big(log(\alpha_\mu) - 0.7, log(\alpha_\mu) + 0.4\big)$;

Infant mortality parameter:

- $\sigma_{i,t} \sim U(0,1)$;

Premature mortality parameters:
Centered parametrization:

- $\mu_{m,t} \sim N(30,5)$, restricted to be positive and $\leq \mu_{M,t}$;
- $\sigma_{m,t} \sim Inv.\Gamma(0.001,0.001)$, restricted to be positive;
- $\gamma_{m,t} \sim N(0,1)$, restricted to $(0, +0.99527)$;

Direct parametrization:

- $\xi_{m,t} = \mu_{m,t} - \frac{\sigma_{m,t}\mu_{m,z,t}}{\sqrt{1-\mu_{m,z,t}^2}}$;
- $\omega_{m,t} = \frac{\sigma_{m,t}}{\sqrt{1-\mu_{m,z,t}^2}}$;
- $\lambda_{m,t} = \frac{\mu_{m,z,t}\sqrt{\pi/2}}{\sqrt{1-\frac{\pi\mu_{m,z,t}^2}{2}}}$
- $c_{m,t} = sgn(\gamma_{m,t})(\frac{2\gamma_{m,t}}{4-\pi})^{1/3}$;
- $\mu_{m,z,t} = \frac{c_{m,t}}{\sqrt{1+c_{m,t}^2}}$;

Adult mortality parameters:
Centered parametrization:

- $\mu_{M,t} \sim N(80,10)$, restricted to be positive and $\geq \mu_{m,t}$;
- $\sigma_{M,t} \sim Inv.\Gamma(0.001,0.001)$, restricted to be positive;
- $\gamma_{M,t} \sim N(0,1)$ restricted to $(-0.99527, +0.99527)$;

Direct parametrization:

- $\xi_{M,t} = \mu_{M,t} - \frac{\sigma_{M,t}\mu_{M,z,t}}{\sqrt{1-\mu_{M,z,t}^2}}$;
- $\omega_{M,t} = \frac{\sigma_{M,t}}{\sqrt{1-\mu_{M,z,t}^2}}$;
- $\lambda_{M,t} = \frac{\mu_{M,z,t}\sqrt{\pi/2}}{\sqrt{1-\frac{\pi\mu_{M,z,t}^2}{2}}}$
- $c_{M,t} = sgn(\gamma_{M,t})(\frac{2\gamma_{M,t}}{4-\pi})^{1/3}$;

- $\mu_{M,z,t} = \frac{c_{M,t}}{\sqrt{1+c_{M,t}^2}}$;

Infant, premature, and adult mortality components:

- $f_I(\sigma_{i,t})$ is the density function of a half-normal (0, $\sigma_{i,t}$);
- $f_m(\xi_{m,t}, \omega_{m,t}, \lambda_{m,t})$ is the density function of a skew-normal ($\xi_{m,t}, \omega_{m,t}, \lambda_{m,t}$);
- $f_M(\xi_{M,t}, \omega_{M,t}, \lambda_{M,t})$ is the density function of a skew-normal ($\xi_{M,t}, \omega_{M,t}, \lambda_{M,t}$);

Probability of dying at age x in year t, mixture without cohort effects, not normalized:

- $f(x;\theta_t) = f(x;\zeta_t, \sigma_{i,t}, \xi_{m,t}, \omega_{m,t}, \lambda_{m,t}, \xi_{M,t}, \omega_{M,t}, \lambda_{M,t}) = \zeta_{1,t} f_I(x;\sigma_{i,t}) + \zeta_{2,t} f_m(x;\xi_{m,t}, \omega_{m,t}, \lambda_{m,t}) / (F_m(100) - F_m(0)) + \zeta_{3,t} f_M(x;\xi_{M,t}, \omega_{M,t}, \lambda_{M,t}) / (F_M(100) - F_M(0))$, with F_m and F_M being the cumulative distribution functions of f_m and f_M, respectively, with the same parameters.

Probability of dying at age x in year t including cohort effects:

- $P(x;\theta_t, t, \alpha) = \frac{\alpha_{t-x} f(x;\theta_t)}{\sum_{x=0}^{100} \alpha_{t-x} f(x;\theta_t)}$.

Uncertainty of parameter estimates

Table A-2: 10th percentile of parameter estimates by region: Median, 1974 and 2022 estimates

Region	Estimate	ζ_1	ζ_2	ζ_3	σ_I	μ_m	σ_m	γ_m	μ_M	σ_M	γ_M
Friuli Venezia Giulia	1974	0.016	0.02	0.96	0.52	20.17	10.86	0.00	70.66	13.67	–0.75
	2022	0.002	0.06	0.92	0.58	54.04	20.47	0.00	82.10	10.02	–0.81
	Median	0.003	0.03	0.96	0.59	30.98	13.37	0.01	76.78	11.81	–0.74
Lazio	1974	0.017	0.03	0.95	0.53	23.99	16.19	0.02	73.85	11.85	–0.67
	2022	0.002	0.08	0.88	0.59	61.86	21.68	0.00	82.59	9.68	–0.79
	Median	0.004	0.05	0.94	0.55	39.62	17.16	0.01	77.99	10.99	–0.70
Lombardy	1974	0.016	0.02	0.96	0.55	17.35	11.34	0.00	70.81	12.67	–0.73
	2022	0.002	0.11	0.86	0.60	70.63	24.38	0.00	83.06	9.15	–0.79
	Median	0.003	0.04	0.95	0.58	32.42	12.77	0.00	77.20	11.40	–0.69
Sardinia	1974	0.019	0.06	0.90	0.58	30.61	24.80	0.02	75.42	12.25	–0.74
	2022	0.001	0.06	0.91	0.57	51.32	19.88	0.01	81.40	10.61	–0.81
	Median	0.003	0.06	0.92	0.58	34.04	16.14	0.01	78.09	11.59	–0.73
Sicily	1974	0.023	0.05	0.92	0.56	27.96	27.10	0.02	75.42	11.88	–0.71
	2022	0.002	0.09	0.88	0.55	62.19	22.90	0.00	81.54	9.72	–0.78
	Median	0.006	0.05	0.93	0.56	36.20	20.56	0.01	77.69	11.04	–0.72

Table A-3: 90th percentile of parameter estimates by region: Median, 1974 and 2022 estimates

Region	Estimate	ζ_1	ζ_2	ζ_3	σ_I	μ_m	σ_m	γ_m	μ_M	σ_M	γ_M
Friuli Venezia Giulia	1974	0.017	0.03	0.96	0.56	21.68	12.40	0.07	70.81	13.81	–0.74
	2022	0.002	0.08	0.94	0.68	61.82	23.45	0.11	82.32	10.24	–0.80
	Median	0.003	0.03	0.97	0.67	36.25	17.30	0.13	76.98	11.99	–0.72
Lazio	1974	0.018	0.04	0.95	0.56	28.98	21.02	0.27	74.03	12.02	–0.65
	2022	0.002	0.12	0.92	0.69	70.55	24.37	0.07	82.83	9.96	–0.77
	Median	0.005	0.06	0.95	0.62	44.27	21.02	0.15	78.22	11.19	–0.68
Lombardy	1974	0.018	0.02	0.96	0.58	18.80	13.30	0.10	70.94	12.78	–0.71
	2022	0.002	0.14	0.89	0.70	78.16	26.73	0.06	83.22	9.36	–0.77
	Median	0.004	0.05	0.96	0.65	34.61	14.23	0.07	77.37	11.54	–0.67
Sardinia	1974	0.020	0.08	0.92	0.61	38.41	31.39	0.27	75.65	12.46	–0.72
	2022	0.002	0.09	0.94	0.69	60.88	23.37	0.13	81.72	10.90	–0.80
	Median	0.005	0.07	0.94	0.63	38.41	20.13	0.11	78.28	11.82	–0.71
Sicily	1974	0.024	0.06	0.93	0.59	35.31	35.16	0.28	75.58	12.04	–0.70
	2022	0.003	0.12	0.91	0.63	69.68	25.42	0.07	81.73	9.96	–0.76
	Median	0.007	0.06	0.94	0.61	41.70	24.08	0.21	77.88	11.23	–0.70