Draft of the paper submitted at the European Population Conference
Bologna, 3-6 June 2026

Life Table Functions and Their Uncertainty: A Matrix Algebra Approach

Carlo G. Camarda
Institut National d'Études Démographiques
carlo-giovanni.camarda@ined.fr

November 1, 2025

Abstract

The construction of confidence intervals for life expectancy at birth has a long tradition in demography, reflecting the central role of uncertainty in mortality estimation, particularly when analyzing small populations or subnational areas. Yet, despite this longstanding interest, no unified statistical framework currently exists for quantifying uncertainty across all life table functions. This paper introduces a coherent, statistically grounded, and computationally refined approach for propagating uncertainty throughout the life table. Building on the Poisson nature of death counts, we derive and compare three complementary methods for quantifying uncertainty in age-specific mortality: the Delta method, the Garwood exact approach, and a Bayesian bootstrap based on Poisson-Gamma conjugacy. These methods are then embedded in a matrix algebra formulation that unifies the computation of all major life table functions, survivorship, deaths, person-years, total years remaining, and life expectancy, and their associated variance–covariance structures. All analytical derivations are presented explicitly, providing both a methodological framework for uncertainty propagation and a formal, elegant reinterpretation of the life table as an interconnected stochastic system. The approach is illustrated using 2020 female mortality data from the 20 administrative districts (*arrondissements*) of Paris, highlighting how uncertainty behaves across populations of varying size and mortality levels. Finally, several extensions are outlined, including the quantification of uncertainty for lifespan variability measures, the adaptation to abridged life tables, and applications to model-based and comparative mortality analyses. These developments lay the groundwork for a unified, probabilistic framework for the study of demographic functions.

Keywords: life table · uncertainty quantification · matrix algebra · formal demography · small-area demography

1 Introduction

Life tables, a cornerstone of formal demography and quantitative population research, provide a concise framework for summarizing age-specific patterns of mortality and survival. Since their introduction by Graunt (1662), they have served as a fundamental tool for assessing variations in mortality across periods and cohorts, as well as for informing evidence-based decisions in public health, social policy, and risk assessment. By translating deaths and exposures into interpretable age-specific rates, life tables support the evaluation of pension systems, the design of targeted health interventions, and predictive modeling in insurance and related domains.

Mortality is inherently stochastic, and observed deaths represent only one realization of this process; consequently, life table functions are subject to unavoidable variability. Quantifying this variability is therefore essential. Classic demographers focused on large populations, in which stochastic variation is relatively small, and often treated mortality as effectively deterministic. As Keyfitz (1966) noted, "The mere fact that over a period of a year the population is likely to be fifty or a hundredfold the deaths will result in a higher order of precision" (p. 312), illustrating that in large populations random fluctuations tend to average out.

However, the increasing availability of detailed mortality data, across smaller populations, specific subgroups, or shorter time intervals, has made it increasingly important to account explicitly for uncertainty. In such contexts, stochastic variation can substantially affect indicators like life expectancy at birth. Similarly, when life tables are derived from model-based estimates or projections, propagating variability prevents overconfidence in predictions. Accurate assessment of this variation is also critical when comparing populations or cohorts, evaluating public health interventions, designing pension systems, calculating insurance premiums, or conducting demographic research and modeling. By explicitly accounting for stochastic variation, researchers and policymakers can make more robust, evidence-based decisions and communicate the reliability of mortality measures with confidence.

Building on the recognition that mortality is inherently stochastic and that variability can substantially affect life table functions, we present a systematic approach to quantify uncertainty across all life table functions. While most previous research has focused primarily on life expectancy at birth, our framework captures uncertainty comprehensively, providing a more nuanced and robust view of demographic dynamics. Importantly, the approach is flexible and can also be applied when life tables are derived from model-based mortality estimates, extending its applicability beyond analyses based solely on observed data.

Traditionally, two main approaches have been used to quantify uncertainty in life table estimates. The first, rooted in the seminal works of Chiang (1960a,b, 1961), adopts an analytical framework to derive confidence intervals for life expectancy and related functions. Chiang's approach, grounded in statistical theory, expresses the variance of life table functions as a function of mortality rates and their covariances, providing a mathematically elegant and computationally efficient solution. The second approach, which has become increasingly prominent with the rise of modern computational power, relies on simulation-based methods that generate large ensembles of life tables to empirically assess variability. Through repeated sampling of mortality schedules, these methods infer uncertainty directly from the dispersion of simulated outcomes (Andreev and Shkolnikov, 2010; Scherbov and Ediev, 2011; Silcocks et al., 2001).

Building on Chiang's analytical foundation and well-established statistical principles, we embed the life table system within a matrix algebra framework that unifies and generalizes the quantification of uncertainty across its functions. This formulation enables classical analytical methods to be applied systematically, reducing or even eliminating the need for large-scale simulations while maintaining mathematical rigor. The resulting framework is flexible and elegant, applicable to life tables derived from both raw and modeled mortality data.

Setting aside concerns about computational efficiency, which modern computing has rendered almost trivial, this formulation integrates uncertainty into life tables in a mathematically rigorous way, providing a formal representation that enhances both interpretability and methodological coherence.

Finally, in the following we focus on life tables constructed for single years of age and on period life tables, which summarize the mortality experience that a hypothetical cohort would face if it were subject, throughout its lifetime, to the age-specific mortality rates observed in a given period. Abridged and cohort life tables are not considered here but will be addressed in the full version of the paper. In this long abstract, we present our framework only for life tables derived from raw data; extensions to modeled data—including two-dimensional cases in which mortality is modeled over both age and time—will be presented later. Conceptually, the construction and the principles of uncertainty quantification remain the same across all these forms of life tables.

As an illustrative example for this study, we consider female mortality in 2020 across the 20 *arrondissements*, the administrative districts of Paris. Observed deaths vary substantially between districts, ranging from 39 in the 2nd arrondissement to 1005 in the more populous 15th arrondissement. Corresponding exposures similarly span a wide range, from approximately 7,663 person-years in the 1st arrondissement to 125,912 in the 15th, reflecting differences in population size and highlighting the heterogeneity across districts.

2 A general formulation

The construction of life tables relies on a fundamental, yet often overlooked, assumption: mortality rates are constant within each age interval, corresponding to a piecewise constant hazard model. Despite its simplicity, this model is widely employed when prior knowledge of the underlying hazard function is limited or when the aim is to describe mortality patterns without imposing a specific parametric form (Friedman, 1982). Within a life table context, this framework rests on the implicit assumption that the hazard of death remains constant within each age interval. This assumption is particularly strong at specific ages, such as at birth and in the last open-age interval, where adjustments are typically applied to improve the accuracy of life table functions. It is also more pronounced in abridged life tables, where relatively large age groups are used.

If it were possible to observe the complete lifetimes of all individuals, the maximum likelihood estimate of the age-specific constant hazard at each age would simply be the ratio of the number of deaths to the total time at risk, summed across individuals within that age interval (i.e., the person-years). In practice, however, demographers typically work with aggregated data, and the denominator is approximated by what is commonly referred to as exposure. Several methods have been proposed for estimating exposure, with the choice of approximation depending on the characteristics of the data and the assumptions one is willing to make. Two widely used approaches in demography are the midpoint approximation, which assumes that individuals are exposed to risk for half

of the interval, and the linear interpolation method, which assumes a constant hazard between ages (Preston et al., 2001, p. 15). When working with single years of age and populations with modest growth rates, both methods yield very similar estimates of age-specific hazards. For a detailed discussion of the various techniques employed, for instance, in the Human Mortality Database (2025), including cases with original data aggregated in broader age groups, see Sections 5 and 6 of Wilmoth et al. (2021).

Given the chosen approximation method, the data can be represented by two vectors, $\boldsymbol{e} = (e_i)$ and $\boldsymbol{y} = (y_i)$, containing the exposures and observed number of deaths, respectively, over a sequence of ages $\boldsymbol{x} = (x_i)$ for $i = 1, \dots, n$. For simplicity, we set the first available age to zero; however, without loss of generality, different starting ages could be considered. Given the equivalence between the likelihood of piecewise exponential survival data and the likelihood of a Poisson distribution, inference for this survival model can be carried out using standard tools for log-linear Poisson regression (Holford, 1980; Laird and Olivier, 1981). Let $\boldsymbol{Y} = (Y_i)$ denote a vector of independent Poisson random variables representing the number of deaths at each age. The observed counts, $\boldsymbol{y} = (y_i)$, are then realizations of these random variables:

$$\boldsymbol{Y} \sim \mathcal{P}(\boldsymbol{e} * \boldsymbol{\mu}), \tag{1}$$

where $*$ denotes the element-wise product. Equivalently, the expected value of each observed count is $\mathbb{E}[y_i] = e_i\, \mu_i$, given by the product of the corresponding exposure e_i and the age-specific force of mortality μ_i. In constructing a life table from raw deaths and exposures, we assume that each age-specific force of mortality is estimated independently.

With a logarithmic link, the estimated coefficients correspond to the log-rates, defined as the element-wise logarithm of the ratio of deaths to exposures:

$$\boldsymbol{\eta} = \ln(\boldsymbol{\mu}) = \ln\left(\frac{\boldsymbol{y}}{\boldsymbol{e}}\right). \tag{2}$$

In a Poisson framework, the variance of observed deaths equals their expected value. For independent, non-parametric age-specific estimates, the variance–covariance matrix is diagonal, with entries corresponding to the expected counts. In our piecewise constant hazard model, we approximate these expected values with the observed deaths, giving

$$\boldsymbol{V}_{\boldsymbol{y}} = \texttt{diag}(\boldsymbol{y}) = \boldsymbol{D}_{y}\,. \tag{3}$$

For notational convenience, we introduce the compact operator $\boldsymbol{D}_{y}$ to denote a diagonal matrix with the elements of $\boldsymbol{y}$ on the main diagonal. In the following, we use $\boldsymbol{V}_{\boldsymbol{y}}$ to represent the variance–covariance matrix, although, more formally, it corresponds to $\boldsymbol{V}_{\boldsymbol{Y}} = \boldsymbol{D}_{\mathbb{E}[\boldsymbol{Y}]}$.

2.1 Uncertainty for the log-rates

To assess the uncertainty of quantities derived from deaths, starting from the estimated log-rates, we first need to translate the variance from the space of observed counts to that of the modeled parameters. This step lays the groundwork for propagating uncertainty through all subsequent life table functions, which are unaffected by cohort size thanks to the inclusion of the exposure values $\boldsymbol{e}$, serving as an offset in the modeling framework. We then consider three complementary approaches for this transformation: the Delta method, the Garwood exact formulation, and a Bayesian bootstrap procedure. Although

these methods are conceptually different, they all provide consistent ways to quantify the variability of the log-rates and, by extension, of any life table function derived from them. For large populations, the three approaches tend to converge, producing nearly identical uncertainty estimates; yet, when dealing with small populations, the choice of method becomes decisive, as stochastic variability plays a much larger role.

2.1.1 Delta Method

The Delta method provides an analytical, large-sample approximation, relying on the asymptotic normality of estimators, for quantifying uncertainty in smooth functions of random variables. In our context, it is first applied to propagate variance from the Poisson death counts to the logarithm of the estimated rates, $\boldsymbol{\eta}$. We then use it to further propagate uncertainty from the estimated log-rates to all other life table functions. This approach is straightforward, computationally efficient, and readily extendable to more complex life table functions via matrix algebra, making it a practical baseline for comprehensive uncertainty assessment.

Given $\boldsymbol{Y}$, the vector of Poisson random variables with variance-covariance matrix $\boldsymbol{V_y}$, and a differentiable function $g(\boldsymbol{Y})$, the Delta method approximates the variance of $g(\boldsymbol{Y})$ as

$$\boldsymbol{V}_{g(\boldsymbol{Y})} \approx \boldsymbol{J}_{g(\boldsymbol{Y})}\ \boldsymbol{V_y}\ \boldsymbol{J}'_{g(\boldsymbol{Y})}, \tag{4}$$

where $\boldsymbol{J}_{g(\boldsymbol{Y})} = \partial g(\boldsymbol{Y})/\partial \boldsymbol{Y}$ is the Jacobian of partial derivatives. For the log-rates $\boldsymbol{\eta} = \ln(\boldsymbol{y}/\boldsymbol{e})$, the Jacobian is simply the diagonal matrix of reciprocals:

$$\boldsymbol{J_\eta} = \boldsymbol{D}_{\frac{1}{\boldsymbol{y}}}.$$

Applying the Delta method then gives

$$\boldsymbol{V_\eta} \approx \boldsymbol{J_\eta}\, \boldsymbol{V_y}\, \boldsymbol{J}'_{\boldsymbol{\eta}} = \boldsymbol{D}_{\frac{1}{\boldsymbol{y}}},$$

which remains diagonal due to the independence of age-specific observations. This example illustrates how uncertainty in the observed counts propagates directly to the log-rates, forming the basis for all subsequent life table functions.

It is important to note that directly applying the Delta method can be problematic when $y_i = 0$, since the reciprocal $1/y_i$ is undefined. In practice, software like `R` handles this in the computation of the variance-covariance matrix by replacing zeros with a very small positive number, typically derived from the machine precision and the scale of the non-zero counts. This substitution is purely a numerical safeguard to allow matrix inversion and computation of $\boldsymbol{V_y}$; it does not represent a statistical estimate of the zero counts. These zero-count cases are handled more rigorously by the other two approaches presented in the following which provide a principled treatment of zero deaths and rare events.

2.1.2 Garwood Exact Confidence Limits

One of the earliest approaches for constructing confidence intervals for Poisson-distributed counts is the Garwood exact method (Garwood, 1936), which derives intervals directly from the chi-squared distribution. Unlike asymptotic methods that rely on normal approximations, the Garwood method exploits the exact sampling distribution of the Poisson variable, ensuring finite-sample coverage for the expected number of deaths in each age

interval. Although this approach requires evaluating chi-squared quantiles for each observation and does not yield a closed-form variance–covariance structure, it provides higher accuracy in settings with sparse data or rare events, where asymptotic approximations may fail.

Formally, for the vector of observed death counts $\boldsymbol{y}$, the Garwood exact confidence interval for the Poisson mean at confidence level $1-\alpha$ is

$$(\boldsymbol{y}^{\mathrm{L}}, \boldsymbol{y}^{\mathrm{U}}) = \left(\frac{1}{2}\,\chi^2_{2\boldsymbol{y},\,\alpha/2},\ \frac{1}{2}\,\chi^2_{2(\boldsymbol{y}+1),\,1-\alpha/2} \right)$$

where $\chi^2_{v,p}$ denotes the pth quantile of the chi-squared distribution with v degrees of freedom. By construction, these intervals provide exact coverage for the Poisson mean, including in small-sample or zero-count cases.

A key advantage of the Garwood formulation is that it constructs intervals for the Poisson mean rather than the observed counts. This distinction allows monotonic transformations, such as the logarithm, to be applied directly while preserving the nominal coverage probability. The corresponding finite-sample confidence intervals for the log-rates are thus

$$(\boldsymbol{\eta}^{\mathrm{L}}, \boldsymbol{\eta}^{\mathrm{U}}) = \left(\ln \frac{\boldsymbol{y}^{\mathrm{L}}}{\boldsymbol{e}},\ \ln \frac{\boldsymbol{y}^{\mathrm{U}}}{\boldsymbol{e}} \right)$$

It is important to note that when $y_i = 0$, the lower bound y_i^{L} is zero, so the corresponding log-transformed lower bound becomes $-\infty$. In such cases, adjustments may be applied to the lower bounds, analogous to those used in the Delta method, to enable numerical computation of the confidence intervals for $\boldsymbol{\eta}$. Once the log-rate intervals are obtained, they can be used to construct the variance–covariance matrix of the log-rates and to propagate uncertainty through all subsequent life table functions. This approach provides a rigorous, exact-sample method for quantifying uncertainty without relying on asymptotic approximations.

2.1.3 Bayesian Bootstrap

A third approach for quantifying uncertainty in the log-rates is a Bayesian bootstrap procedure, which provides a coherent probabilistic representation of uncertainty beyond classical analytical approaches. This method benefits from the conjugacy between the Poisson likelihood and a Gamma prior, a property that ensures the posterior distribution of the age-specific Poisson mean remains Gamma and allows for fast, closed-form posterior sampling (Donovan and Mickey, 2019). Specifically, given observed deaths y_i and exposures e_i, and assuming a Gamma prior with shape α and rate β, the posterior distribution of the age-specific death rate m_i at age x_i is

$$m_i \mid y_i, e_i \sim \Gamma(\alpha + y_i,\ \beta + e_i)\,.$$

For each age, S posterior draws ($S = 1000$ in this study) are sampled from this posterior distribution, denoted $m_i^{(s)}$, $s = 1, \ldots, S$, and then transformed to the log-scale to obtain the corresponding posterior draws of the log-rates. Repeating this procedure across all ages produces a set of posterior draws that comprehensively captures uncertainty in the log-rates and can be subsequently propagated through all derived life table functions.

This Bayesian framework allows flexible specification of priors, and the choice of hyperparameters (α, β) is particularly important in small-sample or zero-count situations.

The Gamma prior for the Poisson mean μ_i can be calibrated to reflect different levels of prior information. A commonly adopted specification is a nearly uninformative prior, such as $\Gamma(0.001, 0.001)$, which contributes minimal information relative to the data and allows the posterior to be almost entirely determined by the observed deaths y_i. The mean of this prior is $\alpha/\beta = 1$, but its variance is extremely large (≈ 1000). Consequently, when $y_i = 0$, the posterior shape parameter remains close to zero, yielding a Gamma posterior heavily concentrated near zero. This leads to very small simulated death rates and, in turn, highly negative log-rates $\eta_i = \log(\mu_i)$. In practice, this prior yields posterior intervals closely aligned with those obtained under the Garwood exact formulation, as both reflect the highly skewed uncertainty structure of the Poisson mean in the low-count regime.

Alternatively, a weakly informative prior such as $\Gamma(0.5, 0.001)$ can be used, which corresponds approximately to the Jeffreys prior, proportional to the square root of the Fisher information. This choice provides a principled, invariant form of noninformativeness that stabilizes posterior estimates in age groups with few or zero deaths, while remaining data-dominant when counts are large. Intuitively, this prior introduces mild regularization, preventing the posterior from collapsing toward zero when $y_i = 0$. In this case, the posterior shape parameter becomes 0.5, producing a distribution still concentrated near zero but less extreme, resulting in more realistic log-rate estimates. From a frequentist perspective, this adjustment behaves similarly to adding 0.5 in the Delta method variance approximation, preventing numerical instability in the computation of log-rates and ensuring finite posterior uncertainty across all ages. In the following, we present results obtained using this approximate Jeffreys prior.

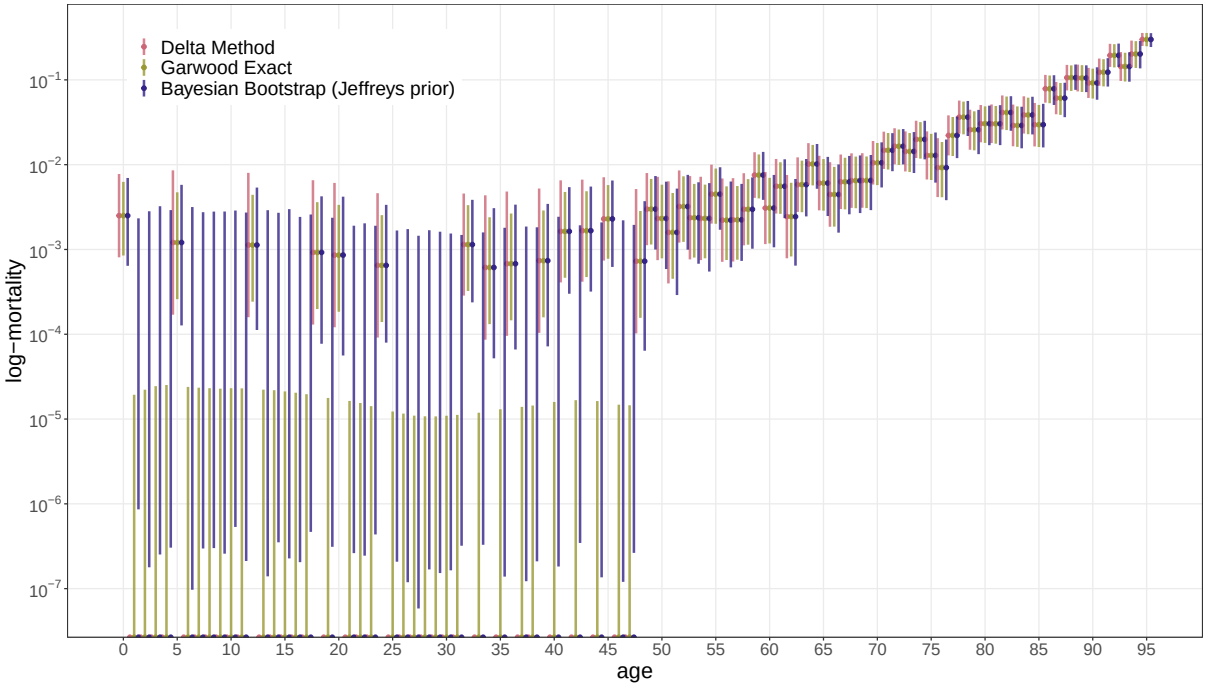

Figure 1: Estimated age-specific log-mortality rates and corresponding 95% confidence intervals obtained using three alternative methods. Data refer to females in the 20th arrondissement of Paris, 2020.

Figure 1 illustrates the estimated log-rates for the 20th arrondissement, which in 2020 had a female population of approximately 100,975 and experienced 722 deaths across

all ages. Across the three methods, the results are highly consistent at ages with many deaths (generally after age 50) and remain broadly similar at ages with positive but low counts (below age 50). In contrast, ages with zero observed deaths exhibit substantial differences between the approaches. Specifically, the Delta method produces $-\infty$ log-rates with undefined variance, the Garwood method provides a finite upper bound but a $-\infty$ lower bound, and the Bayesian bootstrap yields reasonable estimates with fully defined distributions. These posterior draws from the Bayesian bootstrap will be used to construction the variance–covariance matrix for the log-rates ($\boldsymbol{V_\eta}$), which can then be propagated to quantify uncertainty across all life table functions.

2.2 From log-rates to the survival function

In conventional life table construction, once the (log-)mortality rates are estimated, they are converted into age-specific probabilities of death, $\boldsymbol{q}$. These probabilities are then applied to a synthetic cohort (here, assuming a radix of 1 for simplicity) to derive the survival function ($\boldsymbol{l}$), the number of deaths at each age ($\boldsymbol{d}$), person-years lived within each age interval ($\boldsymbol{L}$) and above a given age ($\boldsymbol{T}$), and ultimately life expectancy ($\boldsymbol{e}$). The conversion from mortality rates m_i to probabilities q_i relies on the so-called a_i values, representing the average number of person-years lived within the interval by individuals who die during that age (Preston et al., 2001, p. 43):

$$q_i = \frac{m_i}{1 + (1 - a_i)m_i}. \tag{5}$$

The survival function is typically computed recursively as:

$$l_i = \prod_{j=0}^{i-1}(1 - q_j). \tag{6}$$

Alternatively, a matrix-algebra approach can be introduced, inspired by the continuous formulation:

$$l(x) = \exp\left(- \int_0^x \mu(t)dt \right) \quad \Rightarrow \quad \boldsymbol{l} \approx \exp(\boldsymbol{C}\,\boldsymbol{m}) = \exp(\boldsymbol{C}\,\exp(\boldsymbol{\eta})), \tag{7}$$

where $\boldsymbol{C}$ is a lower-triangular $n \times n$ matrix with entries -1, which effectively accumulates the age-specific mortality rates. It is important to note that, in a discrete-age setting, this matrix-based approximation is not equivalent to the recursive formula in (6). Nevertheless, it is valuable because it establishes a framework for computing the survival function and all subsequent life table quantities in a purely matrix-algebra fashion, which is particularly convenient for uncertainty propagation.

To align the matrix-algebra approximation with discrete-age data, the mortality rates are adjusted to $\tilde{\boldsymbol{m}}$ so that the matrix-exponentiated survival function reproduces the results from the conventional recursive computation. Specifically, for a given age x, the two formulations can be expressed side by side as

$$\begin{aligned} l_x &= \exp(-\tilde{m}_0 - \tilde{m}_1 - \cdots - \tilde{m}_{x-1}), \\ l_x &= (1 - q_0) \cdot (1 - q_1) \cdot \cdots \cdot (1 - q_{x-1}). \end{aligned}$$

Taking the natural logarithm of both right-hand-sides, consistency requires that, for each age x, the adjusted rate satisfies

$$-\tilde{m}_x = \ln(1 - q_x).$$

Substituting the age-specific rate-to-probability relationship from (5) yields

$$-\tilde{m}_x = \ln(1 - a_x m_x) - \ln(1 + m_x - a_x m_x),$$

which does not simplify to a closed-form expression for $\tilde{m}_x$ in terms of the observed rates m_x and the a_x values.

To obtain an explicit and tractable approximation, we apply the first three terms of the Taylor expansion to both logarithmic components, resulting in the adjusted mortality rates:

$$\tilde{m}_x \approx m_x + \frac{a_x^2 m_x^2}{2} + \frac{a_x^3 m_x^3}{3} - \frac{(m_x - a_x m_x)^2}{2} + \frac{(m_x - a_x m_x)^3}{3}. \tag{8}$$

This expression defines the necessary correction to the original mortality rates to ensure consistency between the matrix-based and recursive formulations of the survival function. Even when limited to the first three terms, the approximation error remains negligible, below approximately 0.001% in the example of the 20th arrondissement of Paris. The resulting adjusted rates enable a direct, matrix-based computation of the survival function $\boldsymbol{l}$, which can then be extended to all subsequent life table functions ($\boldsymbol{d}$, $\boldsymbol{L}$, $\boldsymbol{T}$, $\boldsymbol{e}$). Although we use the adjusted rates $\tilde{\boldsymbol{m}}$ in these computations, this deterministic, higher-order correction is very small and does not introduce additional variability; therefore, its impact on the variance-covariance matrix and the propagation of uncertainty through the life table functions can be safely neglected.

In terms of uncertainty quantification, the variance–covariance matrix of the survival function $\boldsymbol{l}$ can be obtained via the Delta method, as expressed in (4). Given the variance–covariance matrix of the log-rates, $\boldsymbol{V}_{\boldsymbol{\eta}}$, and the Jacobian of the survival function with respect to $\boldsymbol{\eta}$ derived from (7), we have:

$$\boldsymbol{J}_{\boldsymbol{l}} = \frac{\partial \boldsymbol{l}}{\partial \boldsymbol{\eta}} = \frac{\partial \exp(\boldsymbol{C}\, \exp(\boldsymbol{\eta}))}{\partial \boldsymbol{\eta}} = \boldsymbol{D}_{\exp(\boldsymbol{C}\exp(\boldsymbol{\eta}))}\, \boldsymbol{C} \boldsymbol{D}_{\exp(\boldsymbol{\eta})} = \boldsymbol{D}_{\boldsymbol{l}}\, \boldsymbol{C}\, \boldsymbol{D}_{\boldsymbol{m}}\,.$$

The approximate variance–covariance matrix of $\boldsymbol{l}$ is then obtained as:

$$\boldsymbol{V}_{\boldsymbol{l}} \approx \boldsymbol{J}_{\boldsymbol{l}}\, \boldsymbol{V}_{\boldsymbol{\eta}}\, \boldsymbol{J}_{\boldsymbol{l}}'\,.$$

Standard errors for the survival function can be directly obtained as the square roots of the diagonal elements of $\boldsymbol{V}_{\boldsymbol{l}}$, and under the usual asymptotic normality assumption, 95% confidence intervals can be constructed using the corresponding standard normal quantile.

Figure 2 presents the estimated survival functions for two arrondissements in Paris. These survival curves were constructed using the $\boldsymbol{a}$ values for French females from the 2020 life table provided by the Human Mortality Database (2025). The $\boldsymbol{a}$ values are set to 0.5 for all ages except for age 0, where $a_1 = 0.14$, and for the open-ended last age group, the standard approximation $a_n = 1/m_n$ was applied. Although different conventions exist for handling the open-ended last age group (see discussions in the epidemiological literature (Eayres and Williams, 2004; Silcocks, 2004) and in the demographic community (Lo et al., 2016)), such choices may slightly alter the resulting survival curves but do not affect the underlying methodology described above. Moreover,these $\boldsymbol{a}$ values will be used consistently in all subsequent computations.

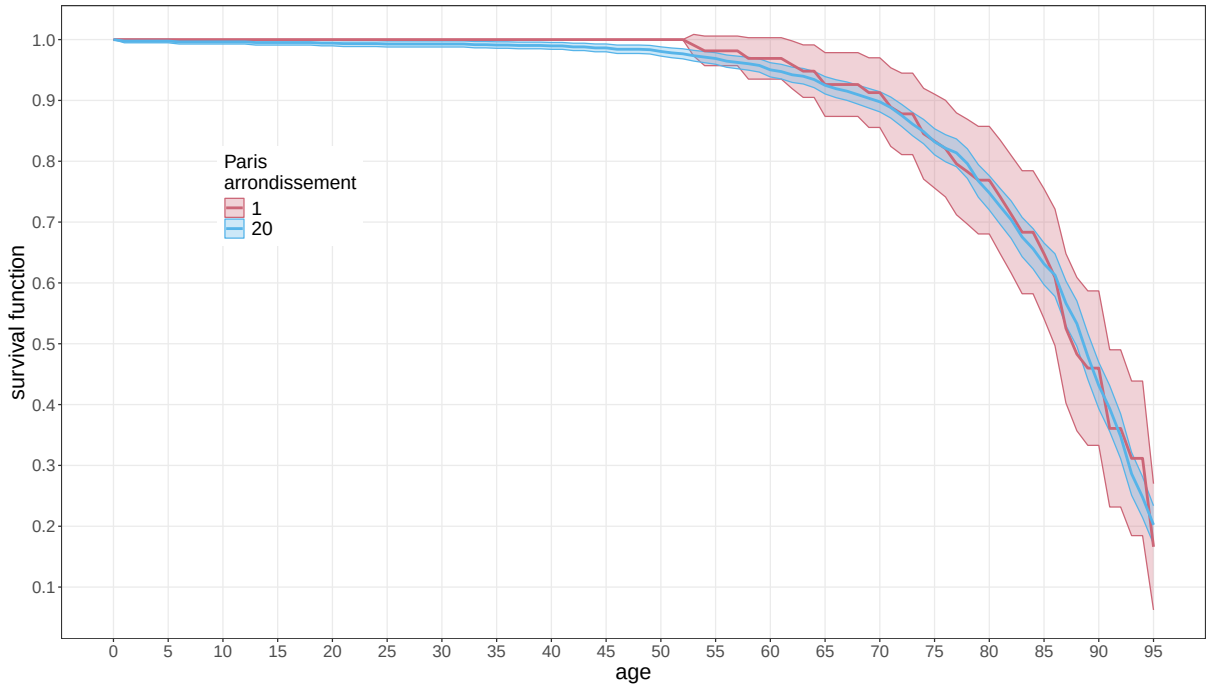

Figure 2: Estimated age-specific survival functions and corresponding 95% confidence intervals for both 1st and 20th arrondissement of Paris, females, 2020.

Alongside the estimated survival functions, Figure 2 also shows the corresponding 95% confidence intervals for the two arrondissements. Although the arrondissements differ in population size, reflected in the width of the confidence intervals, the overall shapes of the survival curves are similar. Notably, the survival function for the first arrondissement falls within the confidence interval of the 20th arrondissement only from approximately age 65 onward.

2.3 Extension to other life table functions

We now extend the framework to the remaining life table functions. Let us define the $n \times n$ matrix $\boldsymbol{E_d}$ as

$$\boldsymbol{E_d} = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & -1 \\ 0 & \cdots & 0 & 0 & 1 \end{bmatrix}.$$

This matrix computes first-order differences between adjacent ages. Using it, life table deaths can be expressed compactly as

$$\boldsymbol{d} = \boldsymbol{E_d}\,\boldsymbol{l}\,.$$

In other words, $\boldsymbol{E_d}$ transforms the vector of survivors $\boldsymbol{l}$ into the vector of deaths $\boldsymbol{d}$ by differencing consecutive entries.

The corresponding variance-covariance matrix for $\boldsymbol{d}$ can be derived via the Delta method as

$$\boldsymbol{V_d} \approx \boldsymbol{J_d}\,\boldsymbol{V_\eta}\,\boldsymbol{J_d'}\,,$$

where the Jacobian of partial derivatives is given by

$$\boldsymbol{J_d} = \frac{\partial \boldsymbol{d}}{\partial \boldsymbol{\eta}} = \frac{\partial (\boldsymbol{E_d l})}{\partial \boldsymbol{\eta}} = \boldsymbol{E_d}\, \boldsymbol{J_l}.$$

Similarly, we can represent the person-years lived within each age interval ($\boldsymbol{L}$) and the total person-years lived above a given age ($\boldsymbol{T}$) as linear transformations of the survival function:

$$\boldsymbol{L} = \boldsymbol{E_L}\, \boldsymbol{l} \qquad \text{and} \qquad \boldsymbol{T} = \boldsymbol{E_T}\, \boldsymbol{E_L}\, \boldsymbol{l},$$

where the matrices $\boldsymbol{E_L}$ and $\boldsymbol{E_T}$ are defined as

$$\boldsymbol{E_L} = \begin{bmatrix} a_1 & 1-a_1 & 0 & \cdots & 0 \\ 0 & a_2 & 1-a_2 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & a_{n-1} & 1-a_{n-1} \\ 0 & \cdots & 0 & 0 & a_n \end{bmatrix}, \qquad \boldsymbol{E_T} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 0 & 1 & 1 & \cdots & 1 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 1 \\ 0 & \cdots & 0 & 0 & 1 \end{bmatrix}.$$

The matrix $\boldsymbol{E_L}$ reproduces, in matrix form, the conventional expression linking person-years lived and survivors (Preston et al., 2001, p. 50):

$$L_i = l_{i+1} + a_i d_i = l_{i+1} + a_i(l_i - l_{i+1}) = a_i l_i + (1-a_i) l_{i+1}.$$

The matrix $\boldsymbol{E_T}$, on the other hand, performs the retro-cumulative summation required to compute the total person-years lived above each age.

Finally, the variance-covariance matrices of $\boldsymbol{L}$ and $\boldsymbol{T}$ follow the same general formulation as for $\boldsymbol{d}$:

$$\boldsymbol{J_L} = \frac{\partial \boldsymbol{L}}{\partial \boldsymbol{\eta}} = \boldsymbol{E_L}\, \boldsymbol{J_l}, \qquad \boldsymbol{J_T} = \frac{\partial \boldsymbol{T}}{\partial \boldsymbol{\eta}} = \boldsymbol{E_T}\, \boldsymbol{E_L}\, \boldsymbol{J_l},$$

with the corresponding variances approximated as

$$\boldsymbol{V_L} \approx \boldsymbol{J_L}\, \boldsymbol{V_\eta}\, \boldsymbol{J_L'}, \qquad \boldsymbol{V_T} \approx \boldsymbol{J_T}\, \boldsymbol{V_\eta}\, \boldsymbol{J_T'}.$$

Finally, we consider the computation of the life expectancy vector, $\boldsymbol{e}$, which quantifies the expected remaining lifetime for individuals who have reached each age:

$$\boldsymbol{e} = \frac{\boldsymbol{T}}{\boldsymbol{l}} = \boldsymbol{D}_{1/\boldsymbol{l}}\, \boldsymbol{T}.$$

The first expression corresponds to the conventional definition of life expectancy as the age-specific ratio between the total person-years lived above a given age and the number of survivors at that age, while the second expression is introduced for notational convenience, as it facilitates the derivation of the Jacobian required for variance propagation.

Specifically the partial derivatives of $\boldsymbol{e}$ with respect to $\boldsymbol{\eta}$ follows by the product and chain rule:

$$\boldsymbol{J_e} = \frac{\partial \boldsymbol{e}}{\partial \boldsymbol{\eta}} = \frac{\partial (\boldsymbol{D}_{1/\boldsymbol{l}}\, \boldsymbol{T})}{\partial \boldsymbol{\eta}} = \boldsymbol{D}_{1/\boldsymbol{l}}\, \boldsymbol{J_T} + \boldsymbol{T} \frac{\partial \boldsymbol{D}_{1/\boldsymbol{l}}}{\partial \boldsymbol{\eta}} = \boldsymbol{D}_{1/\boldsymbol{l}}\, \boldsymbol{J_T} - \boldsymbol{D}_{\boldsymbol{T}/\boldsymbol{l}^2}\, \boldsymbol{J_l}.$$

The variance-covariance matrix of $\boldsymbol{e}$ is then approximated by the Delta method as

$$\boldsymbol{V_e} \approx \boldsymbol{J_e}\, \boldsymbol{V_\eta}\, \boldsymbol{J_e'}.$$

Taken together, these derivations demonstrate that the matrix-based formulation offers a unified, algebraic framework to compute all key life table functions and their associated uncertainty directly from the estimated log-rates, enabling consistent, transparent, and computationally efficient variance propagation.

3 Results

Figure 3 presents four key life table functions across age, together with their 95% confidence intervals, for females in the 1st and 20th arrondissements of Paris in 2020. Although the two districts differ substantially in population size, reflected in the width of their confidence intervals, the overall patterns are largely comparable. Life expectancy in the 20th arrondissement, which is larger and more demographically heterogeneous, is consistently lower than in the 1st arrondissement, though it remains within the latter's confidence bounds. The characteristic bell-shaped pattern of the life table deaths ($\boldsymbol{d}$) is not clearly visible, likely due to the small population size and the relatively low open-age interval (95+), which is limited given the high survival levels observed in 2020 and more generally in recent years. The impact of this upper age truncation is particularly evident in the final odd value of the person-years function ($\boldsymbol{L}$) and the last value of life expectancy ($\boldsymbol{e}$) for the smaller arrondissement, where both show sharper increase than would be expected with a higher open-age threshold.

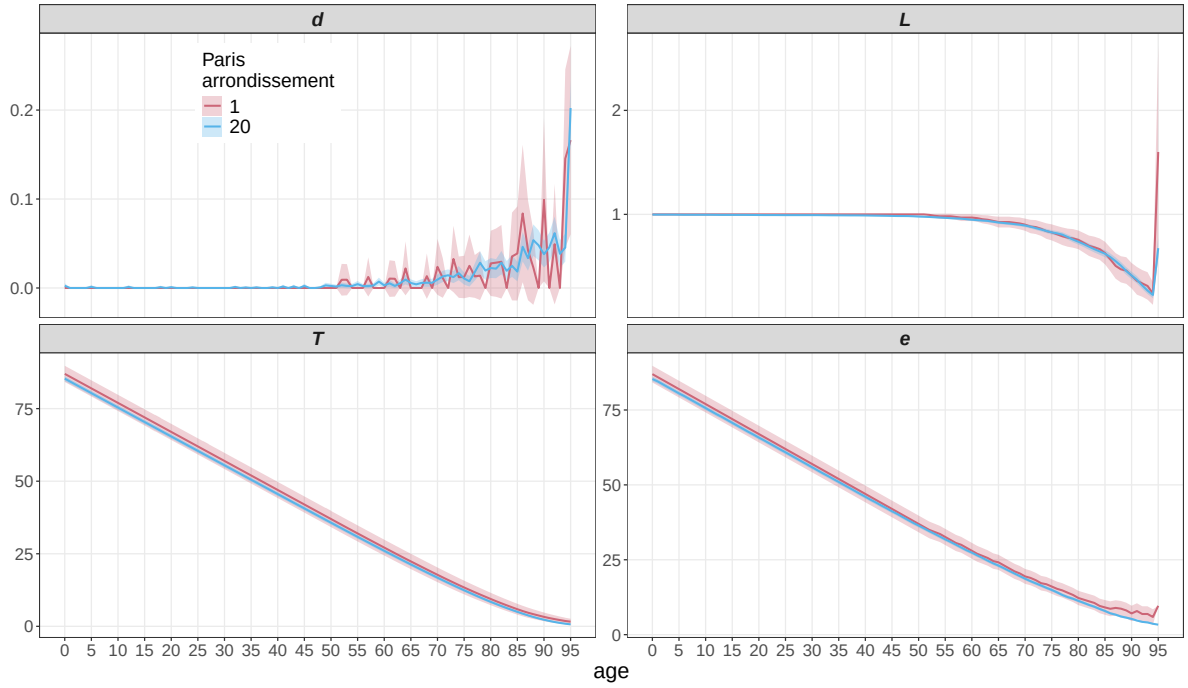

Figure 3: Estimated age-specific life-table deaths ($\boldsymbol{d}$), person-years lived within each age interval ($\boldsymbol{L}$), person-years lived above a given age ($\boldsymbol{T}$), and life expectancy ($\boldsymbol{e}$), with corresponding 95% confidence intervals for each measure, for females in the 1st and 20th arrondissements of Paris, 2020.

Figure 4 presents the estimated female life expectancy at birth in 2020 across the 20 arrondissements of Paris, together with their associated 95% confidence intervals. The estimated levels reveal a substantial intra-urban disparity, with a gap of 6.59 years between the highest and lowest values. The highest life expectancies are observed in the 3rd arrondissement (90.3 years) and the 6th arrondissement (90.0 years), while the lowest are found in the 19th (83.7 years) and 18th (84.9 years). The width of the confidence intervals varies considerably across arrondissements and is largely driven by differences in population size—smaller populations produce wider uncertainty bands. In extreme cases, such as the 2nd arrondissement, the 95% confidence interval spans up to 6.6 years.

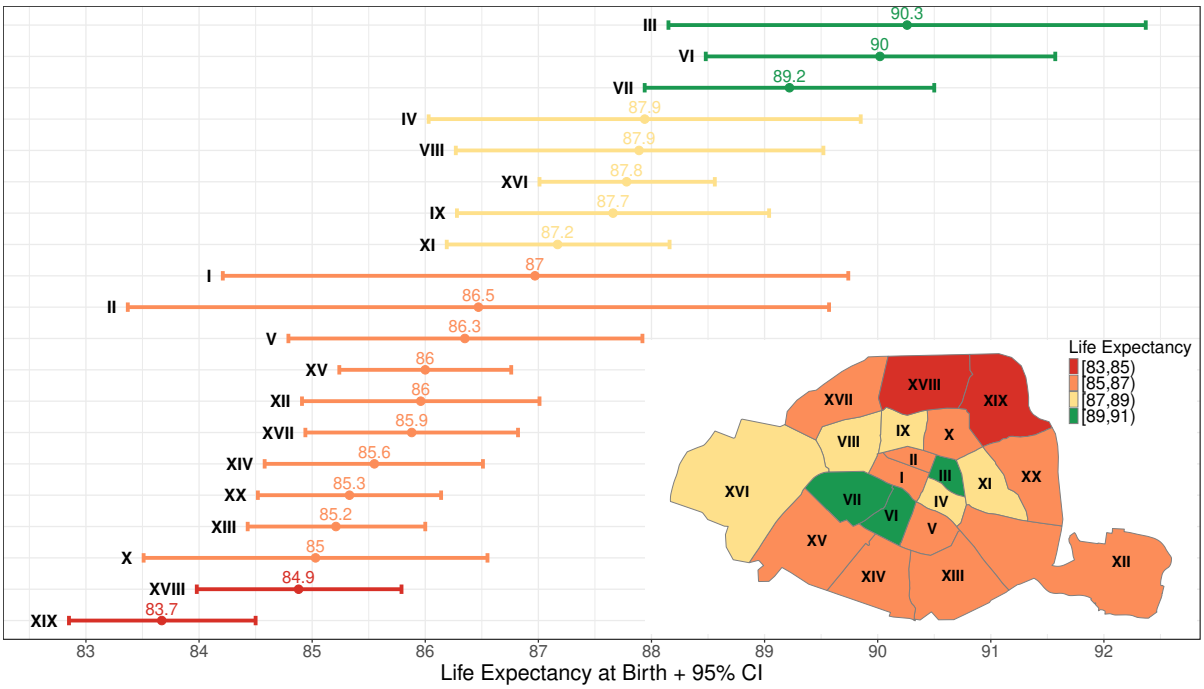

Figure 4: Estimated female life expectancy at birth in 2020 across the 20 arrondissements of Paris, with 95% confidence intervals. The inset map shows the spatial distribution of the arrondissements. Colors in both the bar plot and the map correspond to the levels of life expectancy at birth.

These uncertainty bounds also allow for formal comparisons of mortality levels across arrondissements. When confidence intervals do not overlap, the corresponding differences in life expectancy can be interpreted as statistically significant. For instance, the 19th arrondissement exhibits a life expectancy significantly lower than that of nearly all other arrondissements, except the 18th, 10th, and 13th—whose mortality levels are similar—and the 1st and 2nd arrondissements, where wide confidence intervals preclude precise inference. The inset map in Figure 4 further illustrates the spatial pattern of mortality in 2020, revealing a pronounced east–west and south–north gradient: central, western, and southern arrondissements display the highest longevity, whereas eastern and northern areas experience comparatively lower life expectancy.

4 Outlook

The methodological framework presented in this study establishes a unified, matrix-based approach for deriving life table functions and their associated uncertainties from estimated log-mortality rates. Beyond its immediate application to mortality analysis in small populations, this framework opens several promising directions for future research. A natural next step will involve extending uncertainty quantification beyond the standard life table functions to indicators of lifespan variability—such as the standard deviation of ages at death, life disparity ($\boldsymbol{e}^{\dagger}$), and the Gini coefficient of longevity. These measures capture essential dimensions of inequality in mortality, yet their associated uncertainty is often overlooked—even though it is fundamental to understanding demographic heterogeneity.

Further work will generalize the proposed framework to abridged life tables, which

are the most common data format available for small-area mortality estimation. This adaptation will require reformulating the relevant matrices to accommodate broader age intervals while preserving internal consistency of uncertainty propagation. A systematic investigation will also be undertaken into the sensitivity of uncertainty estimates to the choice of $\boldsymbol{a}$, particularly for the open-ended last age group and for cohort life tables. Such analyses will clarify how assumptions regarding exposure within age intervals influence both point estimates and their variability.

Another important extension will focus on quantifying the uncertainty of differences between two populations, such as the comparison of the 1st and 20th arrondissements presented here. By jointly modeling the covariance structure of both populations, the framework can be used to formally assess whether observed differences in any life table function, at any age, are statistically significant. Moreover, the presented approach naturally lends itself to model-based applications. Since most mortality models are estimated on the log-rate scale, the same procedure can be directly applied when a (possibly approximate) variance–covariance matrix of $\boldsymbol{\eta}$ is available from model output. This opens the possibility of propagating model-based uncertainty across both age and time dimensions in two-dimensional mortality models.

A further extension will explore how the same framework can be applied in model-based settings, where mortality is typically modeled on the log-rate scale under a Poisson framework. In such cases, the approach can be directly implemented using an estimated or approximate variance–covariance matrix of $\boldsymbol{\eta}$ obtained from the fitted model, allowing full propagation of model-based uncertainty through all life table functions. Because statistical models introduce smoothness and structural dependence across age, time, or population, we expect the resulting uncertainty to be reduced and more stable than that obtained from purely empirical estimates. Building on this, we plan to investigate joint modeling of multiple populations, which would enable the explicit estimation of cross-population covariance structures and provide a coherent basis for assessing whether observed differences in life table functions—such as those between the 1st and 20th arrondissements—are statistically significant.

Finally, the explicit Jacobians derived for each life table function offer a powerful analytical framework for sensitivity analysis, enabling researchers to quantify how changes in mortality at specific ages propagate through to all downstream demographic quantities. Although this study focused on mortality, the same algebraic structure can be readily extended to any demographic process formulated within a life table framework, including fertility, morbidity, and migration. The forthcoming research, to be presented at the European Population Conference, will expand on these developments to deliver a unified and computationally efficient toolkit for the analytical and probabilistic study of demographic functions, treating them explicitly as stochastic processes rather than deterministic summaries.

References

Andreev, E. M. and V. M. Shkolnikov (2010). Spreadsheet for calculation of confidence limits for any life table or healthy-life table quantity. techreport TR-2010-005, MPIDR Technical Report. 4 pages.

Chiang, C. L. (1960a). A Stochastic Study of the Life Table and Its Applications: I. Probability Distributions of the Biometric Functions. *Biometrics 16*(4), 618–635.

Chiang, C. L. (1960b). A Stochastic Study of the Life Table and Its Applications. II. Sample Variance of the Observed Expectation of Life and other Biometric Functions. *Human Biology 32*(3), 221–238.

Chiang, C. L. (1961). A Stochastic Study of the Life Table and Its Applications. III. The Follow-Up Study with the Consideration of Competing Risks. *Biometrics 17*(1), 57–78.

Donovan, T. M. and R. M. Mickey (2019). The Shark Attack Problem: The Gamma-Poisson Conjugate. In *Bayesian Statistics for Beginners: a step-by-step approach*. Oxford University Press.

Eayres, D. and E. S. Williams (2004). Evaluation of methodologies for small area life expectancy estimation. *Journal of Epidemiology and Community Health 58*(3), 243–249.

Friedman, M. (1982). Piecewise Exponential Models for Survival Data with Covariates. *The Annals of Statistics 10*, 101–113.

Garwood, F. (1936). Fiducial Limits for the Poisson Distribution. *Biometrika 28*(3-4), 437–442.

Graunt, J. (1662). *Natural and Political Observations Made upon the Bills of Mortality*. London: American edition edited by Walter F. Willcox, John Hopkins Press, Baltimore, 1939.

Holford, T. R. (1980). The analysis of rates and survivorship using log-linear models. *Biometrics 36*, 299–305.

Human Mortality Database (2025). *Max Planck Institute for Demographic Research (MPIDR, Germany), University of California, Berkeley (USA), and French Institute for Demographic Studies (INED, France)*. Available at `www.mortality.org`. Data downloaded on October 2025.

Keyfitz, N. (1966). Sampling Variance of Standardized Mortality Rates. *Human Biology 38*, 309–317.

Laird, N. and D. Olivier (1981). Covariance analysis of censored survival data using log-linear analysis techniques. *Journal of the American Statistical Association 76*, 231–240.

Lo, E., R. Bourbeau, D. Vatnik, and A. Benedetti (2016). Variance models of the last age interval and their impact on life expectancy at subnational scales. *Demographic Research 35*(15), 399–454.

Preston, S. H., P. Heuveline, and M. Guillot (2001). *Demography. Measuring and Modeling Population Processes*. Blackwell.

Scherbov, S. and D. Ediev (2011). Significance of life table estimates for small populations: Simulation-based study of estimation errors. *Demographic Research 24*(22), 527–550.

Silcocks, P. B. (2004). Improving estimation of the variance of expectation of life for small populations. *Journal of Epidemiology and Community Health 58*(7), 611–612.

Silcocks, P. B., D. A. Jenner, and R. Reza (2001). Life expectancy as a summary of mortality in a population: statistical considerations and suitability for use by health authorities. *Journal of Epidemiology and Community Health 55*(1), 38–43.

Wilmoth, J., K. Andreev, D. Jdanov, D. Glei, and T. Riffe (2021). Methods protocol for the human mortality database. version 6. Technical report, University of California, Berkeley, and Max Planck Institute for Demographic Research, Rostock. With the assistance of C. Boe, M. Bubenheim, D. Philipov, V. Shkolnikov, P. Vachon, C. Winant, M. Barbieri. Available at https://www.mortality.org/File/GetDocument/Public/Docs/MethodsProtocolV6.pdf. Retrieved on October 28, 2025.