On the smoothing of death curves using mixtures of probability distributions

Lucia Zanotto, Adelchi Azzalini

1 Introduction

To study and analyze mortality, models are frequently employed because they help to understand the characteristics and the evolution of the phenomenon. Even if non parametric models allow more accurate data fitting, parametric ones have the advantage of facilitating interpretation, comparison and forecasting (Congdon, 1993). In particular, the estimated parameters can be used as indicators of mortality pattern and employed to quantify the differences among groups of individuals. Moreover, the trends of the computed parameter can be examined in order to follow recent transformations and to predict future (or past) mortality scenarios (Canudas-Romo et al., 2018). Most of the models provided in the literature are mathematical functions which fit the death rates (Gompertz, 1825; Kannisto, 1994; Makeham, 1860; Siler, 1979) or the odds ratio of probability of dying (Heligman and Pollard, 1980).

Recently more attention has been focused on the distribution of deaths by age (Canudas-Romo, 2010; Cheung et al., 2005; Van Raalte and Caswell, 2013; Wilmoth and Horiuchi, 1999), which provides key insights on longevity and lifespan variability (Basellini and Camarda, 2016) and it can be approximated by probability distributions because it has the advantage of being a density function (Mazzuco et al., 2018). In this framework, Zanotto et al. (2017) proposed a mixture model, which fits the entire distribution of deaths by age. The model is inspired by Pearson's theory about mortality components (Pearson, 1897) and it is a combination of three distributions, which fit infant, premature and adult mortality, respectively. The shape of the the model is very flexible so it can be applied to several mortality schedules with satisfactory results.

However the estimation of the parameters can be problematic because of computational issues, in particular due to identification problems, producing local irregularities in the trends of the estimated parameters. Also when period data are analyzed instead of cohort ones, raw fluctuations in the coefficient evolutions are not appropriate if they are not justified by exceptional events (as for instance wars) because mortality changes slowly. Since the time evolution of the estimated parameters is very useful to identify paths, to formulate hypothesis and explanations, and to reach conclusions, it is convenient to enforce regular trends, which are clearly easier to interpret. To this end, one approach is to consider the parameters of the mixture model as time-related functions. So, instead of estimating the parameters of the model each year separately, the coefficients of the time-related functions are calculated using the deaths in the entire time series. The value of a mixture parameter for a single year is easily obtained combining the coefficients of its time-related function and the year of interest. If the shape of the selected time-dependent functions are sufficiently regular, the trends of the mixture parameters are automatically smooth, without unreasonable fluctuations and irregularities. Moreover, even if the parameters of the mixture model are calculated from the time-related functions and not straight estimated, the fit of the model is still appropriate: the estimated curves are close to the ones computed year by year separately, and, in some cases, the adaptation is better because gaps and not admissible values are excluded.

2 Data

The above-mentioned problem of irregular trends is particularly evident when the parameters of the mixture model are estimated using the death distributions calculated on male input data (Population size and Deaths) of USA. Actually, life tables computed starting from raw period data show rough fluctuations presumably caused by the fact that the reference population is a fake cohort and no adjustment is made. The estimation process is therefore complex because of the presence of local maxima, which result in a more irregularity of the parameters' trends especially for those having estimation problems. Moreover, the last open age class between 1959-1980 and 2000-2009 is 85+, while for the other years it is 100+. Especially for the time-window 2000-2009, the adult mode of the death curve is not clearly visible in the data. This generates several issues in the maximization of the likelihood because the values of the coefficients related to adult mortality are outside the possible range. These two reasons lead to select male data of this country as a good example of problematic parameters' trends, so that the advantages of smoothing techniques applied to obtain regular trajectories, can be immediately evident.

3 Method

In the life table, the distribution of deaths by age can be seen as a probability density function. For this reason, Pearson (1897) proposed a mixture of distributions with different shapes and characteristics to approximate the death curve. Following his idea, a three-component mixture model has been introduced by Zanotto et al. (2017), who consider the whole distribution of deaths made up ot three types of mortality: infant, premature and adult. To approximate the first part of the curve referring to infant deaths, an Half Normal distribution was suggested, with its scale parameter fixed and equals to 1 to capture deaths at age 0, even when they are only a few. The asymmetric shape of the adult mortality was estimated with a Skew Normal distribution, introduced by Azzalini (1985). Another Skew Normal was employed to fit the central part of the curve (accidental and premature deceases), which can assume several patterns, depending on the historical period and the country. The three selected distributions are then

$$
\begin{aligned}
f_I(x) &= \overbrace{\frac{\sqrt{2}}{\pi}\exp(-x^2)}^{\text{Infant mortality}} \qquad (x \geq 0)\,, && (1)\\
f_m(x;\xi_m,\omega_m,\lambda_m) &= \overbrace{\frac{2}{\omega_m}\phi\left(\frac{x-\xi_m}{\omega_m}\right)\Phi\left(\lambda_m\frac{x-\xi_m}{\omega_m}\right)}^{\text{Premature mortality}} \qquad (x \in \mathbb{R})\,, && (2)\\
f_M(x;\xi_M,\omega_M,\lambda_M) &= \overbrace{\frac{2}{\omega_M}\phi\left(\frac{x-\xi_M}{\omega_M}\right)\Phi\left(\lambda_M\frac{x-\xi_M}{\omega_M}\right)}^{\text{Adult mortality}} \qquad (x \in \mathbb{R})\,, && (3)
\end{aligned}
$$

with ξ_m and $\xi_M \in \mathbb{R}$, ω_m and $\omega_M \in \mathbb{R}^+$, λ_m and $\lambda_M \in \mathbb{R}$. Combining these three distributions with two mixture parameters $\eta \in [0,1]$ and $\alpha \in [0,1]$, which indicate the probability of infant and adult deaths, respectively, a model with eight coefficients was obtained:

$$
\begin{aligned}
\delta(x,\theta) = \eta \cdot f_I(x) &\\
&+ (1-\eta)\cdot\alpha\cdot f_m(x;\xi_m,\omega_m,\lambda_m) \\
&+ (1-\eta)\cdot(1-\alpha)\cdot f_M(x;\xi_M,\omega_M,\lambda_M),
\end{aligned}
\tag{4}
$$

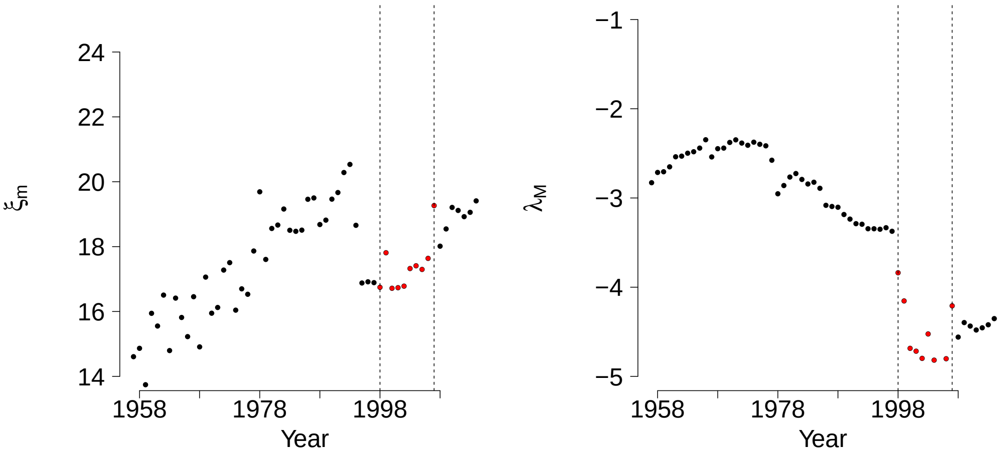
where $\theta = (\eta,\alpha,\xi_m,\omega_m,\lambda_m,\xi_M,\omega_M,\lambda_M)$. Equation (4) is an improper distribution because the support of the Skew Normals is defined also for $\mathbb{R}^-$, while the death curve is only positive-valued. However, the probability mass for ages $x < 0$ is negligible. To estimate the vector θ, the maximization of the likelihood is required, but the function can not derive directly from the model in equation (4) because deaths in the life tables are grouped into age intervals $(x, x+1)$:

$$
d_x(\theta) = \int_x^{x+1} \delta(u;\theta)\, du. \tag{5}
$$

Thus, the parameter values θ can be computed using the likelihood function of a multinomial distribution, which models the probability of the number of deaths occurring in the age interval $(x, x+1)$

$$L(\theta) = \prod_{x=0}^{\Omega} d_x(\theta)^{D_x}, \tag{6}$$

where D_x are the real death counts in $(x, x+1)$ and Ω is the highest attained age at death. For each year, the model parameters are estimated maximizing equation (6), obtaining vectors of 8 values. In most cases, the parameter trends are regular and smooth, but there is a set of situations where the coefficients exhibit non-negligible irregularity. Since mortality changes slowly, raw fluctuations in the coefficients' paths are not appropriate if they are not justified by exceptional events. Moreover, in these cases, the standard numerical optimization algorithms maximizing the likelihood function are often not able to identify the global maximum. As example, the trends of two parameters are reported in Figure 1. The path of the coefficient related to the mode of premature mortality, ξ_m, is very floating during all the period and it is also partially affected by the truncation of the data at age 85+ between years 2000-2009, where most of the red points seem to have a smaller value than the expected one. Moreover there is a sharp change between years 1995 and 1997, where the value of the coefficient goes from 20.5 to 16.8 without any proper explanation. The parameter λ_M, which indicates the skewness of adult component, shows a smooth trend except when the last open age class is 85+: in these years all the points estimated are too small. This results in highly asymmetrical curves, incompatible with the distribution of deaths by



(a) (b)

Figure 1: Trends of two different parameters of the mixture model estimated maximizing the likelihood function year by year separately.

age.

Information regarding past and future need to be taken into account in order to preserve regularity along time. However, this is not possible by estimating the parameters θ for each year separately from the other years. In order to ensure regular trends, we consider a different route where every coefficient of the mixture is expressed as a function of time t:

$$\theta_i^{(t)} = f(t; \psi^{(i)}), \tag{7}$$

where $i = 1, \ldots, 8$ denotes the parameter of the mixture mortality model, ψ is a vector including all the parameters of the time-dependent functions and $\psi^{(i)}$ indicates the coefficients of the time-related function specific for the parameter i.

A practical example, which can also clarify the smoothing technique, is provided below. To select the function form to assign to the trends of the θ parameters of the USA between 1959 and 2016, their evolution in the chosen period was observed. For $\xi_m, \omega_m, \lambda_m, \omega_M$ and η a polynomial of second degree was set, for ξ_M and α a linear regression was enough, while a polynomial of third degree was fixed for λ_M:

$$
\begin{aligned}
\operatorname{logit}\left(\eta^{(t)}\right) &= \eta_0 + \eta_1 \cdot t + \eta_2 \cdot t^2, & \operatorname{logit}\left(\alpha^{(t)}\right) &= \alpha_0 + \alpha_1 \cdot t, \\
\xi_m^{(t)} &= \xi_{m0} + \xi_{m1} \cdot t + \xi_{m2} \cdot t^2, & \xi_M^{(t)} &= \xi_{M0} + \xi_{M1} \cdot t, \\
\log\left(\omega_m^{(t)}\right) &= \omega_{m0} + \omega_{m1} \cdot t + \omega_{m2} \cdot t^2, & \log\left(\omega_M^{(t)}\right) &= \omega_{M0} + \omega_{M1} \cdot t + \omega_{M2} \cdot t^2, \\
\lambda_m^{(t)} &= \lambda_{m0} + \lambda_{m1} \cdot t + \lambda_{m2} \cdot t^2, & \lambda_M^{(t)} &= \lambda_{M0} + \lambda_{M1} \cdot t + \lambda_{M2} \cdot t^2 + \lambda_{M3} \cdot t^3,
\end{aligned}
\tag{8}
$$

where $t \in [1958, 2016]$. In the specific case of USA data, ψ is a vector of 23 coefficients:

$$
\begin{aligned}
\psi = (&\eta_0, \eta_1, \eta_2, \alpha_0, \alpha_1, \\
&\xi_{m0}, \xi_{m1}, \xi_{m2}, \omega_{m0}, \omega_{m1}, \omega_{m2}, \lambda_{m0}, \lambda_{m1}, \lambda_{m2} \\
&\xi_{M0}, \xi_{M1}, \omega_{M0}, \omega_{M1}, \omega_{M2}, \lambda_{M0}, \lambda_{M1}, \lambda_{M2}, \lambda_{M2}),
\end{aligned}
\tag{9}
$$

so that, when $\theta_i = \xi_m$ the corresponding $\psi^{(i)}$ is the vector $(\xi_{m0}, \xi_{m1}, \xi_{m2})$.

The estimation of value of the time-related coefficients ψ is provided considering a comprehensive procedure which embraces all years of the given population, therefore the new likelihood to maximize is the following:

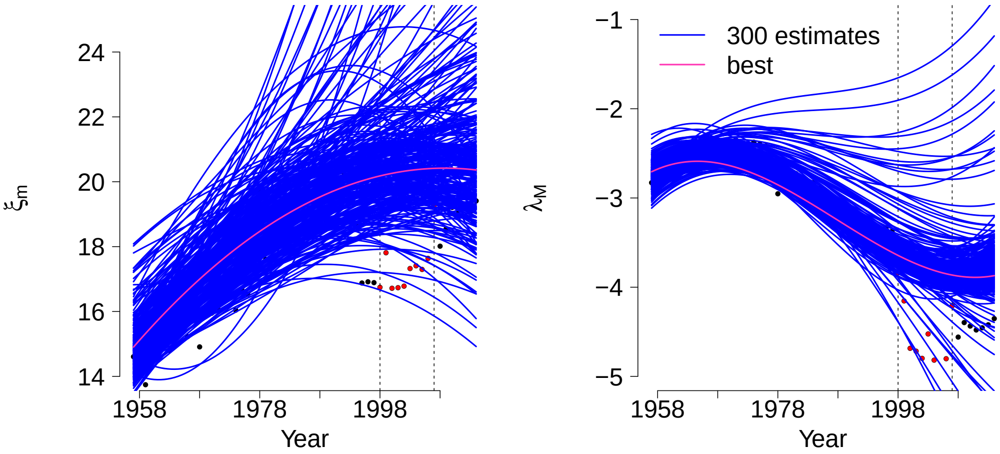
$$L^*(\psi) = \prod_t L\left(\theta^{(t)}\right), \tag{10}$$

where $L(\cdot)$ refers to equation (6) and $\theta^{(t)} = (\theta_1^{(t)}, \theta_2^{(t)}, \ldots, \theta_8^{(t)})$ is a vector of 8 parameters computed using equation (7) and the year t. Instead of evaluating θ, the vector of eight model parameters, separately at each year, equation (10) provides directly the coefficient of the parameters' trends ψ. The vector of parameters to specify the shape of mixture model (4) for the year t ($\theta^{(t)}$) can be derived from ψ. The smoothing strategy, here suggested, has the advantage of reducing the number of parameters to estimate: only the time coefficients ψ need to be computed, instead of the eight values θ for each period. Moreover, no more than one maximization is required, since all the curves are calculated starting from the chosen functions of time: the combination of time and time-dependent coefficients ψ arise a vector of 8 parameters $\theta^{(t)}$ for each year.

To estimate the 23 time-dependent coefficients ψ specified for the USA in the polynomials (8), equation (10) is maximized trough the algorithm `optim` implemented on `R`. Since ω_m, ω_M are defined only positive and the range of the two mixture coefficients α and η is $[0, 1]$, sometimes the numerical optimization algorithm reaches combinations of points which define not admissible values. These two restrictions, which produce problem in the maximization of the likelihood, can easily solved by employing a ri-parametrization. This is the reason why for ω_m and ω_M a logarithm transformation is used, while the logit function is selected for α and η. Additionally, the starting points need to be chosen carefully to facilitate the convergence of the algorithm. Considering the trends of the parameters θ estimated year by year separately, a rough approximated estimate of time-related coefficients is computed fitting the polynomials of first, second and third degree with a linear model. Next, to improve upon the initial estimation, a refinement step is undertaken. For each coefficient of ψ, a set of starting values is defined sampling randomly from a Normal distribution using their linear model estimates and standard error as mean and standard deviation. The algorithm is then run using all the random combinations as starting points. The set of the parameters ψ with the higher likelihood value is finally chosen. The number of random sets is fixed at 300, which is the minimum quantity that ensures the convergence on the global maximum.

4 Results

The convergence of the algorithm to maximize the likelihood is quite fast, but not always the global maximum is reached. For this reason a set of different starting vectors is necessary, even if the time consuming increases significantly. As an example, in Figure 2, the 300 curve estimated for the trends of ξ_m and λ_M are drew. The identification of the right polynomial $(\xi_{m0}, \xi_{m1}, \xi_{m2})$ for the path of ξ_m, the shape parameter of premature component, has some obstacles as it is possible to see in Figure 2a, where there are several curves that are clearly inconsistent. The coefficients of early mortality $(\alpha, \xi_m, \omega_m$ and $\lambda_m)$ also in the estimation year by year have shown several identification problems which are reflected in the zig-zag trends. Cer-



(a) (b)

Figure 2: Polynomials estimated using the 300 random starting points for two parameters of the mixture model. The functions whose time-related coefficients reach the higher likelihood value are highlighted.

tainly, this issue has affected also the estimates of the curve for ξ_m tendency. Moreover, since the functional form for the trend is assigned based on the observation of the values computed year by year, it is maybe possible that a parabola is not the best option. Regarding the skewness parameter of adult mortality, λ_M, the curves of polynomial estimated with the different starting points are close to each others with only few exceptions, as it is possible to see in Figure 2b. In any case for both the coefficients, the trends traced by the polynomial with the higher likelihood value is coherent with the points. Moreover the two curves are not affected by the truncation of the last open age class at 85+ between 2000 and 2009. The advantage to estimate all year together is clearly visible in Figure 2: the trends obtained using the time-related coefficients ψ is clear, easy to understand and more interpretable. In the case of more regular paths, for instance, for parameter ξ_M and η, the identification of the functional form of the polynomial is easier and also the estimation of its time-dependent coefficients. In these cases the 300 curves almost overlap each others. Instead, the behaviors of $\omega_m, \lambda_m, \alpha$ and ω_M are similar to Figure 2b.

Starting from ψ, the time-related coefficients of the polynomials, it is possible to compute the vector of 8 parameters for each year, $\theta^{(t)}$, and to compare the curve of the mixture model estimated year by year with the one obtained applying the comprehensive procedure. Since the period covered by the data is 58 years, a selection of 4 significant cases is reported in Figure 3. In 1960 (second year of the time series) and in 1990 the two curves almost overlap, in particular in the second graph, where no differences are visible. In Figure 3a the model estimated year by year seems to capture better the senescent deaths (after the mode of the death curve), while the new methodology fits more accurately the adult ones (before the mode). The 2007 is one of the years in which the last open age class is truncated at age 85+. As you can see in Figure 3c, the mixture model computed with the classic procedure tends to estimate a too skewed curved, which is inconsistent also considering the shape of the distribution of deaths after 2010, where the last open age class is again 100+. Instead, the parameters estimated from the time-related coefficients ψ allow to draw a more robust model, which is not affected by the range of the last class of deaths' counts. In the last year of the time series, 2016, the shape of the two models appear again very close, but the one estimated taking into account all the years approximate better the deaths around the mode of the distribution.

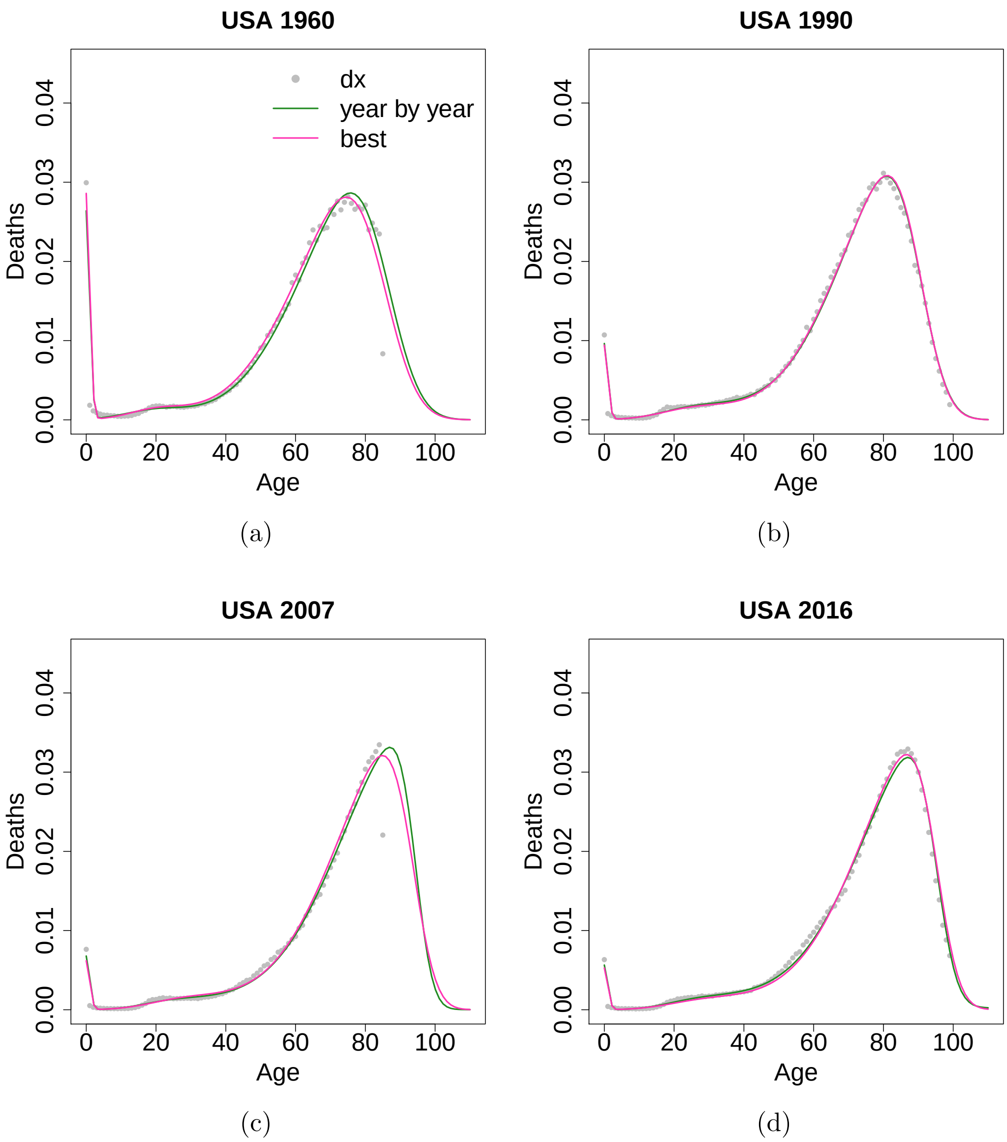


Figure 3: Comparison between the curve of the mixture model estimated year by year separately and the one calculated as result of the functions of time.

5 Conclusion

A smoothing technique to obtain regular parameters' trends of a mixture mortality model is here presented. Fitting the model considering each year on it own generates, in most of the cases, raw fluctuations in the parameter evolutions because, in the estimates, the time component is completely omitted. Instead of computing the vector of parameters year by year, informations regarding past and future need to be taken into account. The goal is obtained specifying for all the parameters of the model time-dependent functions, whose coefficients are estimate directly, maximizing the likelihood using the deaths of the entire available period.

By doing so, the number of unknown quantities to estimate is smaller: instead of calculating a vector of model parameters for each year of the time series, only the coefficients of the time-related functions need to be computed. Moreover a single maximization is required because the time-dependent coefficients are estimates all at once. The parameters' trends obtained with the new procedure are smooth, so they provide a clear indication about mortality evolution, and easier to interpret than the ones computed by fitting the model year by year. Furthermore, the fit of the mixture model whose parameters are reconstructing starting from the time-dependent coefficients, show a satisfactory adaptation which is close and in some case better than the one obtained with the estimates year by year.

To reach satisfactory estimates of the time-related coefficients, al least 300 vectors of different starting points are required to identify the global maximum of the likelihood function. Thus, the estimation of smoothing trends is time-consuming while fitting the model year by year is faster. Furthermore, the selection of the polynomials to assign to each model parameter is based on the trajectory observed on the estimates year by year, in the belief that most of them are correctly identified. Finally, how the choice of the polynomials of the time-related functions influences the estimates of the parameter trends is not established: the effects of an improper specification of the functional form is not yet studied. Although the above mentioned criticisms, the smoothing procedure allows to reach the target set, ensuring both parameters' trends without irregularities and suitable fit of the mixture model for each year of the time series.

References

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

Basellini, U. and Camarda, C. G. (2016). Modeling and forecasting age at death distributions. In *PAA Annual Meeting*.

Canudas-Romo, V. (2010). Three measures of longevity: Time trends and record values. *Demography*, 47(2):299–312.

Canudas-Romo, V., Mazzuco, S., and Zanotto, L. (2018). Chapter 10 - measures and models of mortality. In Rao, A. S. S. and Rao, C., editors, *Integrated Population Biology and Modeling, Part A*, volume 39 of *Handbook of Statistics*, pages 405 – 442. Elsevier.

Cheung, S. L. K., Robine, J.-M., Tu, E. J.-C., and Caselli, G. (2005). Three dimensions of the survival curve: Horizontalization, verticalization, and longevity extension. *Demography*, 42(2):243–258.

Congdon, P. (1993). Statistical graduation in local demographic analysis and projection. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 156(2):237–270.

Gompertz, B. (1825). On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philosophical Transactions of the Royal Society of London*, 115:513–583.

Heligman, L. and Pollard, J. H. (1980). The age pattern of mortality. *Journal of the Institute of Actuaries*, 107(01):49–80.

Kannisto, V. (1994). Development of oldest-old mortality 1950-1990: evidence from 28 developed countries. *Monographs on Population Aging*, 1.

Makeham, W. M. (1860). On the law of mortality and the construction of annuity tables. *The Assurance Magazine, and Journal of the Institute of Actuaries*, 8(6):301–310.

Mazzuco, S., Scarpa, B., and Zanotto, L. (2018). A mortality model based on a mixture distribution function. *Population Studies*, pages 1–10.

Pearson, K. (1897). *Chances of Death, and Other Studies in Evolution*, volume 1. London: Edward Arnold.

Siler, W. (1979). A competing-risk model for animal mortality. *Ecology*, 60(4):750–757.

Van Raalte, A. A. and Caswell, H. (2013). Perturbation analysis of indices of lifespan variability. *Demography*, 50(5):1615–1640.

Wilmoth, J. R. and Horiuchi, S. (1999). Rectangularization revisited: variability of age at death within human populations. *Demography*, 36(4):475–495.

Zanotto, L., Canudas-Romo, V., and Mazzuco, S. (2017). Evolution of premature mortality. *Under review*.