Uncovering adult modal age at death in populations with grouped data

Carlo G. Camarda* Nadine Ouellette†

Extended abstract submitted to the
ALAP-XX ABEP Congress, Foz do Iguau, Brazil, 17-22 October 2016

Abstract

The adult modal age at death is a measure of the most frequent length of life among adults. It has been demonstrated to be an important lifespan measure in longevity research, capable of shedding light on some specific aspects of old-age mortality that are not necessarily captured well by other widely used measures of old-age survival. The current P-spline method for estimating the modal age at death is flexible and highly effective but the model requires population size and mortality data that are detailed by single years of age. For several countries and regions of the world, these data are solely available for broader age groupings. In this paper, we introduce a generalized version of the earlier method based on the Penalized Composite Link Model to estimate the modal age at death in instances where population estimates and/or mortality data are not provided by single years of age. We start by illustrating the new method and assessing its performance using data from the Human Mortality Database. We then uncover sex-specific trends in the modal age at death between 1996 and 2010 in Brazil, a country where population estimates and mortality data are readily available by 5-year age groups only from the WHO Mortality Database and Latin American Human Mortality Database. We offer concluding remarks about the vast range of possible applications and extensions of the newly proposed method.

Keywords: Composite Link Model; Grouped data; Latent distribution; Longevity; Modal age at death; Poisson; Smoothing.

*Corresponding author: Institut National d'Études Démographiques. 133, Bd. Davout, 75020 Paris, France; Tel. +33(0)1 5606 2155, email: `carlo-giovanni.camarda@ined.fr`

†Institut National d'Études Démographiques, Paris; email: `nadine.ouellette@ined.fr`

1 Introduction

In the past fifteen years, there has been a growing interest in the adult modal age at death, which corresponds to the most frequent age at death and stands as one of the three central tendency measures of the age distribution of deaths. Unlike the life expectancy at birth and median age at death, the modal age at death is a central value solely influenced by old-age mortality, thus making it a key measure in the current era of longevity extension (Horiuchi et al., 2013). Due to this property, patterns of trends and differentials in the modal age at death can be noticeably different from those in the life expectancy at birth and the median age at death, and thereby shed light on some specific aspects of mortality in later stages of the lifespan (Canudas-Romo, 2010; Cheung and Robine, 2007; Cheung et al., 2009; Horiuchi et al., 2013; Kannisto, 2001; Office of National Statistics, 2012). The modal age at death may also be of substantial theoretical importance in aging research because recent studies have revealed its critical role in major mathematical models of adult mortality such as the Gompertz, logistic, and Weibull models (Horiuchi et al., 2013; Missov et al., 2015).

While the life expectancy at birth and median age at death can be computed from complete or even abridged life tables, determining the modal age at death is more challenging because of the typical flatness and irregular pattern of life table deaths in this area. For this reason, a flexible nonparametric method for estimating the modal age at death was recently proposed (Ouellette and Bourbeau, 2011). Instead of smoothing the age distribution of life table deaths to locate it modal value, this P-spline method for Poisson counts builds on observed deaths and population exposures (i.e., person-years lived) to obtain a smooth age pattern of mortality from which the corresponding smooth age-at-death distribution can then be derived. The model is such that both death counts and population exposures have to be provided for single years of age. For several countries and regions of the world, however, data are solely given for broader age groups (e.g., 5-year age groups). A generalized version of the current P-spline approach for estimating the modal age at death is thus highly desirable. In this paper, we introduce such a generalization by adapting an earlier method for estimating smooth distributions from coarsely grouped data (Rizzi et al., 2015) to accommodate the estimation of the adult modal age at death in cases where population and/or mortality data are not available by single years of age.

2 Sources of data

We used three sources of data for our analyses. To illustrate the method and evaluate its performance, we first used data for Sweden taken from the Human Mortality Database

(HMD). We extracted male population exposure and death counts detailed by single years of age for the year 1950. We then aggregated the exposure and death counts into 5-year age groups. With these two series of HMD data, we were able to compare the estimated modal age at death obtained with 1-year versus more coarsely grouped input data.

The method proposed in this paper for estimating the modal age at death was then applied to sex-specific population and mortality data for Brazil, available by 5-year age groups only, and which covers calendar years ranging from 1996 to 2010. Mid-year U.N. population estimates for Brazil were obtained from the Latin American Human Mortality Database and annual death counts from the WHO Mortality Database.

We focus on ages 10 and older because mortality in infancy and early childhood presents unique features that are not of interest for estimating the adult modal age at death by any means.

3 Methods

Assuming that both exposures and death counts are grouped into broader than 1-year age classes, we need to proceed in two successive steps. First, we redistribute the population's amount of exposure into single years of age and then, we use this result to estimate death counts as well as the force of mortality in each single years of age. The first step can be overlooked when death counts only are available in coarse age groups.

We start by presenting the ungrouping procedure for the population exposures. Let $\breve{e}_i, i = 1, \ldots, I$ be the observed grouped exposures in each i-th group. We aim to estimate the latent distribution of exposures by single years of age: $e_j, j = 1, \ldots, J$. We assume that the observed grouped data are realization of a Poisson distribution with a composed mean:

$$\boldsymbol{e} \sim \mathcal{P}(\boldsymbol{C}\,\breve{\boldsymbol{e}})\,, \tag{1}$$

where $\boldsymbol{C}$ is the $I \times J$ composition matrix that maps the observed grouped data $\breve{\boldsymbol{e}}$ to the latent distribution of $\boldsymbol{e}$:

$$\boldsymbol{C} = \begin{bmatrix} 1 & \ldots & 1 & 0 & \ldots & \ldots & \ldots & \ldots & 0 \\ 0 & \ldots & 0 & 1 & \ldots & 1 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & 0 & \ddots & 0 & 0 & \vdots & \vdots \\ 0 & \ldots & 0 & 0 & \ldots & 0 & 1 & \ldots & 1 \end{bmatrix}.$$

The elements of $\boldsymbol{C}$ are equal to 1 if age j falls within the age group i and zero otherwise. This model is a Composite Link Model (Thompson and Baker, 1981), which can be estimated using a penalized likelihood approach (Eilers, 2007), and it was presented in earlier works for estimating grouped continuous data (Camarda et al., 2012; Lambert and

Eilers, 2009; Rizzi et al., 2015).

Since $J > I$, we deal with an underdetermined problem and certain assumptions about the distribution of $\boldsymbol{e}$ are needed. Instead of searching for a parametric description of the latent distribution for exposures, it seems natural to exploit the natural order in the solution and to assume smoothness. This is not just a computational trick: as we know little from grouped observations, it is reasonable to assume a smooth curve which, itself, has little detail. In the absence of specific reasons for irregularities in the exposures distribution, we cannot expect to estimate fine details from limited data: we accept smooth results and we can obtain some useful insights from data that, as collected, can not provide any response.

Specifically, we describe the latent distribution as a linear combination of equally-spaced B-splines and coefficients:

$$\ln(\boldsymbol{e}) = \boldsymbol{B}\,\boldsymbol{\beta},$$

and we enforce smoothness of the coefficients-vector $\boldsymbol{\beta}$ by penalizing the associated log-likelihood function with a roughness term measured by the second-order differences of the neighboring coefficients (Eilers and Marx, 1996).

Figure 1 provides an illustration - the case of Swedish males aged 10 and above in 1950. Recall that we extracted the exposures data from the HMD by single years of age from 10 to 105, then we grouped the data by 5-year age groups (10-14, 15-19, ..., 90-94, 90+). The dotted red line depicts the original detailed data and the bars show the grouped data. From the grouped data, the figure reveals that we were able to reproduce well the original detailed distribution of population exposures (see the blue solid line).

The method for ungrouping death counts is only slightly more complicated than the exposures one. Again, we assume that observed deaths $\breve{d}_i$ in each 5-year age groups $i = 1, \ldots, I$ are realizations from a Poisson distribution with a composed mean as in (1). However, the expected values in the composed mean are the product of exposures by single years of age and the actual force of mortality at the corresponding age, μ_j. Because we have previously obtained an estimation of e_j, our aim is now to describe the force of mortality, μ_j.

The information about e_j can be incorporated in the composite matrix $\boldsymbol{C}$. Assuming the grouping structure as in the previous Swedish example with estimated exposures from age 10 to 105, the composite matrix can be written as follows:

$$\boldsymbol{C} = \begin{bmatrix} e_{10} & \ldots & e_{14} & 0 & \ldots & \ldots & \ldots & \ldots & 0 \\ 0 & \ldots & 0 & e_{15} & \ldots & e_{19} & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & 0 & \ddots & 0 & 0 & \vdots & \vdots \\ 0 & \ldots & 0 & 0 & \ldots & 0 & e_{90} & \ldots & e_{105} \end{bmatrix}.$$

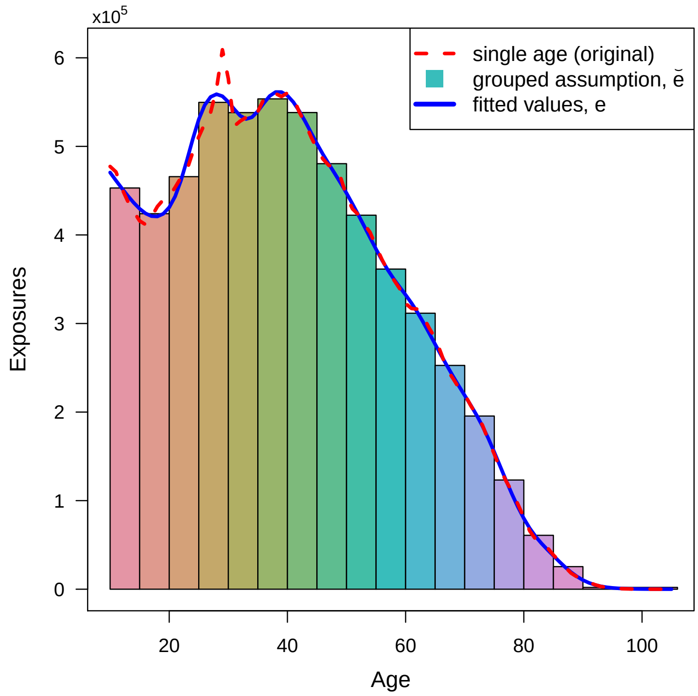

Figure 1: Population exposures. By single years of age (original data), grouped into 5-year age classes, and estimated from grouped data. Swedish males aged 10-105, 1950.

We have again a composite link model. Therefore, a description of the force of mortality by means of P-splines and a penalized likelihood smoothing approach for the estimation, allows us to estimate the latent force of mortality over the age domain as well as the corresponding age distribution of deaths.

Figure 2 illustrates the outcomes for the force of mortality as well as the age distribution of deaths among Swedish males in 1950. As in the case of population exposures, we grouped data by 5-year age groups (10-14, 15-19, ..., 90-94, 90+) based on the original HMD data by single years of age. Both original and grouped observations are plotted on the figure, along with the estimated $\boldsymbol{\mu}$ and $\boldsymbol{d}$. It shows that our approach is capable of reproducing adequately the underlying age-trajectory of mortality by using the grouped data only.

It should be noted that grouping structures for exposures and deaths can differ, as long as the estimation of the latent distribution of exposures is done at the details targeted for the estimation of the latent force of mortality.

Finally, using the estimated coefficients obtained for the force of mortality, we can evaluate the associated B-splines at any desired fine grid. This allows us to compute the corresponding smoothed survival function by standard numerical integration techniques. The product of the force of mortality and the survival function yields the density function, from which we can obtain the modal age at death (for more details, see Ouellette and Bourbeau (2011)).

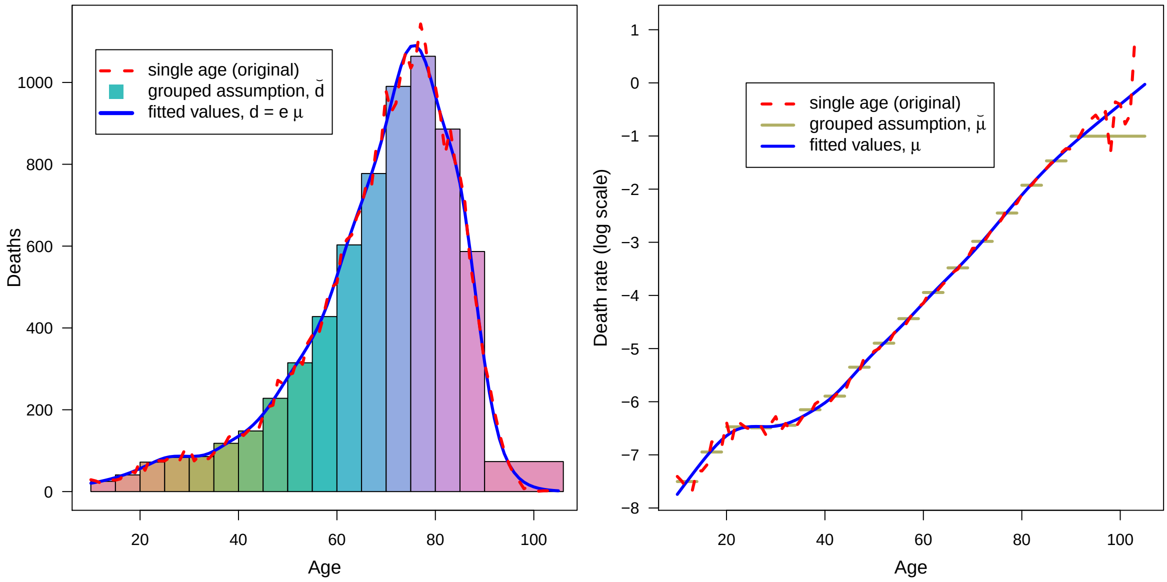

Figure 2: Deaths (left panel) and death rates (right panel). By single years of age (original data), grouped into 5-year age classes, and estimated from grouped data. Swedish males aged 10-105 in 1950.

Figure 3 shows the smooth density function for Swedish males in 1950 estimated using solely the data on population exposure and death counts by 5-year age groups. The corresponding modal age at death is 78.59 years, or only 0.17 years smaller than the modal age at death computed from data by single years of age, at 78.76 years.

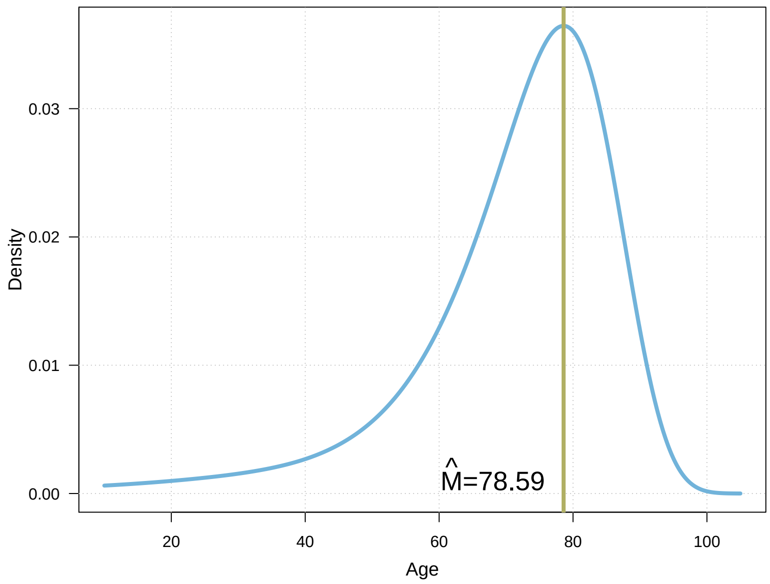

Figure 3: Smooth density function describing the age-at-death distribution. Standardized with respect to exposure-to- risk and estimated from grouped data. Swedish males aged 10-105 in 1950.

We obtained comparable outcomes when testing the method on other combinations of HMD countries and sex, thereby suggesting that the proposed methodology for estimating modal age at death in populations with grouped data performs well.

4 Application

We used the method described above to uncover trends in modal age at death since 1996 in Brazil, a country where death counts and corresponding population's amount of exposure to the risk of death are not readily available by sigle years of age. Figure 4 provides the outcomes for females in calendar year 2000.

The bottom left panel shows, as a step function, the observed death rates for ages classes 10-14, 15-19, . . ., 90-94, 95+. The superimposed solid blue line in this panel gives the fitted death rates obtained after applying the penalized composite link model, first to smoothly ungroup the binned population exposures over age (top left panel), and, similarly, to smoothly ungroup the binned deaths counts (top right panel). The smooth density describing the age distribution of death can then be derived, as shown in the bottom right panel, together with the estimated modal age at death of 84.94 years.

We applied the method separately for each sex to successive calendar years up to 2010. Figure 5 reveals that the sex-specific modal ages at death have increased between 1996 and 2010 in Brazil. Females and males show an increment of about 2 years and 1 year, respectively, over the entire study period. In addition to having an advantage in terms of pace of increase, female modal age at death values are consistently greater than those for males (difference of 5 years in 1996 and 6 years in 2010). Based on the trends depicted in the figure, future rises in modal age at death are expected for each sex because there is currently no signs of a slowdown.

5 Concluding remarks

It is often the case that population and/or mortality data detailed at the individual level or by single years of age are not readily available. Several developing countries and historical populations, among others, offer coarsely grouped data only. In these instances, the current flexible methodology based on Poisson P-splines proposed by Ouellette and Bourbeau (2011) to estimate the adult modal age at death and to investigate its variations over time cannot be used. The method proposed in this paper fills this gap by generalizing the earlier method.

In the next few months, we intend to take advantage of the new method to study trends in modal age at death not only in the instances listed above, but also to expand our work on cause-specific modal age at death (Diaconu et al., 2015). Indeed, for most countries

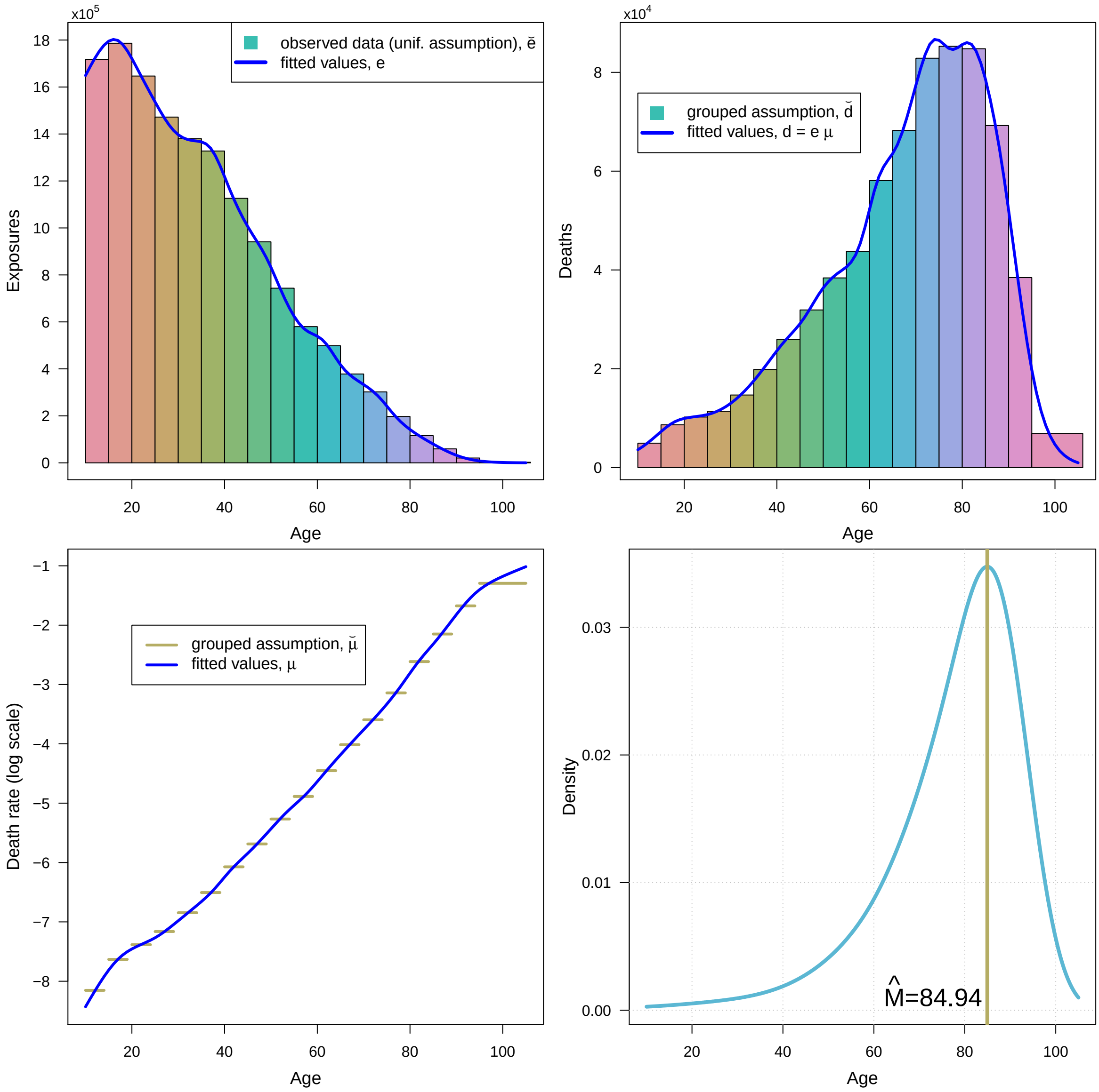

Figure 4: Exposure population (top left), deaths (top right), death rates (bottom left), and smoothed density function (bottom right). Brazilian females ages 10-105 in 2000.

of the world, data on deaths by cause of death are provided by 5-year age groups at best. We will also explore the possibility of generalizing the new method further by accounting for smoothing simultaneously along the age and time dimensions. Such a generalization may prove useful for studying regional disparities in modal age at death (Ouellette et al., 2012) in countries where grouped data is accessible at a fine geographic scale.

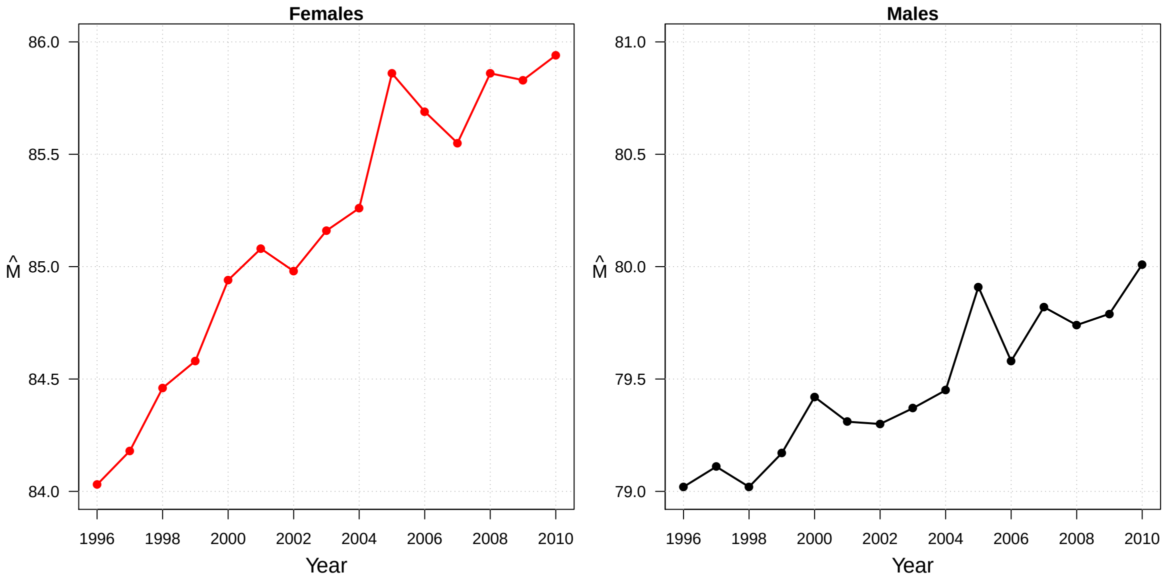

Figure 5: Estimated modal age at death based on smoothed density functions estimated from grouped data.

Acknowledgment

This paper was written as part of the following funded research projects: "AXA Project on Mortality Divergence and Causes of Death" and "Project ANR-12-FRAL-0003-01 DIMOCHA".

References

Camarda, C. G., P. H. C. Eilers, and J. Gampe (2012). Additive Decomposition of Vital Rates from Grouped Data. In A. Komárek and S. Nagy (Eds.), *Proceedings of the 27th International Workshop of Statistical Modelling*, pp. 57–62.

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

Cheung, S. L. K. and J.-M. Robine (2007). Increase in common longevity and the compression of mortality: The case of Japan. *Population Studies 61*(1), 85–97.

Cheung, S. L. K., J.-M. Robine, F. Paccaud, and A. Marazzi (2009). Dissecting the compression of mortality in Switzerland, 1876-2005. *Demographic Research 21*, 569–598.

Diaconu, V., N. Ouellette, C. Camarda, and R. Bourbeau (2015). Insight on "typical" longevity: An analysis of the modal lifespan by leading causes of death in Canada. Submitted.

Eilers, P. H. C. (2007). Ill-posed Problems with Counts, the Composite Link Model and Penalized Likelihood. *Statistical Modelling 7*, 239–254.

Eilers, P. H. C. and B. D. Marx (1996). Flexible Smoothing with *B*-splines and Penalties (with discussion). *Statistical Science 11*, 89–102.

Horiuchi, S., N. Ouellette, S. L. K. Cheung, and J.-M. Robine (2013). Modal age at death: Lifespan indicator in the era of longevity extension. *Vienna Yearbook of Population Research 11*, 37–69.

Human Mortality Database (2015). University of California, Berkley (USA), and Max Planck Institute for Demographic Research (Germany). Available at `http://www.mortality.org`. Accessed 30 November 2015.

Kannisto, V. (2001). Mode and dispersion of the length of life. *Population (An English Selection) 13*(1), 159–171.

Lambert, P. and P. H. C. Eilers (2009). Bayesian density estimation from grouped continuous data. *Computational Statistics & Data Analysis 53*, 1388–1399.

Latin American Human Mortality Database (2015). B. Piedad Urdinola and Bernardo L. Queiroz. Available at `http://www.lamortalidad.org`. Accessed 30 November 2015.

Missov, T. I., A. Lenart, L. Nemeth, V. Canudas-Romo, and J. W. Vaupel (2015). The gompertz force of mortality in terms of the modal age at death. *Demographic Research 32*, 1031–1048.

Office of National Statistics (2012). Mortality in England and Wales: Average life span. Tech. rep. London: Office of National Statistics.

Ouellette, N. and R. Bourbeau (2011). Changes in the age-at-death distribution in four low mortality countries: A nonparametric approach. *Demographic Research 25*, 595–628.

Ouellette, N., C. Camarda, and R. Bourbeau (2012). Regional disparities in Canadian adult and old-age mortality: A comparative study based on smoothed mortality ratio surfaces and age at death distributions. *Canadian Studies in Population 39*(3-4), 79–106.

Rizzi, S., J. Gampe, and P. H. C. Eilers (2015). Efficient estimation of smooth distributions from coarsely grouped data. *American Journal of Epidemiology 182*(2), 138–147.

Thompson, R. and R. J. Baker (1981). Composite Link Functions in Generalized Linear Models. *Applied Statistics 30*, 125–131.

World Health Organization (2015). Detailed data files of the WHO Mortality Database. `http://www.who.int/entity/healthinfo/statistics/mortality/_rawdata/en/index.html`. Accessed 30 November 2015.