DIFFERENCES IN CAUSE-SPECIFIC LIFESPAN DISTRIBUTIONS: THE CONFOUNDING EFFECT OF AGE

Viorela Diaconu[1], Robert Bourbeau[1], Nadine Ouellette[2], and Carlo G. Camarda[2]

[1] *Department of Demography, Université de Montréal*
[2] *Institut National d'Études Démographiques (INED)*

Introduction

In human populations, great dissimilarities can be observed between cause-specific lifespan distributions. Variability in the typical length of life occurs from the interaction between the lethality of the disease and individual's ability to fight the disease. These two components depend in turn on the age at which the illness becomes clinically obvious. A 50-year old individual has a higher probability of surviving from a heart attack than from a lung cancer and a lower chance of battling any of these illnesses at an older age. As one's capacity of remaining alive depends on its genetic baggage and lifestyle choices, inter-individual differences are observed in the organism's capacity of resisting to a certain pathogen. Consequently, some causes of death are dispersed over a narrower or a wider age interval. The analysis of differences of cause-specific lifespan distributions is therefore affected by the rate of speed at which mortality occurs. In order to remove this "tempo bias", age needs to be standardized according to a typical length of life indicator, such as the mean, the median, or the modal age at death.

The age-standardization approach has been widely-used in studies on interspecies lifespan. As Buffon (1976) stipulated in its general law on species longevity, the less time a tree or a species takes to grow, the faster they perish and the shorter their life duration. Comparing interspecies typical length of life such as measured by the mean, median, or modal age at death, on the same time-scale ignores the stark differences in the aging process of various populations. Researchers thus proposed to standardize time in order to remove its confounding effect on the indicators of length of survival time. Various normalizing approaches have been identified in the biodemographic literature. Inter alia, the rescaling of time using life expectancy at birth (Pearl and Miner, 1935), the modal age at death (Horiuchi, 2003), life expectancy at maturity (Baudisch, 2011) or in some particular cases the mortality curve of a baseline species (Caswell and al., 1998).

Accordingly, the objective of this paper is to compare the cause-specific age-at-death distributions on the absolute and standardized by the associated modal age at death time scales. We focus only on six leading causes of death in Canada, that are cardiovascular diseases, heart diseases, and three types of cancers (breast, prostate, colorectal, and trachea, bronchus, and lung). This comparative analysis is twofold. Firstly, for a given calendar year and sex, we juxtapose the age-at-death

distributions by each cause on an absolute and relative to the modal age scales. In doing so, we highlight the importance of controlling for the tempo at which mortality progresses when drawing conclusions regarding the dispersion of ages at death by cause. Secondly, we examine the changes underwent by the age-at-death distribution of each cause overtime, that is between 1974 and 2007. We thus identify the causes of death who have experienced a compression of mortality and those for which this phenomenon manifested itself to a greater extent.

Data

Through the "Data Liberation Initiative", a program initiated by Statistics Canada to improve access to data resources at Canadian postsecondary institutions, we were granted access to the Canadian Vital Statistics database for calendar years 1974 to 2007. This database provides information on observed death counts by single calendar year, sex, single year of age and underlying cause of death. Causes of death are classified according to the World Health Organization "International Statistical Classification of Diseases and Related Health Problems" (ICD). To ensure the comparability between ICD-8, ICD-9 and ICD-10 revisions, we restricted the analysis on larger groups of causes instead of more detailed ones. In this fashion, we minimize the impact of the different revisions on our results.

Information on exposure to risk by calendar year, sex, and single year of age were extracted from the Human Mortality Database.

Methods

For a given calendar year and sex, let d_i^k represent observed death counts by single year of age i (above age 10) and cause of death k and e_i denote the population's amount of exposure to the risk of dying for every single year of age i. The cause-specific central death rate at age i, $m_i^k = d_i^k / e_i$, is generally quite close to the cause-specific instantaneous death rate in the middle of the single-year age interval, i.e. $m_i^k \simeq \mu_{i+1/2}^k$ (Thatcher et *al.*, 1998, Appendix A). Under the assumption of a constant force of mortality over each one-year of age-interval, i.e. $\mu^k(x) = \mu_i^k$ for all $x \in [i, i+1)$, death counts by single year of age and cause of death are assumed to be realizations from a Poisson distribution with mean $e_i \cdot \mu_i^k$. This "piece-wise constant" assumption is the core basis for the estimation of empirical central death rates (Camarda, 2008).

The density function of a particular cause k describing its age-at-death distribution for a given calendar year and sex is given by:

$$\begin{aligned} f^k(x) &= \mu^k(x) S(x) \\ &= \mu^k(x) \left(\exp\left[- \int_0^x \{\mu^1(u) + \mu^2(u) + \cdots + \mu^K(u)\}\, du \right] \right) \end{aligned}$$

where $\mu^k(x)$ is the force of mortality of cause *k* and *S(x)* is the overall (all-cause) survival function. The second equality holds under the assumption that causes of death are mutually exclusive and mutually exhaustive (Preston et *al.*, 2001).

Cause-specific forces of mortality and corresponding density functions are obtained using a nonparametric smoothing method based on *P-splines* (penalized B-splines), which we specifically adapted to the context of cause of death analysis (Diaconu et *al*. 2013a, Diaconu *et al*., 2013b). Therefore, for a given calendar year and sex smoothed force of mortality for cause *k* is computed as:

$$\hat{\mu}^k(x) = \exp\left(\boldsymbol{B}(x)\hat{\boldsymbol{\alpha}}^k\right), \quad (2)$$

where $\boldsymbol{B}$ is the B-spline basis matrix and $\hat{\boldsymbol{\alpha}}^k$ is the vector of estimated cause-specific coefficients for each B-spline included in $\boldsymbol{B}$. It follows that smoothed cause-specific density curves can be then derived as:

$$\hat{f}^k(x) = \hat{\mu}^k(x)\left(\exp\left[-\int_0^x (\hat{\mu}_i^1(u) + \hat{\mu}_i^2(u) + \cdots + \hat{\mu}_i^k(u))du\right]\right). \quad (3)$$

The *P-spline* method has proved to be a highly effective approach for smoothing mortality rates and consequently for obtaining smooth forces of mortality (Currie et *al*., 2004; Camarda, 2008, 2012). Once smooth cause-specific density functions are obtained, we estimate the modal age at death of each cause by numerically computing the correspondent density function (Ouellette and Bourbeau, 2011).

As previously mentioned, comparison of cause-specific age-at-death distributions is tainted by the tempo at which mortality progresses. In order to remove this effect, we standardize the age scale according to the cause-specific modal age at death, that is, standardized age is given by $x_s^k = x/M^k$ where x is the age and M^k is the modal age at death for a given cause of death. As Horiuchi (2003), we chose the modal age at death as the indicator of the typical life span because it is solely determined by old-age mortality (Canudas-Romo, 2010; Horiuchi and al. 2013). Moreover, unlike the other two measures of length of life this indicator is not influenced by juvenile mortality. This standardization approach shifts the chronological age in such a way that all distributions, regardless of their peculiar features, are re-centered at their respective modal age of 1.

Preliminary Results

Visual inspection of figure 1 indicates that when the age-at-death distributions are plotted on a absolute time scale, individuals diagnosed with cardiovascular diseases die, at every age, later on than those diagnosed with colorectal cancer.

Given the bell-shape of the density function, the same proportion of deaths can be observed at ages on the left-hand and on the right-hand side of the curve's inflection point. For example, in 1975 (left panel) a proportion of deaths from colorectal cancer of about 0.01 corresponds to an age at death of 54.5 and of 90.4. Similarly, for cardiovascular diseases this same proportion is noticed at ages 61.7 and 94.3. When comparing the ages at death of these two causes, we notice that the gap amounts to about 7.2 years (purple line) and 3.9 years (green line). Thus suggesting, that on an absolute time scale, cardiovascular diseases experience an "aging of mortality" relative to colorectal cancer, which seems to be of a higher magnitude at relatively younger ages.

A similar analysis for calendar year 2005 revealed that death from cardiovascular diseases occurs later in life when compared to colorectal cancer. However, it seems that the gap between the ages at death of these illnesses widened overtime. As previously, when we keep the proportion of deaths at about 0.01, individuals diagnosed with cardiovascular diseases live about 8.7 years more than those diagnosed with colorectal cancer (orange line). As they get older, this difference amounts to about 4.2 years (light blue line). The wider gap observed in 2005 relative to 1975, is caused by a more important shift of the mortality curve of cardiovascular diseases than the one of colorectal cancer.

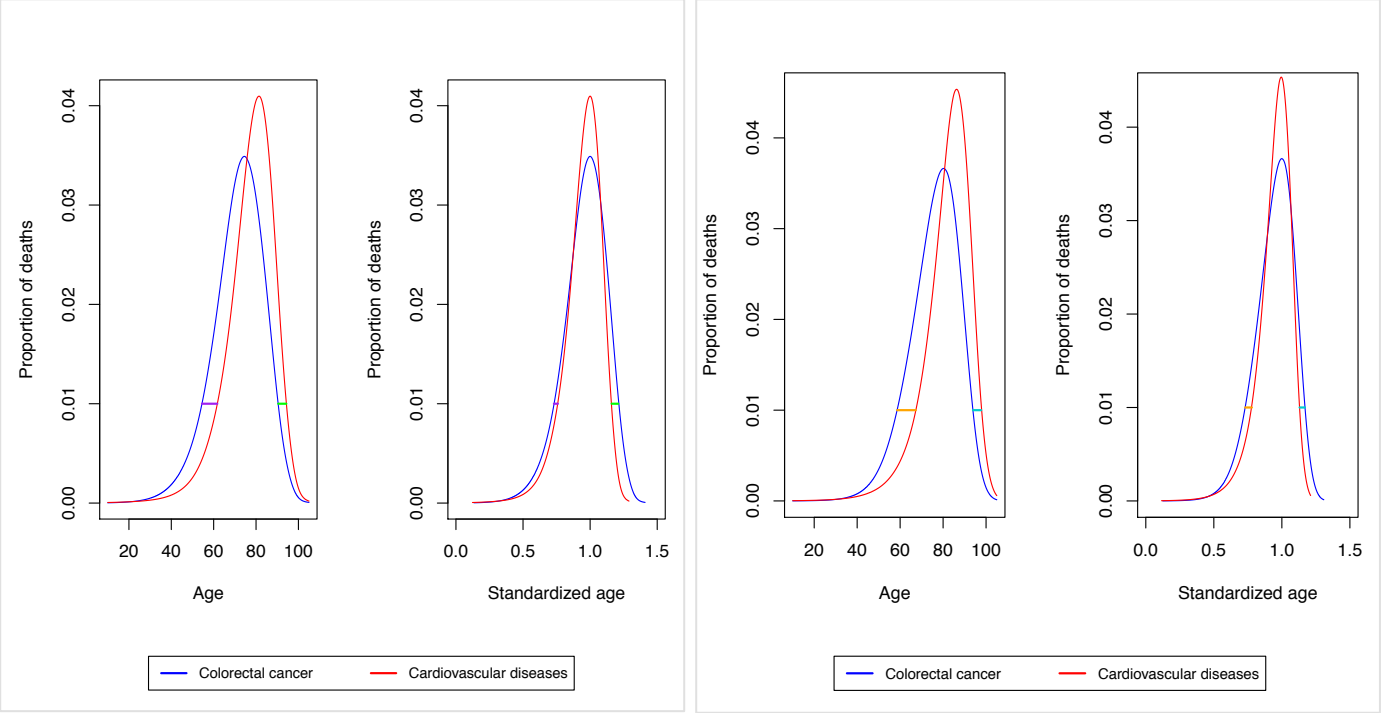
After removing the "tempo bias", it appears that in fact the deaths from cardiovascular diseases do not occur at older ages as it was suggested by the analysis on the absolute scale. In addition, the differences previously observed in the ages at death of these two illnesses for a given proportion of deaths, say 0.01, are now smaller (for 1975, purple and green lines; for 2005 orange and light blue lines). In fact, after rescaling, the density function of cardiovascular diseases is almost entirely contained within the bounds of the colorectal cancer curve. Thus suggesting that mortality from this illness experienced a higher compression of mortality over the entire lifespan. When examining the curves on the right-hand side of the modal values, we notice that at a given proportion of deaths, an individual perishes later on from colorectal cancer than from cardiovascular diseases. Moreover, the older the individuals gets, the faster he succumbs to cardiovascular diseases (always in comparison to colorectal cancer).

The higher concentration of deaths from cardiovascular diseases within a shorter age interval, relative to that of colorectal cancer, can also be seen in 2005. However, it seems that the compression of mortality at ages below the modal values was higher than the one in 1975 (purple vs orange line). This result indicates that overtime the ages at death from cardiovascular diseases became less dispersed when compared to those from colorectal cancer.

A similar comparative analysis will also be conducted using United States' mortality data. By doing, so we are able firstly, to verify if the trends observed in the Canadian context are also noticed elsewhere and, secondly, to identify country-specific features of the age-at-death distribution of a given cause of death.

Figure 1: Age-at-death distributions of colorectal cancer and cardiovascular diseases, absolute and relative time scale, Males, 1975 (left panel) and 2005 (right panel).

A. Year 1975 B. Year 200



Note: The area under the curves over the adult age range is unity, that is $\sum_x [f^k(x)/\sum_x f^k(x)]$

References:

BAUDISCH, A. (2011). The pace and shape of ageing. Methods in Ecology and Evolution, vol. 2(4): 375-382.

BUFFON, G.-L.L. (1776). « Supplément à l'histoire des animaux quadrupèdes. Des Mulets » in *Histoire naturelle*, Suppléments, tome troisième, 1776, Paris, Imprimerie Royale.

CAMARDA, C. G. (2008). Smoothing Methods for the analysis of mortality development. Thèse de doctorat, Universidad Carlos III de Madrid (Spain) and Max Planck Institute for Demographic Research (Germany).

CAMARDA, C. G. (2012). MortalitySmooth: An R package for smoothing Poisson counts with P-splines. *Journal of Statistical Software*, 50(1): 1-24.

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

CASWELL, H. et *al*. (1998). Harbor Porpoise and fisheries: an uncertainty analysis of incidental mortality. *Ecological Applications*, 8: 1226-1238.

CURRIE, I.D. et *al*. (2004). Smoothing and forecasting mortality rates. *Statistical Modelling* 4(4): 279–298.

DIACONU, V. et *al*. (2013a). An innovative analysis of the changes in the age-at-death distribution by leading causes of death in Canada, paper presented at the Population Association of America, 10-13 April 2013, New Orleans, United States.

DIACONU, V. et *al*. (2013b). Insights on longevity: An analysis of the modal life span by leading causes of death in Canada, paper presented at the XXVII IUSSP International Population Conference, 26-31 August 2013, Busan, South Korea.

HORIUCHi, S. (2003). Interspecies differences in the life span: Humans versus Invertebrates. in *Life Span: Evolutionary, Ecological, and Demographic Perspective*, Supplement to *Population and Development Review*, vol. 29, 2003.

HORIUCHI, S. *et al*. (2013). Modal age at death: Lifespan indicator in the era of longevity extension. *Vienna Yearbook of Population Research*, 11: 37-69.

HUMAN MORTALITY DATABASE (2011). University of California, Berkley (USA), and Max Planck Institute for Demographic Research (Germany). Available at http://www.mortality.org or http://www.humanmortality.de.

OUELLETTE, N. and BOURBEAU, R. (2011). Changes in the age-at-death distribution in four low mortality countries: A nonparametric approach. *Demographic Research,* 25(19): 595-628.

PEARL, R. and MINER, J.R. (1935). Experimental studies on the duration of life XIV: The comparative mortality of certain lower organisms. *Quarterly Review of Biology*, 10: 60-79.

PRESTON, S. et *al*. (2001). *Demography: Measuring and Modeling Population Processes*. Oxford: Blackwell Publishers.

THATCHER, A. R. *et al*. (1998). *The force of mortality at ages 80 to 120*. Odense Denmark, Odense University Press.