Paper submitted for the 2016 meeting of the EPC December 2015

LIFESPAN DISPARITY BY LEADING CAUSES OF DEATH IN CANADA AND THE U.S. FROM 1975 TO 2011

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

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

A vast body of existing literature has examined the relationship between socioeconomic status and lifespan variation. However, analyses along a cause-of-death dimension are quite scarce. In this paper, we examine lifespan variation trends over the 1975-2011 period for five leading causes of death in Canada and the U.S. Using flexible P-splines adapted to the context of cause-of-death analysis, we estimate smooth cause-specific age-at-death distributions and subsequently derive the modal age at death (*M*). Because the spread in ages at death differs greatly by cause, lifespan distributions must therefore be compared on a similar time scale. We thus rescale the smoothed distributions according to their corresponding *M*. Preliminary results for Canada and the U.S. show that the five leading causes differ greatly not only in terms of relative lifespan variation trends but also in terms of levels. Moreover, gaps in levels between causes are more pronounced for females than for males.

Introduction

The variability in the length of life observed in any given human population arises notably from physiological and socioeconomic inequalities among individuals. A vast body of existing literature has examined the relationship between socioeconomic conditions and the variability of the age at death. One of the main common findings emerging from these studies is that individuals from a lower socioeconomic background experience shorter lives and face greater uncertainty in their timing of death (Brown et al. 2012, Edwards and Tuljapurkar 2005, Meara et al. 2008, Skolnikov et al. 2003, van Raalte et al. 2011, 2014). Analyses of lifespan variation along the physiological dimension are, on the other hand, quite scarce. Of course, an individual's physiological conditions at a given age play a central role in determining its ability to resist a certain pathogen. Moreover, because the body's vital functions decline with age and because no two individuals are the same, deaths resulting from certain diseases are concentrated into a narrow age range while others are more widely dispersed. Consequently, age distributions of deaths by cause differ greatly in terms of their variability.

Comparison of variability in cause-specific lifespan distributions should therefore take into account the particular age domain over which deaths from each cause occur. One strategy is thus to compare the cause-

specific age-at-death distributions relative to a central lifespan indicator (i.e., mean, median, or mode). The time-standardization approach has been widely used in studies on interspecies lifespan. As Buffon (1776) 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. Researchers thus proposed that patterns of mortality be time-standardized. Various normalizing approaches have been suggested in the biodemographic literature. Inter alia, life expectancy at maturity (Baudisch, 2011), the modal age at death (Horiuchi, 2003), or in some particular cases the mortality curve of a baseline species (Caswell and al., 1998).

Variability in the length of life has been examined in the light of various measures of dispersion such as the interquartile age range (*IQR*), the variance (*VAR*), the standard deviation (*SD*), the e-dagger (e†), Keyfitz's entropy (*H)*, the Theil entropy index (*T*), the Gini coefficient (*G*), and so on. Using historical mortality data for Sweden, Japan, and the United States, Wilmoth and Horiuchi (1999) showed a strong positive correlation between several of these measures. Consequently, they supported the use of the *IQR* since it is easier to calculate and to interpret. However, as showed by Shkolnikov and colleagues (2003), the *IQR* fails to respect one of the three basic properties of a desirable inequality indicator. According to their study, the Gini coefficient is most desirable for measuring how spread out deaths are across the age range. A recent study by van Raalte and Caswell (2013) showed that the various measures of variability are highly sensitive to mortality changes at different ages and that large differences were observed in their values over the age range of 60-80 years. Another widely-used measure of variability, particularly in studies focusing on changes in the distribution of deaths at older ages, is the standard deviation of ages at death above the mode (*SD(M+)*). But, this indicator presents at least three theoretical limitations. Firstly, it is strongly dependent on the modal age at death. Indeed, analyses of trends in *M* and *SD(M+)* in several low mortality countries revealed a negative correlation between these two measures (Cheung and Robine 2007, Kannisto 2001, Robine and Cheung, 2008). Secondly, by examining trends in *SD(M+)* one indirectly assumes that changes underwent by the death distribution on the right-hand and on the left-hand side of *M* are symmetric. However, this is a strong hypothesis given that the distribution of deaths is left-skewed in human populations. Finally, this indicator of dispersion is greatly influenced by deaths occurring at extreme ages. Because mortality data at these ages is quite erratic, this measure is thus subject to high random variation.

In this paper, our first objective is to compare standardized cause-specific age-at-death distributions for the five following leading causes of death in Canada and the U.S.: (1) cardiovascular diseases, (2) heart diseases, (3) trachea, bronchus and lung cancer, (4) colorectal cancer, (5) breast cancer (females) and prostate cancer (males). By including these two countries in our analysis we are able to verify if trends in cause-specific lifespan variation are country-specific or if a similar pattern is observed. Secondly, we want to examine relative cause-specific lifespan variation trends over the 1975-2011 period. Given the multitude of dispersion indicators to choose from, each with their underlying properties and limitations, we must therefore identify the most suitable one for measuring variability of the age at death by cause and eventually adapt it to the cause-of-death context.

The results of this study will allow us to, firstly, determine which causes of death display a relatively higher variability of the age at death and, secondly, uncover causes for which uncertainty in the timing of death has changed since 1975. Our findings on lifespan variability will be of great relevance to individuals and societies, more generally, as uncertainty in the timing of death affects the level of investment in education (Lee 2003), the age at retirement (Kalemli-Oczan and Weil 2002), savings behavior (Dynan et al. 2002, Edwards 2009), the costs of public pensions, as well as those of insurance and annuities (Edwards and Tuljapurkar 2005).

Data

We use observed death counts by single years of age (10 and above)[1], sex, and underlying cause of death, covering the period 1975-2011 in Canada and the U.S. The Canadian cause-specific mortality series are taken from the Vital Statistics Death Database, administered by Statistics Canada, which gathers demographic and medical information from all Canadian provincial and territorial vital statistics registries on all deaths that occurring in the country. Equivalent data for the U.S. comes from the National Vital Statistics System of the National Center for Health Statistics.

Causes of death are classified according to the World Health Organization International Classification of Diseases (ICD). Given that our study period extends from 1975 to 2011, it covers three revisions of the international classification, namely ICD-8 (1969-1978), ICD-9 (1979-1999), and ICD-10 (since 2000). The adoption of ICD-10 led to a significant increase in the number of codes, as existing causes of death are classified in greater detail and newly recognized diseases were added. To ensure the consistency of categories throughout the three revisions of the ICD, we restricted our analysis to large groups of causes. In this way, we minimize the impact of the successive revisions on our results.

The cause-specific data on deaths are supplemented by estimates of population exposure by single years of age (10 and above), sex, and single calendar years (1975 to 2011) for Canada and the U.S., taken from the Human Mortality Database.

Methods

For a given calendar year and sex, let d_{ik} represent observed death counts by single years of age i and cause of death k, and e_i denote the population's amount of exposure to the risk of dying for each age i. Assuming that the force of mortality is constant over each one-year interval, i.e. $\mu_k(x) = \mu_{ik}(x)$ for all $x \in [i, i+1)$, death counts by single years of age and cause of death, d_{ik}, are assumed to be realizations from a Poisson distribution with mean $e_i \cdot \mu_{ik}$ (Brillinger 1986). Generally, μ_{ik} can be estimated by the central death rate $m_{ik} = d_{ik}/e_i$. Although this straightforward approach is quite simple, it does not allow deriving with great precision the cause-specific density functions and the associated modal ages at death. To overcome these limitations, we smooth age- and cause- specific death rates.

Among the various nonparametric smoothing techniques, we opted for the P-splines method, which combines the concept of (fixed-knot) B-splines with a roughness penalty. Developed by Eilers and Marx (1996), this method has been proven highly effective for smoothing mortality rates and hence for obtaining smooth forces of mortality (Currie et al. 2004, Camarda 2008, 2012). For an introduction to P-spline and their use for the estimation of the all-cause modal age at death see Ouellette and Bourbeau (2011).

[1] Lifespan variation is greatly dependent on whether infant and juvenile mortality is included or excluded from the analysis. When variability is examined across the entire age range, the compression of deaths is largely due to reductions in infant and child mortality. Variance measured across the entire age range may therefore undermine progress made in postponing deaths to older ages. Hence, we restricted our analysis to ages 10 and older.

Therefore, for each country, calendar year, and sex, we compute smooth cause-specific density functions, $\hat{f}_k(x)$, using flexible P-spline adapted to the context of cause-of-death (Diaconu et al. 2013) and derive the cause-specific modal ages at death, $\widehat{M}_k$. Following earlier work by Horiuchi (2003) on interspecies life span differences, we then rescale the smooth cause-specific density functions by their corresponding modal ages at death as such:

$$\hat{f}_k^s(x) = \frac{\widehat{M}_k}{100}\hat{f}_k(x).$$

Unlike the mean and median, the modal age at death is solely determined by changes in old-age mortality (Canudas-Romo 2008, Horiuchi et al. 2013). Given this special feature, *M* has become a prominent lifespan indicator in countries where the longevity extension is primarily due to improvements in old-age survival.

Given the attractiveness of *M* as an indicator of longevity extension, one possibility is to use its associated measure of dispersion, that is the standard deviation of ages at death above *M* (*SD(M+)*), to assess cause-specific changes in variability of age at death since 1975. However, as previously mentioned this indicator presents some theoretical limitations. Hence, for the time being, we are monitoring changes in the timing of death using the number of deaths at the mode ($\hat{f}_k^s(\widehat{M}_k)$). Indeed, a higher peak for the lifespan distribution indicates less variability of the age at death. We opted for this measure because it is easy to calculate and to interpret. Moreover it allows us to have a general overview of changes in the timing of death since 1975.

Preliminary results

Figure 1 shows, for illustration purposes, relative lifespan distributions for the five leading causes of death among Canadian females in 1975. Visual inspection of the rescaled distributions in Figure 1 reveals a higher peak for cerebrovascular diseases than any other cause under study. Thus, deaths from cerebrovascular diseases concentrate in a narrower age interval while those attributed to the remaining causes of deaths are dispersed more widely. Differences are also noted when examining the age distribution of deaths for the three types of female cancers. Indeed, breast and lung cancer deaths spread over a wider age interval than for colorectal cancer.

Figure 2 illustrates relative variation of cause-specific lifespan distributions by sex for the five leading causes of death in Canada and the U.S. since 1975 using $\hat{f}_k^s(\widehat{M}_k)$. The Canadian results reveal that variability of the age at death for heart diseases decreased at a faster pace for both males and females than any other leading cause (Fig. 2a). Conversely, the compression of deaths for the three types of cancer was very slow among males until the mid-1980s but accelerated afterwards, especially for prostate cancer. As for females, variability of the age at death for breast and colorectal cancers appears to have decreased at a faster pace than that of lung cancer. The five causes under study differ greatly not only in terms of trends but also in terms of levels. Indeed, the spread of deaths over ages is relatively wide for breast cancer, but narrow for prostate cancer. Lung and colorectal cancer deaths, on the other hand, show great variability for both sexes.

In the U.S., similarly to the Canadian context, the relative proportion of deaths at *M* for heart diseases increased at a faster pace for both males and females (Fig. 2b). U.S. trends in relative lifespan distributions for

the three types of cancers reveal that uncertainty in the timing of death decreased more rapidly for lung cancer, especially for females. Moreover, it appears that variability of the age at death for colorectal cancer remained fairly stagnant for both males and females since 1975. Looking more attentively at male results, we notice a decreasing trend in $\hat{f}_k^s(\widehat{M}_k)$ for cerebrovascular diseases since the early 2000s. Moreover, the number of deaths at *M* for prostate cancer has stopped increasing in more recent years, thus indicating a possible leveling off in lifespan variation for this cause. As in Canada, the five leading causes of death among American males and females differ greatly in terms of dispersion across the age range. Indeed, deaths for prostate cancer are concentrated into a narrow age range while those for breast cancer are more widely dispersed.

In both countries, gaps in levels between causes are more pronounced for females than for males. Based on the relative proportion of deaths at *M*, causes of death rank as follow for Canadian and American females since 1975 in terms of their relative variability (lowest to highest): 1) cerebrovascular diseases, 2) heart diseases, 3) colorectal cancer, 4) lung cancer, and 5) breast cancer. For males, the classification of the five causes varied slightly throughout the period under study but in general the causes rank as follows: 1) prostate cancer, 2) cerebrovascular diseases, 3) heart diseases, and 4) colorectal and lung cancer.

We are currently in the progress of testing the various dispersion measures and identifying their potential limits for measuring lifespan variation by cause of death. Once this analysis completed, we will choose the most suitable one for monitoring changes in the timing of death since 1975.

References:

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

BRILLINGER, D. R. (1986). The natural variability of vital rates and associated statistics. *Biometrics* 42: 693-734.

BROWN, D. C., HAYWARD, M. D., MONTEZ, J. K., HUMMER, R. A., et al. (2012). The significance of education for mortality compression in the United States. *Demography* 49(3): 819-840.

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. [PhD thesis], Madrid: Universidad Carlos III de Madrid, Department of Statistics.

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

CANUDAS-ROMO, V. (2008). The modal age at death and the shifting mortality hypothesis. *Demographic Research* 19: 1179-1204.

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

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

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

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

DIACONU, V. et *al.* (2013). 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.

DYNAN, K. E., SKINNER, J., and ZELDES, S. P. (2002). The importance of bequests and life-cycle saving in capital accumulation: A new answer. *American Economic Review* 2(92): 274-278.

EILERS, P. H. C. and MARX, B. D. (1996). Flexible smoothing with B-splines and penalties (with discussion). *Statistical Science* 11(2): 89-102.

EDWARDS, R. D. (2013). The cost of uncertain lifespan. *Journal of Population Economics* 26: 1485-2522.

EDWARDS, R. D. and TULJAPURKAR, S. (2005). Inequality in life spans and a new perspective on mortality convergence across industrialized countries. *Population and Development Review* 31(4): 645-674.

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.

KALEMLI-OZCAN, S. and WEIL D. N. (2002). Mortality Change, the Uncertainty Effect, and Retirement. *NBER Working Paper* 8742.

KANNISTO, V. (2001). Mode and dispersion of the length of life. *Population* 1: 159-171.

LEE, R. (2003). The demographic transition: Three centuries of fundamental change. *Journal of Economic Perspectives* 17(4): 167-190.

MEARA, E. R., RICHARDS, D. S., and CUTLER, D. M. (2008). The Gap Gets Bigger: Changes in Mortality And Life Expectancy, By Education, 1981–2000. *Health Affairs* 27(2): 350–360.

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.

ROBINE, J.-M. and CHEUNG, S. L. K. (2009). Nouvelles observations sur la longévité humaine. *Revue économique* 5(59): 941-953.

SHKOLNIKOV, V. M., ANDREEV, E., and BEGUN A. Z. (2003). Gini coefficient as a life table function. Computation from discrete data, decomposition of differences and empirical examples. *Demographic Research* 8: 305-358.

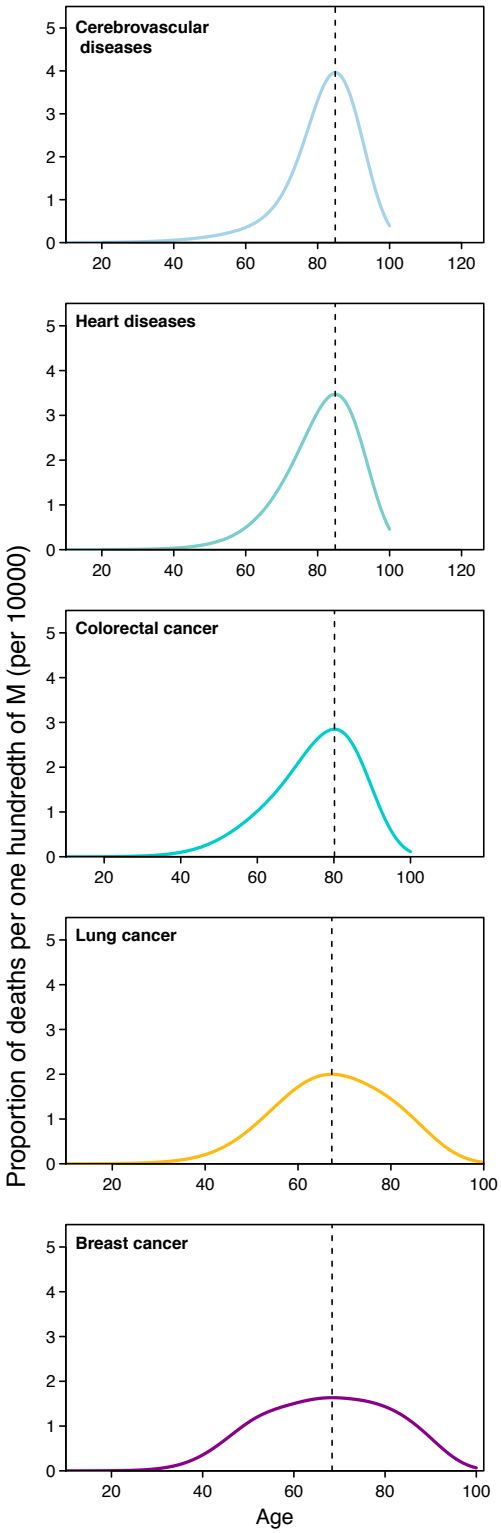
van RAALTE, A. and CASWELL, H. (2013). Perturbation analysis of indices of lifespan variability. *Demography* 50: 1615-1640.

van RAALTE, A., KUNST A. E., DeBOOSERE, P., et al (2011). More variation in lifespan in lower educated groups: Evidence from 10 European countries. *International Journal of Epidemiology* 40: 1703-1714.

van RAALTE, A., MARTIKAINEN, P., and MYRSKYLÄ, M. (2014). Lifespan variation by occupational class: compression or stagnation over time? *Demography* 51(1): 73-95.

WILMOTH, J. R. and HORIUCHI, S. (1999). Rectangularization revisited: Variability of the age at death within human populations. *Demography* 36(4): 475-495.

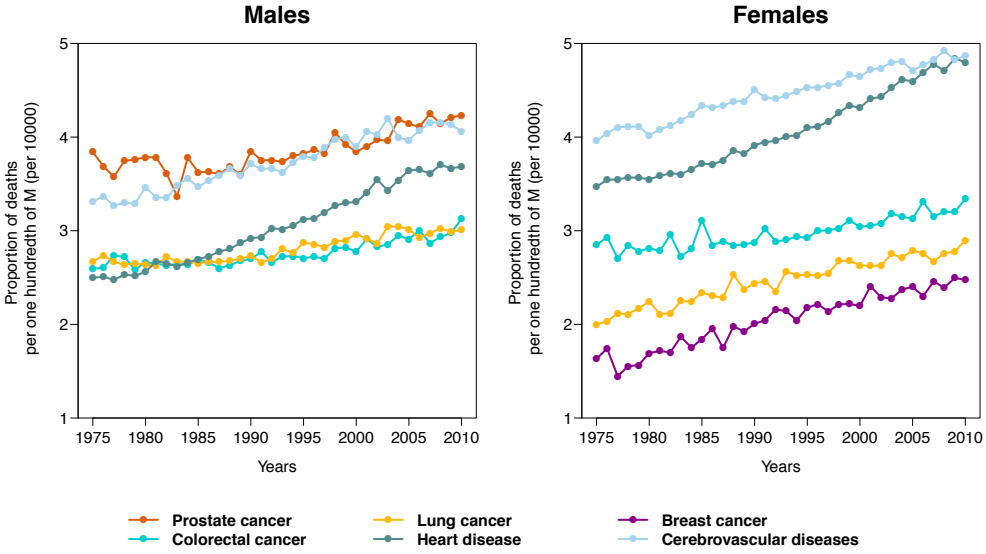
Figure 1: Rescaled lifespan distributions for five leading causes of death among Canadian females in 1975



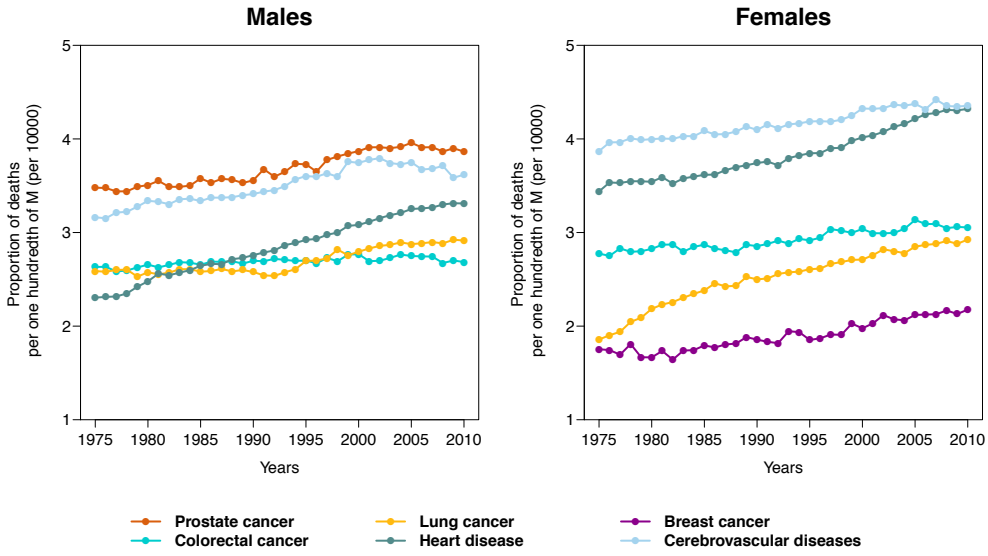
Source: Author's calculations based on the Canadian Vital Statistics Death Database and the Human Mortality Database.

Figure 2: Relative variation of cause-specific lifespan distributions for five leading causes of death, 1975-2011

a. Canada



b. U. S.



Source: Author's calculations based on the Canadian Vital Statistics Death Database, the U.S. National Vital Statistics System Death Database, and the Human Mortality Database.