Cause of death decomposition of cohort survival comparisons: TCAL

Vladimir Canudas-Romo, Timothy Adair and Stefano Mazzuco

Abstract

Period and cohort life expectancy correspond to mortality information from a given year or from one specific cohort. Although widely used, they have limitations, and their time-trends show disparities. If the interest is to know the mortality experience of a population, these two measures will hide the historical mortality path experienced by all cohorts present at a given time. The Truncated Cross-average Length of Life, or TCAL, is a period measure including all available cohort mortality information, irrespective of whether or not cohorts have complete data. This demographic tool is particularly useful for comparing cohort mortality between populations, and in this project we extend it to comparisons by causes of death. The strength of the approach is that it allows us to identify mortality differences in cohorts which are currently alive, as well as to identify which ages and which causes of death contribute to the mortality differentials between populations.

Cause of death decomposition of cohort survival comparisons: TCAL

Introduction

The length of life of a population is most commonly estimated by calculating life expectancy. Although widely used this measure also has limitations, the most important being that it constructs the mortality schedule of a synthetic cohort following the mortality schedule of a given time and, as such, this measure hides the mortality history of cohorts present at a given time.[Preston et al. 2001] To solve this problem the cross average length of life, or CAL, was developed to include the mortality history of a population.[Bouard 1986 and Guillot 2003] This was further extended for countries with only partial mortality information using the truncated version or TCAL.[Canudas-Romo & Guillot 2015]

Demographic measures are meaningful when compared across time or between populations. Life expectancy, CAL, and TCAL are summary measures and cross-country and time differences are determined by more specific – and often intertwined, components. Thus, decomposition methods become useful tool to quantify specific contributions to longevity measures. In this sense, some methods have been suggested to disentangle the age- and cause of death- contribution to differences in life expectancy Andreev (1982) stepwise replacement decomposition, Arriaga (1984) discrete decomposition and Pollard's (1982, 1988), Vaupel & Canudas-Romo (2003), and Horiuchi et al (2008) continuous decomposition. The methodology to decompose CAL and TCAL includes one extra element by disentangling age- and cohorts-contributions to the disparity between two populations.[Canudas-Romo & Guillot 2013] This methodology has been successfully used to compare populations (whites vs blacks in the USA, Nepomuceno & Van Ralte 2018, and an entire region of the world, Nepomuceno & Canudas-Romo 2016). However, the methodology is yet to be extended to calculate cause of death decomposition.

The current study aims to present a new method for decomposing the differences in TCALs by causes of death. The data for such undertaking requires bridging several International Classification of Diseases over time, to have a sufficiently long enough history of mortality. Furthermore, the cause of death information is usually only available for 5-year age groups, which hinders the possibility to identify specific cohorts.[WHO 2018] Both of these limitations were tackled in this study by focusing on a selected group of high developed countries which compete for the highest life expectancies in the world,

namely Japan, New Zealand, Norway, Sweden, and Switzerland.[Adair, Kippen and Lopez 2017] Of particular interest in our study will be to compare all these countries competing for the highest longevity with Japan, which for some decades has had the highest life expectancy in the world for both sexes [Oeppen & Vaupel 2002] but more recently had a slowing mortality decline because of relatively high ischemic heart disease (IHD), lung cancer, chronic respiratory disease and suicide death rates.[Adair, Kippen and Lopez 2017]] The second issue of the single-age causes of death information was solved by distributing the period 5-year data into single age using a well-established method developed by Rizzi et al. (2015).

Data

We use two sources of information: life tables and causes of death. For the first source we used the single age-specific death rates from the Human Mortality Database (HMD), while for the second we used the World Health Organization (WHO) database and the Global Burden of Diseases databases (GBD). All these sources are free available.

The HMD rigorously scrutinizes for quality control the data provided. WHO is an archive of the causes of death information provided by country members, however data quality is not part of the tasks of this institution. The GBD has redistributed causes of death information and is harmonizing data across time and countries. However the latter only offers a short time series of information on causes of death back in time. We are testing the use of the WHO data for countries where we have both WHO and GBD as an opportunity carry on sensitivity analysis of our results and test the possibilities for extending the historical information available. The countries that we are investigating are among those with the most complete causes of death series and thus we don’t expect great disparities between one database or the other. To further, take into account the uncertainty of our causes of death results, confidence intervals have been calculated.

Methods

TCAL is a period age-aggregated measure that summarizes historical mortality information about all cohorts present at a given time and it is not limited to populations with complete cohort mortality data. [Canudas-Romo and Guillot, 2015] TCALs can be compared to enable investigation of survival disparities between populations including all the available information for cohorts present in a given time.

TCAL(t,Y1) is defined between two years: *t* is the year when we are interested to estimate the measure, and Y_1 is the earliest year for the available mortality series; and computed as,

$$TCAL(t, Y_1) = \int_0^{\omega} l_c(x, t, Y_1)dx, \qquad (1)$$

where $l_c(x, t, Y_1)$ is the survival function for cohorts reaching age *x* in year *t*, which were born in year *t-x*. For cohorts born before the year Y_1, only partial cohort mortality data are available, we then assume a set of death rates for years before the year Y_1. The selection of death rates for years before Y_1, is an arbitrary one, and *TCAL* values will vary depending of this selection. However, as explained by Canudas-Romo and Guillot (2015), the selection of death rates for years before Y_1 have no weight in the comparisons between *TCALs* of two populations, as long as these rates before year Y_1 are used in both populations. Thus, as our interest is in the mortality gap between populations, the *TCAL* differences will be consistent if we use the same set of death rates in years before Y_1 for all examined countries. In order to eliminate any confounding effects of death rates before the year Y_1, we assume death rates equal to zero for all examined countries for years before Y_1, focusing our comparisons only on the cohort information available.

To compare two populations at time *t*, both *TCALs* must be truncated at the same year Y_1, which means that mortality series for all countries and for Japan must start at Y_1. Thus, the *TCAL* comparison between each country and Japan reveals which population experienced higher mortality levels in historical mortality data. Lower *TCAL* values correspond to populations that experienced higher cohort mortality levels. The difference in *TCALs* between Japan (JPN) and country (*i*) is then

$$TCAL_{JPN}(t, Y_1) - TCAL_i(t, Y_1) = \int_0^{\omega} l_{JPN}(x, t, Y_1) - l_i(x, t, Y_1)dx, \qquad (2)$$

where the integral corresponds to cohorts aged from 0 to ω, or oldest cohorts of the open age interval, and present at time *t* and both populations have the same set of age-specific death rates in years before Y_1. The cohort survival differences on the right side of equation (2) allow us to identify the mortality contribution of each cohort present in year *t*. The difference between *TCALs* is comparable to differences in life expectancy in showing the number of years one population is lagging behind another.

As *TCAL* condenses the available cohort mortality history into one measure, equation (2) shows that any difference between *TCAL*s allows identification of the cohort-specific contributions to the mortality gap. Furthermore, age-cohort contributions, $\Delta(a, t-x, i)$, to the difference between $TCAL_i$ of population *i* and that of Japan $TCAL_{JPN}$ can be estimated as

$$\Delta(a, t-x, i) = \left[\frac{l(x,t,Y_1,i)+l(x,t,Y_1,JPN)}{2}\right] ln\left[\frac{{}_1p_a(t-x,JPN)}{{}_1p_a(t-x,i)}\right], \qquad (3)$$

where $l(x, t, Y_1, i)$ is the survival function for the cohort aged *x* at time *t* in population *i*, ${}_1p_a(t-x, i)$ is the probability of surviving from age *a* to *a+1* for the cohort born in year *t-x* in population *i*. Finally, instead of the integrals in equation (2), the summation over cohorts and ages of the age-cohort contributions, $\Delta(a, t-x, i)$, returns the difference in *TCALs*

$$TCAL_{JPN}(t, Y_1) - TCAL_i(t, Y_1) \approx \sum_{x=1}^{\omega}\sum_{a=0}^{x-1}\Delta(a, t-x, i). \qquad (4)$$

Through this decomposition, we compare mortality between birth cohorts born from different populations. To further include cause of death contribution we obtain the proportion of deaths at age *x* of cause *j* respect to all deaths, denoted as *c(x,i,j)* for population *i*. From these proportions, cause *j* specific probabilities of surviving from age *a* to *a+1* for the cohort born in year *t-x* in population *i*, or ${}_1p_a(t-x, i, j)$ were calculated as

$${}_1p_a(t-x, i, j) = \left[{}_1p_a(t-x, i)\right]^{c(x,i,j)}, \qquad (5)$$

and substituted in equation (3) to obtain the age- cohort- cause-specific contribution $\Delta(a, t-x, i)$ to the difference in *TCAL*s, changing equation (4) for

$$TCAL_{JPN}(t, Y_1) - TCAL_i(t, Y_1) \approx \sum_{j=1}^{n} \sum_{x=1}^{\omega} \sum_{a=0}^{x-1} \Delta(a, t-x, i, j). \quad (6)$$

with *n* different causes of death.

The main limitation of the method is data availability. In principal we would be interested to present as much cohort data as possible. However, for many regions of the world this is not possible. On top of data quantity constraints, the quality of the data has become better over time and thus any measure that has a cohort perspective will carry on some of that quality bias from its older information.

Since cause-of-death data comes in five years age groups with a 95+ as the last age class while our interest is identifying individual cohorts, we smoothed causes of death data using the methodology suggested by Rizzi et al (2015). R-package.

Confidence intervals are calculated through a stratified bootstrap algorithm (2000 iterations): this means that for each iteration, conditional to age and year, death counts by causes of death are resampled using a multinomial distribution and TCAL decomposition for all causes has been computed. In this way, for every simulation we have a complete set of cause-specific decomposition of TCAL difference, preserving the constraints of adding to the total difference.

Results

Figure 1 presents 6 Lexis surfaces for the *TCAL*s comparisons between the male populations of Japan and Sweden. The Lexis surfaces correspond to the Total overall difference as well as the cause-specific contributions to this difference. These diagrams facilitate comparisons of cohort survival between populations. Diagonally, the Figures illustrate the mortality pattern of each cohort from birth (or the age when data is first available) to old age during the period between 1950 and 2015 (year-range when data on life tables and causes of death coincide in both countries). During this period, Sweden had higher overall survival than Japan, captured in the negative *TCAL* difference of -0.3 years. The Lexis surface illustrates how each cohort contributes over age to this mortality gap. For example, a long-lasting effect of lower mortality at younger ages for Swedish cohorts from the 1950s and 1960s is seen as marked by the orange-pink color (Figure 1 for total difference in TCALs). The low mortality during infancy and childhood remained for several decades until the 1990s and 2000s when Japan lower mortality at birth is present; visually this can be seen in the diagonals showing the higher survival for Sweden (yellows and oranges) transferring to Japan (blues). By contrast, Swedish males born in the 1920s and 1930s faced a particular disadvantage when mortality data from 1950 until 2015 (between ages 70 to 90) are taken into account (again blue for Japanese survival advantage at older ages).

Research into Japanese historical mortality shows that declines in non-communicable diseases (NCDs) have been an important contributor to survival advantages compared with other countries. This analysis identifies that it is particularly cardiovascular diseases which alone contribute to 1.26 years of Japanese advantage (seen in Figure 1 for CVD by the blue dominating at ages 60+ and since the 1980s). But it is Sweden not Japan which has the highest TCAL of the two, thus other causes of death are contributing to this differential and turning the balance in favor of Sweden: Neoplasms -0.62, external causes (homicides, accidents and suicides) -0.15, diseases of the digestive system -0.22, diseases of the respiratory system -0.71. The latter two causes show a particularly interesting dominance of Swedish male with lower mortality for every single cohort present today (seen in the dominance of yellow in Figure 1).

Figure 2 complements this finding by including the cause-specific contribution to the difference in TCALs for Japan versus Sweden for all causes included, as well as their 95% Confidence Intervals (sizes of the bubbles). Additionally to this, Figure 2 also compares Japanese males versus Norwegians

(difference of 0.57 years), Swiss males (diff = -0.02 years), and New Zealand males (diff = 1.04 years). The extreme high CVD contribution found in the Japanese and New Zealand comparison is also present in these other country comparisons. Opposing this are the cancers and respiratory diseases which show negative results, or on favor of countries other than Japan. This is consistent with a recent comparison of period life expectancy in Japan with Australia that identified Japan having relatively high male lung cancer and chronic respiratory mortality which was attributed to Japan's higher smoking rates [Adair, Kippen and Lopez 2017]. Similar comparisons can be taken for other cause-contributions. It should be pointed out the small and consistent confidence intervals of our results which show that small uncertainty is obtained when including all the mortality history of the population present at a given time.

Conclusion

How useful is it to know that life expectancy is low or high, when this measure only corresponds to mortality of one given time or a specific cohort? Today, digital information allows us to have long time series of mortality readily available. We propose a cause of death decomposition methodology which includes all the available mortality information for cohorts present in two populations. Our next steps will include detail uses of this new methodology.

Reference

Adair T, Kippen R and Lopez A(2017). A recent comparative history of life expectancy trends in Japan and Australia. Paper presented in Cape Town, South Africa at the IUSSP population conference 2017.

Andreev E (1982). Metod komponent v analize prodoljitelnosty zjizni.[The method of components in the analysis of length of life]. *Vestnik Statistiki, 9*, 42-47.

Andreev E, Shkolnikov VM and Begun A. (2002). Algorithm for decomposition of differences between aggregate demographic measures and its application to life expectancies, healthy life expectancies, parity-progression ratios and total fertility rates. *Demographic Research, 7*, 499-522.

Arriaga E (1984). Measuring and explaining the change in life expectancies. *Demography, 21*(1), 83-96

Brouard N (1986). Structure et dynamique des populations. La pyramide des années à vivre, aspects nationaux et exemples régionaux [Structure and dynamics of populations. [The later-life pyramid, national aspects and regional examples], Espaces, Populations, Sociétés 4(2): 157-168.

Canudas-Romo V and Guillot M (2015). Truncated cross-sectional average length of life: A measure for comparing the mortality history of cohorts. *Population Studies, 69*(2), 147-159.

Global Burden of Diseases mortality database (2018). Avaialble at http://www.healthdata.org/gbd/data

Guillot M (2003). The cross-sectional average length of life (CAL): A cross-sectional mortality measure that reflects the experience of cohorts, Population Studies 57(1): 41–54.

Horiuchi S, Wilmoth JR and Pletcher SD (2008). A decomposition method based on a model of continuous change. *Demography*, *45*(4), 785-801.

Human Mortality Database 2018. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.moratlity.org

Nepomuceno M and Canudas-Romo V (2016). Cohort Survival Comparison between Central-Eastern Europe and High-longevity countries: The TCAL Approach. Paper presented in Mainz, Germany at the EAPS population conference 2016.

Nepomuceno M and van Raalte AA (2018). Rising Midlife Mortality Among U.S. Whites: A Cohort Survival Comparison. Paper presented in Denver, Colorado at the PAA population conference 2018.

Oeppen, J & Vaupel JW (2002). Broken limits to life expectancy. Science 296(5570):1029-1031.

Pollard JH (1982). The expectation of life and its relationship to mortality. *Journal of the Institute of Actuaries*. *109*(2), 225-240.

Pollard JH (1988). On the decomposition of changes in expectation of life and differentials in life expectancy. *Demography*. *25*(2), 265-276.

Preston S, Heuveline P and Guillot M (2001). *Demography: measuring and modeling population processes*. Blackwell Publishers.

Rizzi S, Gampe J & Eilers PHC (2015). Efficient estimation of smooth distributions from coarsely grouped data. American Journal of Epidemiology 182(2):138-147.

Vaupel JW and Canudas-Romo V. (2003). Decomposing change in life expectancy: A bouquet of formulas in honor of Nathan Keyfitz's 90th birthday. *Demography, 40*(2), 201-216.

World Health Organization mortality Database (2018). Available at

www.who.int/healthinfo/mortality_data/en/

Figure 1. Lexis surface of male cohort survival comparisons between Japan and Sweden and cause of death contribution to their difference in TCALs.

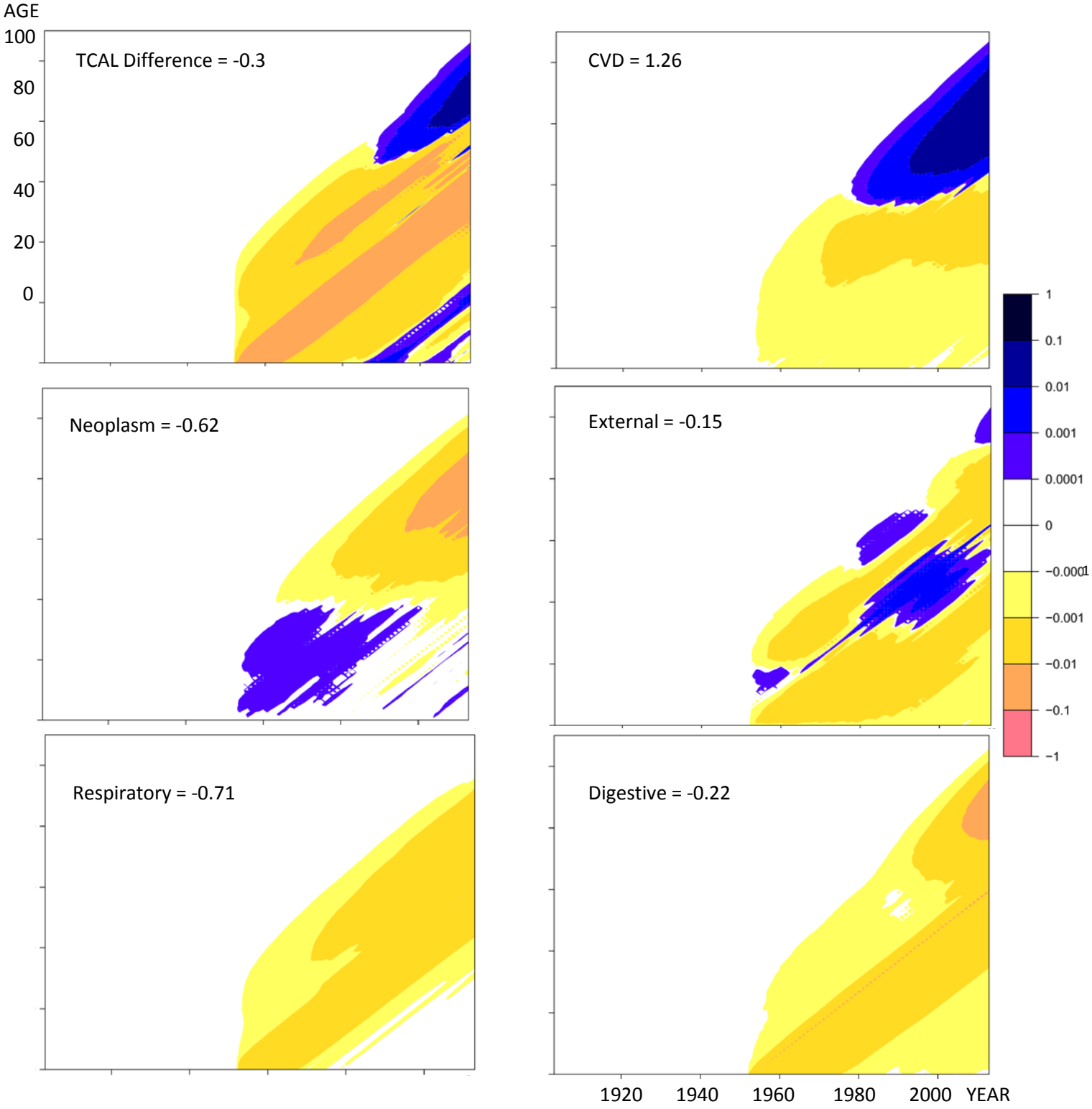
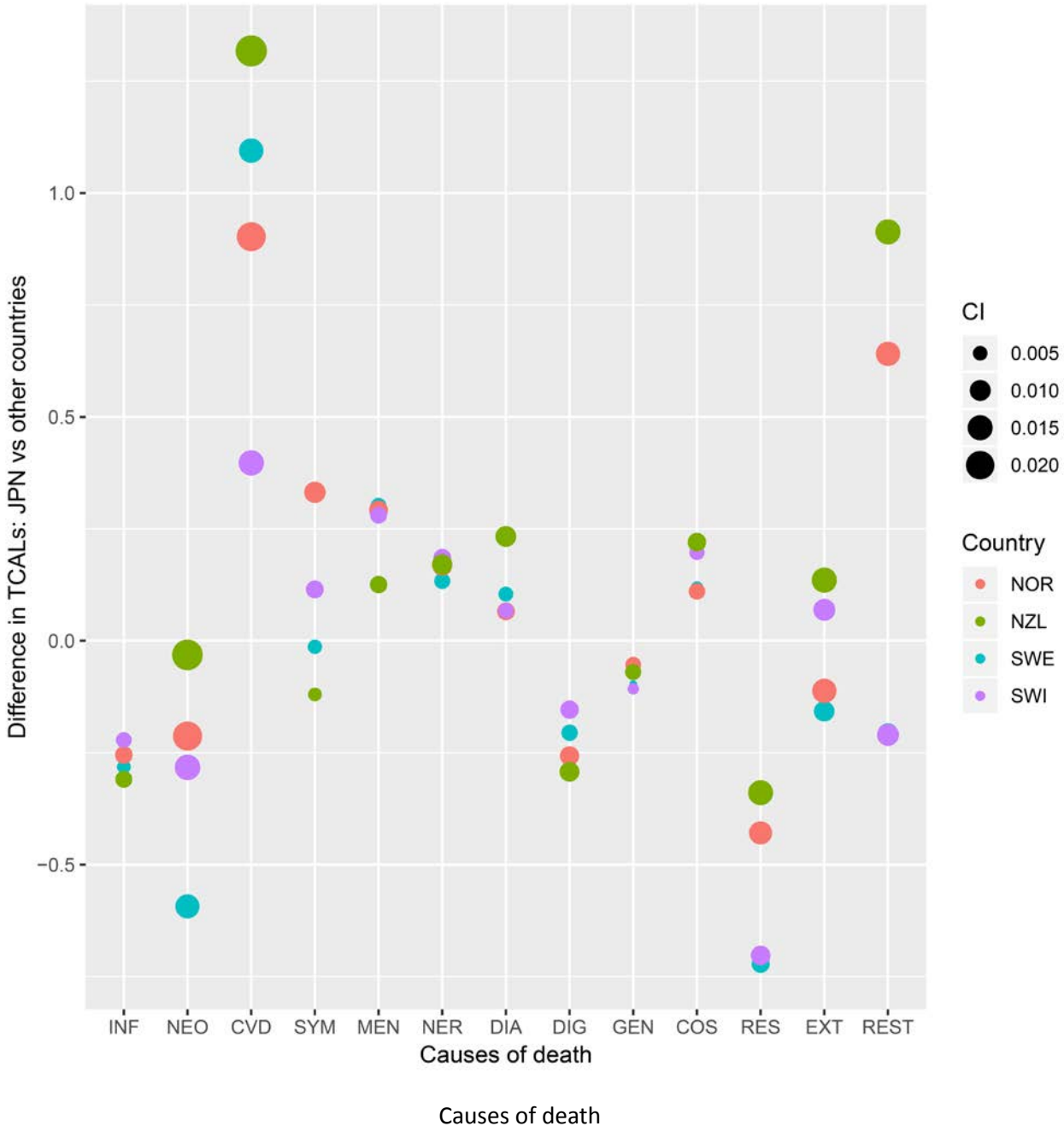


Figure 2. Cause of death contribution to the differences in TCALs between Japan and other high longevity countries male populations by causes of death in 2015.



Causes of death

Sources: WHO mortality database and Human Mortality Database.

Notes for Figure 2.
1) Countries included are: NOR-Norway, NZL-New Zealand, SWE-Sweden, and SWI-Switzerland
2) Causes of death: INF- Infectious and parasitic diseases, NEO-Neoplasms, CVD-Diseases of the circulatory system, SYM-Symptoms not elsewhere classified, MEN-Mental and behavioural disorders, NER-Diseases of the nervous system, DIA-Endocrine, nutritional and metabolic diseases, DIG-Diseases of the digestive system, GEN-Diseases of the genitourinary system congenital malformations, RES-Diseases of the respiratory system, EXT-External causes, REST- causes not included in the others.
3) CI corresponds to confidence intervals for the results at X+0.005, +0.010, +0.015 and +0.020