Title

Uncertainty quantification of years of potential life lost—based estimates from mortality data summarized as death counts within age intervals

Authors

Xu, Jay J

Belin, Thomas R

Ramirez, Christina M

Publication Date

2021-03-01

DOI

10.1016/j.annepidem.2020.11.002

Peer reviewed

Uncertainty Quantification of Years of Potential Life Lost-Based Estimates from Mortality Data Summarized as Death Counts Within Age Intervals

Jay J. Xu[a,*], Thomas R. Belin[a,b], Christina M. Ramirez[a]

[a] *Department of Biostatistics, Jonathan and Karin Fielding School of Public Health, University of California, Los Angeles, Los Angeles, CA, United States of America*
[b] *Department of Psychiatry and Biobehavioral Sciences, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, United States of America*

Keywords: Epidemiology, Monte Carlo Simulation, Public Health, Uncertainty Quantification, Years of Potential Life Lost

Introduction

Disease burden is often encapsulated in mortality rates. However, given the connection between age and death, mortality rates typically are most influenced by deaths among older adults. Alternative epidemiological measures of mortality burden have been proposed that explicitly weight deaths at younger ages more heavily [1, 2, 3]. One such measure is years of potential life lost [4] (YPLL), which has been used in diverse contexts to quantify premature mortality burden (e.g. [5, 6]). YPLL for an individual fatality i is defined to be an upper reference age $\mathcal{A}$ – usually close to a broadly applicable life expectancy – minus age at death a_i if the difference is positive and zero otherwise.

$$\text{YPLL}_i = \max\{\mathcal{A} - a_i, 0\} \tag{1}$$

Mortality data are often summarized as death counts within age intervals due

*Corresponding author
Email address: jayxu33@ucla.edu (Jay J. Xu)

to privacy issues, such as national death counts from the COVID-19 pandemic (e.g. [7]), precluding exact calculation of YPLL values. In such settings, the standard approach is to operationally replace each individual's age at death with its interval midpoint for the purposes of calculating aggregate YPLL, a method that implicitly assumes that ages at death within intervals are uniformly distributed [8]. However, no attempt is typically made in applied epidemiological studies using this "midpoint method" to quantify the uncertainty in YPLL-based estimates due to the administrative interval censoring of ages at death (e.g. [9, 10, 11, 12]). To address this limitation, we propose a straightforward and easily implementable Monte Carlo (MC) simulation procedure that yields point estimates of YPLL-based quantities with associated uncertainty quantification consistent with the assumption of uniformly distributed ages at death within intervals.

Proposed Monte Carlo Simulation Procedure

Suppose we wish to estimate a scalar quantity θ that is a function F of individual YPLL values. Let $d_1, d_2, ..., d_K$ denote the death counts in K mutually exclusive, collectively exhaustive, and chronologically-ordered age intervals $(L_1 = 0, U_1), (L_2, U_2)..., (L_K, U_K = \infty)$, respectively, from the study population. The proposed procedure simulates plausible ages at death for each individual at each MC iteration, and an estimate of θ is obtained from their corresponding YPLL values. The collection of MC estimates of θ are then combined into an overall point and interval estimate for θ, reflecting the uncertainty in the true ages at death. Observe that choosing $\mathcal{A}$ such that $\mathcal{A} < L_K$ obviates the need to simulate ages at death from (L_K, ∞). The proposed procedure can be described as follows:

1. Let $\{(L_1, U_1), (L_2, U_2)..., (L_r, U_r)\}$ denote the maximal set of intervals

such that $L_1 < L_2 < \cdots < L_r \leq \mathcal{A}$, and $r < K$. For each death i in interval (L_j, U_j), $j = 1, 2, ..., r$, independently simulate an age at death a_{ij} from a continuous uniform distribution corresponding to the interval:

$$\tilde{a}_{ij} \overset{ind}{\sim} \mathcal{U}\left(L_j, U_j\right), i = 1, 2, ..., d_j \tag{2}$$

2. Compute a point estimate of θ from the $\sum_{j=1}^{r} d_j$ simulated ages:

$$\hat{\theta} = F\left(\left\{\max\left\{\mathcal{A} - \tilde{a}_{ij}, 0\right\}\right\}_{j=1,i=1}^{r,d_j}\right) \tag{3}$$

3. Repeat steps 1 and 2 to obtain $\mathcal{B}$ point estimates of θ, denoted $\left\{\hat{\theta}^{(1)}, \hat{\theta}^{(2)}, ..., \hat{\theta}^{(\mathcal{B})}\right\}$. The overall point estimate of θ is given by:

$$\hat{\theta} = \frac{1}{\mathcal{B}} \sum_{b=1}^{\mathcal{B}} \hat{\theta}^{(b)} \tag{4}$$

Furthermore, a $(1 - \alpha) \times 100\%$ interval estimate of θ is given by the $\frac{\alpha}{2}$ and $1 - \frac{\alpha}{2}$ quantiles of $\left\{\hat{\theta}^{(1)}, \hat{\theta}^{(2)}, ..., \hat{\theta}^{(\mathcal{B})}\right\}$:

$$\mathcal{I}_{1-\alpha} = \left(\hat{\theta}_{\frac{\alpha}{2}}, \hat{\theta}_{1-\frac{\alpha}{2}}\right) \tag{5}$$

where $\hat{\theta}_q$ denotes the $q \in (0, 1)$ quantile of $\left\{\hat{\theta}^{(1)}, \hat{\theta}^{(2)}, ..., \hat{\theta}^{(\mathcal{B})}\right\}$.

The proposed procedure and the midpoint method yield estimates of total YPLL equal in expectation, but challenges can arise when using the midpoint method to estimate age-standardized YPLL rates. The midpoint method requires the age intervals in the mortality data to be aligned with those in the study and standard population – which may require combining age intervals – to calculate age-specific YPLL rates in the study population that are then

applied to the standard population. If analyses are based on a coarser representation of the mortality data, study population, or standard population in order to align age intervals, it is expected that resulting estimates will be less accurate. If the age intervals cannot be aligned by combining age intervals, it is not straightforward how to proceed with the midpoint method. A key advantage of the proposed procedure is its flexibility; because simulated ages are continuous, their corresponding YPLL values can be aggregated with respect to an arbitrary collection of age intervals for which age-specific YPLL rates can be calculated and subsequently applied to the standard population.

Illustrative Data Example

We demonstrate the proposed procedure on COVID-19 death counts by age and sex from the Directorate-General of Health (DGS) in Portugal [13], calculating a point and 95% interval estimate of the age-adjusted male-to-female YPLL rate ratio (RR). The cumulative death counts by age and sex as of July 30, 2020 are presented in the July 31, 2020 daily DGS epidemiological report (pg. 4), included as Supplementary File 1. While the number of male (867) and female (868) deaths are almost equal, a greater concentration of male deaths occur at earlier ages, which will be reflected in YPLL calculations.

We use 2020 estimates of the sex-stratified age distribution of Portugal from `PopulationPyramid.com` [14] (summarized in 5-year intervals up to age 100) to calculate sex- and age-specific YPLL rates, and we use their combined overall age distribution as the standard population (see Supplementary File 2 for the corresponding data). Direct age adjustment [15] is used to calculate the male and female age-adjusted YPLL rates, and their quotient is the age-adjusted male-to-female YPLL RR. Consistent with the OECD Health Statistics standard [16], we use an upper reference age of $\mathcal{A} = 70$ years, and we perform $\mathcal{B} = 1000$

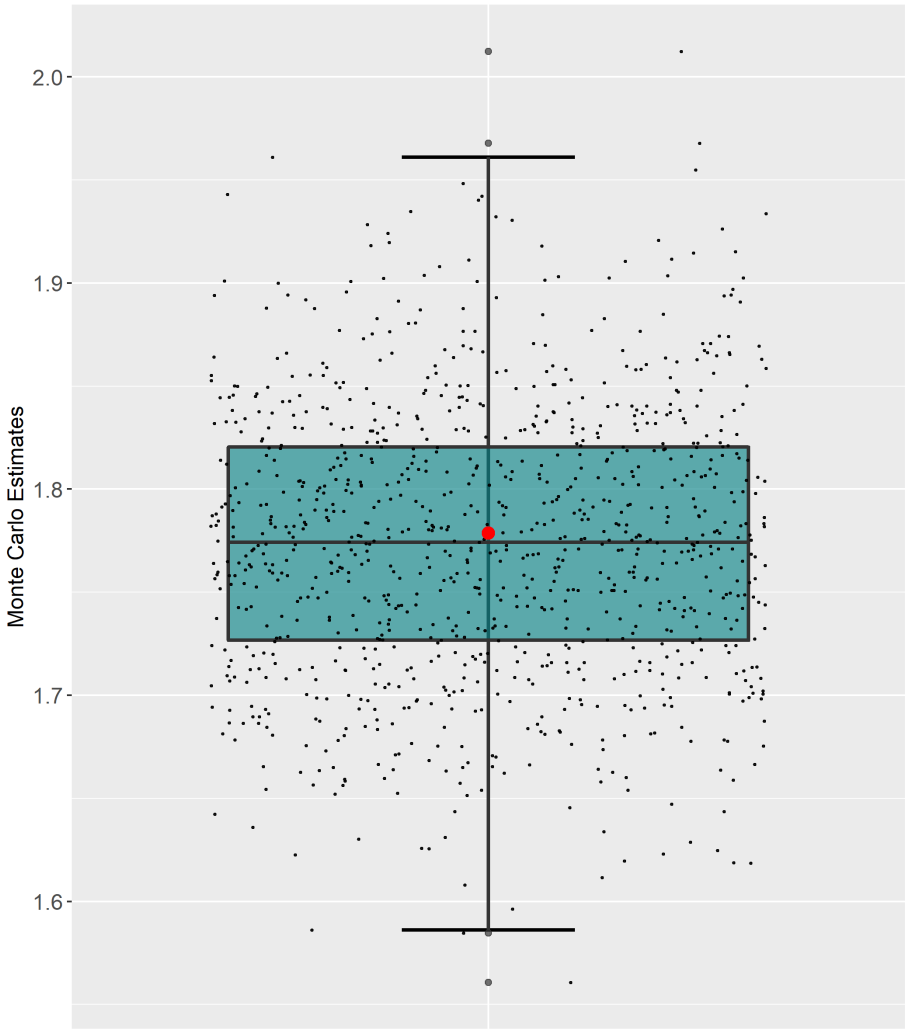

Figure 1: Boxplot of the 1,000 Monte Carlo estimates of the age-adjusted male-to-female YPLL rate ratio with respect to cumulative COVID-19 deaths in Portugal as of July 30, 2020 accompanied by jittered points denoting the individual Monte Carlo estimates. The red dot represents the estimate obtained from the midpoint method.

MC iterations.

Figure 1 displays a boxplot of the 1000 MC estimates of the age-adjusted male-to-female YPLL RR along with the midpoint method estimate for comparison. The overall point estimate of the age-adjusted male-to-female YPLL RR

is 1.77 with an associated 95% interval estimate of (1.65, 1.90), meaning that after accounting for differences in the male and female population age distributions, males experienced COVID-19-attributable YPLL at a rate estimated to be 77% higher than females with a 95% interval estimate for the true relative increase of 65-90%. The midpoint method, which requires the Portugal population age intervals to be combined into 10-year intervals, yields an estimate of 1.78 with no uncertainty quantification. `R` code implementing this reproducible illustrative data example is provided as Supplementary File 3.

Conclusions

In summary, we developed a straightforward and easily implementable MC simulation procedure to quantify the uncertainty of YPLL-based estimates from mortality data summarized as death counts within age intervals. We demonstrated the procedure on COVID-19 mortality data from Portugal, calculating a point and 95% interval estimate for the age-adjusted male-to-female YPLL RR with respect to cumulative deaths as of July 30, 2020. We encourage future applied epidemiological research quantifying the mortality burden of public health problems in terms of YPLL to quantify the uncertainty of their estimates.

References

[1] R. P. Wise, J. R. Livengood, R. L. Berkelman, R. A. Goodman, Methodological Alternatives for Measuring Premature Mortality, American Journal of Preventive Medicine 4 (5) (1988) 268–273.

[2] L. Bonneux, How to measure the burden of mortality?, Journal of Epidemiology & Community Health 56 (2) (2002) 128–131.

[3] B. Devleesschauwer, A. H. Havelaar, C. M. De Noordhout, J. A. Haagsma, N. Praet, P. Dorny, L. Duchateau, P. R. Torgerson, H. Van Oyen, N. Spey-

broeck, Calculating disability-adjusted life years to quantify burden of disease, International Journal of Public Health 59 (3) (2014) 565–569.

[4] J. W. Gardner, J. S. Sanborn, Years of Potential Life Lost (YPLL) — What Does it Measure?, Epidemiology (1990) 322–329.

[5] M. D. Wong, A. K. Chung, W. J. Boscardin, M. Li, H.-J. Hsieh, S. L. Ettner, M. F. Shapiro, The Contribution of Specific Causes of Death to Sex Differences in Mortality, Public Health Reports 121 (6) (2006) 746–754.

[6] S. Darke, C. Marel, K. L. Mills, J. Ross, T. Slade, M. Tessson, Years of potential life lost amongst heroin users in the Australian Treatment Outcome Study cohort, 2001–2015, Drug and Alcohol Dependence 162 (2016) 206–210.

[7] Covid-19 México, `https://datos.covid-19.conacyt.mx/` (Accessed July 31, 2020).

[8] J. Romeder, J. McWhinnie, Potential Years of Life Lost Between Ages 1 and 70: An Indicator of Premature Mortality for Health Planning, International Journal of Epidemiology 6 (2) (1977) 143–151.

[9] M. Arcà, F. Di Orio, F. Forastiere, C. Tasco, C. Perucci, Years of Potential Life Lost (YPLL) before Age 65 in Italy, American Journal of Public Health 78 (9) (1988) 1202–1205.

[10] J. W. Curran, W. M. Morgan, A. M. Hardy, H. W. Jaffe, W. W. Darrow, W. R. Dowdle, The epidemiology of AIDS: current status and future prospects, Science 229 (4720) (1985) 1352–1357.

[11] V. A. Albert, H. K. Koh, A. C. Geller, D. R. Miller, M. N. Prout, R. A. Lew, Years of potential life lost: Another indicator of the impact of cuta-

neous malignant melanoma on society, Journal of the American Academy of Dermatology 23 (2) (1990) 308–310.

[12] A. K. Mitra, M. Payton, N. Kabir, A. Whitehead, K. N. Ragland, A. Brown, Potential Years of Life Lost Due to COVID-19 in the United States, Italy, and Germany: An Old Formula with Newer Ideas, International Journal of Environmental Research and Public Health 17 (12) (2020) 4392.

[13] Relatório de Situação, https://covid19.min-saude.pt/relatorio-de-situacao/ (Accessed July 31, 2020).

[14] PopulationPyramid.net: Population Pyramids of the World from 1950 to 2100, https://www.populationpyramid.net/ (Accessed July 31, 2020).

[15] D. D. Celentano, M. Szklo, Gordis Epidemiology, Elsevier, 2019.

[16] Potential years of life lost, https://data.oecd.org/healthstat/potential-years-of-life-lost.htm (Accessed July 31, 2020).