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

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

[a]Department of Biostatistics, Jonathan and Karin Fielding School of Public Health, University of California, Los Angeles

[b]Department of Psychiatry and Biobehavioral Sciences, David Geffen School of Medicine, University of California, Los Angeles

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 epidemiologic measures of mortality burden have been proposed that explicitly weight deaths at younger ages more heavily [1–3]. One such measure is years of potential life lost [4] (YPLL), which has been used in diverse contexts to quantify premature mortality burden [5,6]. YPLL for an individual fatality i is defined to be an upper reference age $\mathscr{A}$—typically close to a broadly applicable life expectancy—minus age at death a_i if the difference is positive and zero otherwise.

$$YPLL_i = \max\{\mathscr{A} - a_i, 0\} \tag{1}$$

Mortality data are often summarized as death counts within age intervals because of privacy issues, such as national death counts from the coronavirus disease 2019 (COVID-19) pandemic [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 purpose 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

*Corresponding author. Department of Biostatistics, Jonathan and Karin Fielding School of Public Health, University of California, Los Angeles, CA 90095. jayxu33@ucla.edu (J.J. Xu).

in applied epidemiologic studies using this "midpoint method" to quantify the uncertainty in YPLL-based estimates because of the administrative interval censoring of ages at death [9–12]. 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 $d_1, d_2, \ldots, d_K$ denote the death counts in K mutually exclusive, collectively exhaustive, and chronologically ordered age intervals $(L_1 = 0, U_1), (L_2, U_2)\ldots,(L_K, U_K = \infty)$, 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 is then summarized as an overall point and interval estimate of θ, reflecting the uncertainty in the true ages at death. Observe that choosing $\mathscr{A}$ such that $\mathscr{A} \leq L_K$ obviates the need to simulate ages at death from (L_K, ∞). The proposed procedure can be described as follows:

1. Let $\{(L_1, U_1), (L_2, U_2),\ldots, (L_r, U_r)\}$ denote the maximal set of intervals such that $L_1 < L_2 < \cdots < L_r < \mathscr{A}$, and $r < K$. For each death i in interval (L_j, U_j), $j = 1,2, \ldots,r$, independently simulate an age at death a_{ij} from a continuous uniform distribution corresponding to the interval:

$$\tilde{a}_{ij} \stackrel{ind}{\sim} \mathscr{U}(L_j, U_j), i = 1, 2, \ldots, d_j. \tag{2}$$

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

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

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

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

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

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

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

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 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 Directorate General of Health epidemiologic report (pg. 4), included as Supplementary File 1. Although the number of male ($n = 867$) and female ($n = 868$) deaths are almost equal, a greater concentration of male deaths occurs at earlier ages, which will be reflected in YPLL calculations.

We use 2020 estimates of the sex-stratified age distribution of Portugal from PopulationPyramid.net [14] (summarized in 5-year intervals up to age 100 years) 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 Organisation for Economic Co-operation and Development Health Statistics standard [16], we use an upper reference age of $\mathscr{A} = 70$ years, and we perform $\mathscr{B} = 1000$ 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 before age 70 at a rate estimated to be 77% higher than females with a 95% interval estimate of 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 epidemiologic research quantifying the mortality burden of public health problems in terms of YPLL to quantify the uncertainty of their estimates.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

[1]. Wise RP, Livengood JR, Berkelman RL, Goodman RA. Methodological alternatives for measuring premature mortality. Am J Prev Med1988;4(5):268–73. [PubMed: 3224004]

[2]. Bonneux LHow to measure the burden of mortality?J Epidemiol Community Health2002;56(2):128–31. [PubMed: 11812812]

[3]. Devleesschauwer B, Havelaar AH, De Noordhout CM, Haagsma JA, Praet N, Dorny P, et al.Calculating disability-adjusted life years to quantify burden of disease. Int J Public Health2014;59(3):565–9. [PubMed: 24752429]

[4]. Gardner JW, Sanborn JS. Years of potential life lost (YPLL) — What does it measure? Epidemiology1990:322–9. [PubMed: 2083312]

[5]. Wong MD, Chung AK, Boscardin WJ, Li M, Hsieh HJ, Ettner SL, et al.The contribution of specific causes of death to sex differences in mortality. Public Health Rep2006;121(6):746–54. [PubMed: 17278410]

[6]. Darke S, Marel C, Mills KL, Ross J, Slade T, Tessson M. Years of potential life lost amongst heroin users in the Australian Treatment Outcome Study cohort, 2001–2015. Drug Alcohol Depend2016;162:206–10. [PubMed: 27021806]

[7]. Covid-19 Mexico. https://datos.covid-19.conacyt.mx/. [Accessed 31 July 2020].

[8]. Romeder J, McWhinnie J. Potential years of life lost between ages 1 and 70: an indicator of premature mortality for health planning. Int J Epidemiol1977;6(2):143–51. [PubMed: 892979]

[9]. Arca M, Di Orio F, Forastiere F, Tasco C, Perucci C. Years of potential life lost (YPLL) before age 65 in Italy. Am J Public Health1988;78(9):1202–5. [PubMed: 3407820]

[10]. Curran JW, Morgan WM, Hardy AM, Jaffe HW, Darrow WW, Dowdle WR. The epidemiology of AIDS: current status and future prospects. Science1985;229(4720):1352–7. [PubMed: 2994217]

[11]. Albert VA, Koh HK, Geller AC, Miller DR, Prout MN, Lew RA. Years of potential life lost: another indicator of the impact of cutaneous malignant melanoma on society. J Am Acad Dermatol1990;23(2):308–10. [PubMed: 2212128]

[12]. Mitra AK, Payton M, Kabir N, Whitehead A, Ragland KN, Brown A. Potential years of life lost due to COVID-19 in the United States, Italy, and Germany: an old formula with newer ideas. Int J Environ Res Public Health2020;17(12): 4392.

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

[14]. PopulationPyramid.net. Population pyramids of the world from 1950 to 2100. https://www.populationpyramid.net/. [Accessed 31 July 2020].

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

[16]. Potential years of life lost. https://data.oecd.org/healthstat/potential-years-of-life-lost.htm. [Accessed 31 July 2020].
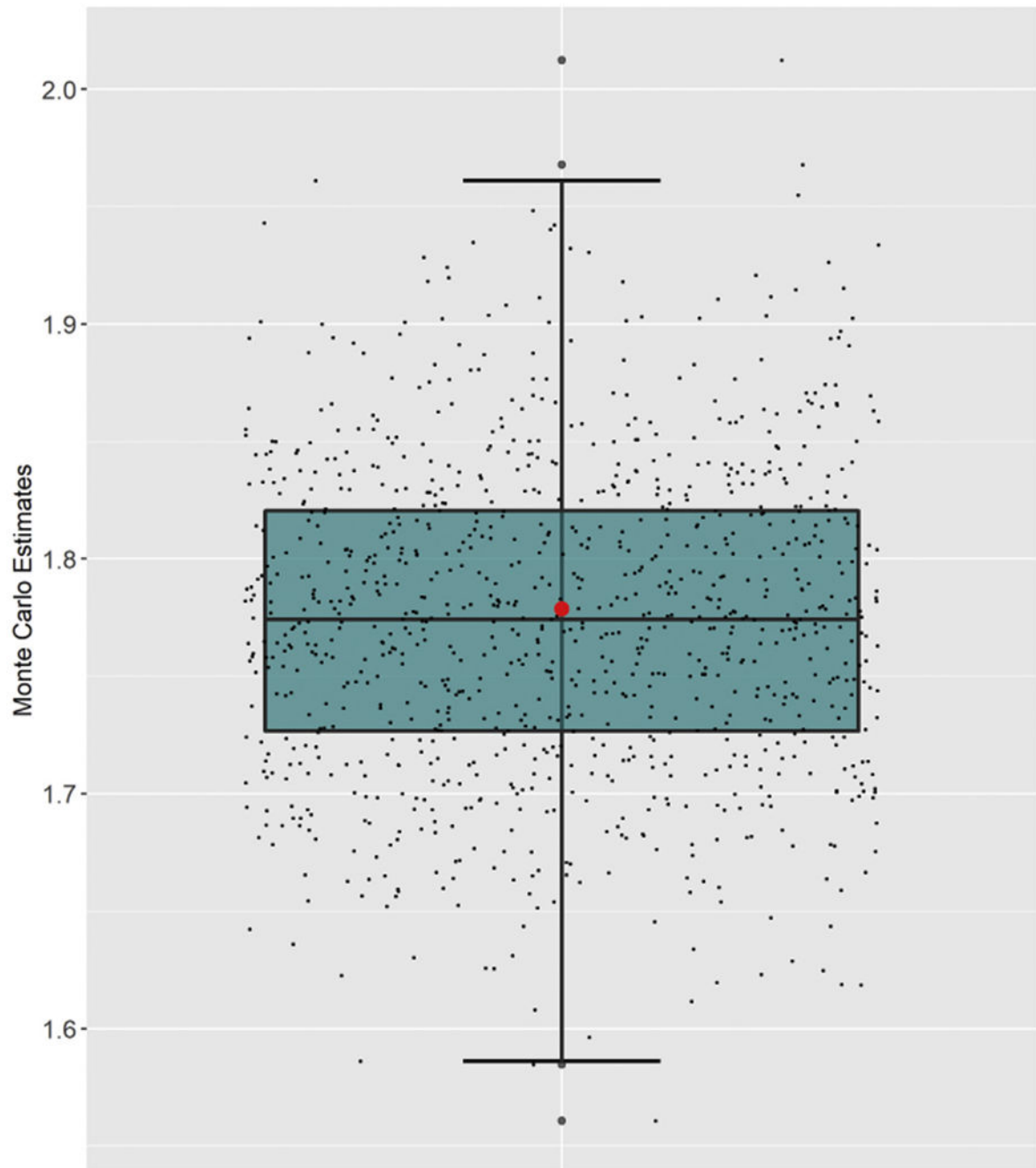
Fig. 1.
Boxplot of the 1000 Monte Carlo estimates of the age-adjusted male-to-female YPLL rate ratio with respect to cumulative COVID-19 deaths in Portugal as of July 30, 2020, accompanied by jittered points denoting the individual Monte Carlo estimates. The red dot represents the estimate obtained from the midpoint method. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)