

RESEARCH

Open Access



# Generating age-specific mortality statistics from incomplete death registration data: two applications of the empirical completeness method

Tim Adair<sup>1\*</sup>  and Alan D Lopez<sup>2</sup>

## Abstract

**Background:** The study aims to assess two approaches that apply the empirical completeness method to generate age-specific mortality statistics from incomplete death registration systems.

**Methods:** We use the empirical completeness method to calculate all-age death registration completeness, which is used with a model life table to generate mortality statistics and age-specific completeness using (1) the conventional method and (2) the equivalent deaths method. The results are compared with a capture-recapture (C-RC) study and three alternative mortality estimates for Brazilian states, and C-RC studies in Thailand, Oman and Vietnam, which independently estimate the level and age pattern of mortality or completeness.

**Results:** The empirical completeness method produces similar estimates of all-age completeness of registration to the C-RC studies. Compared with C-RC studies, at 15-59 years, the conventional method's estimates of mortality and completeness are more concordant, while at 60-84 years the equivalent death method's estimates are closer. Estimates of life expectancy from the two approaches each have similar concordance with the C-RC studies. For male adult mortality in Brazilian states, there is relatively strong average correlation of this study's estimates with three alternative estimates.

**Conclusions:** The two approaches produce mortality statistics from incomplete data that are mostly concordant with C-RC studies, and can be most usefully applied to subnational populations.

**Keywords:** Mortality, Vital statistics, Civil registration, Data quality, Completeness, Capture-recapture

## Introduction

Routine, timely and accurate indicators of the level and pattern of mortality in a population are best informed by a complete civil registration and vital statistics (CRVS) system. Summary all-cause and cause-specific mortality indicators produced by CRVS data track progress

towards national and subnational goals; 16 Sustainable Development Goals (SDGs) and 24 corresponding indicators require data that are most effectively produced by an accurate and timely CRVS system [1]. These mortality statistics are reliant on registered deaths disaggregated by age and sex that enable, for example, age-standardisation of rates and production of life tables. Analysis of the risk of mortality by age groups and sex can also reveal much of policy relevance about a population's health, even in the absence of reliable cause of death data.

\* Correspondence: [timothy.adair@unimelb.edu.au](mailto:timothy.adair@unimelb.edu.au)

<sup>1</sup>Melbourne School of Population and Global Health, The University of Melbourne, Level 5, Building 379, 207 Bouverie Street, Carlton, Victoria 3010, Australia

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

However, CRVS systems are incomplete in many countries, which limits their utilisation as evidence for health policy [2]. Use of unadjusted incomplete death registration data to generate mortality statistics will result in an under-estimation of these key mortality indicators. Measurement of the completeness of death registration is therefore important to understand how well deaths are captured by the CRVS system in national and subnational populations and to generate adjusted mortality statistics. Completeness of registration can be estimated using a range of methods, each with strengths and limitations; these include death distribution methods (DDMs), the empirical completeness method, and capture-recapture (C-RC) of deaths (or direct methods) [3–7]. The empirical completeness method, which models observed versus expected completeness based on modelling the key drivers of mortality levels in a population, has advantages over DDMs, including the use of relatively limited data that are readily available at the national and subnational levels, timely estimates according to when the most recent CRVS data are available (compared with the most reliable DDMs that estimate completeness for the most recent intercensal period) and a lack of reliance on often unrealistic assumptions about population dynamics (e.g. closed migration) that can especially bias subnational estimates [7].

DDMs and the empirical completeness method estimate a constant level of completeness above 5 years of age (or all ages for the empirical completeness method), and in isolation do not reveal information on age patterns of mortality that are necessary to produce mortality statistics [5–10]. A conventional method to generating life tables from incomplete death registration data is to estimate completeness of registration, calculating the  ${}_{45}q_{15}$  (probability of dying between age 15 and 60 years, or adult mortality) from death registration data which is then adjusted using this level of completeness, and inputting the adjusted  ${}_{45}q_{15}$  and an estimate of  ${}_5q_0$  (probability of dying from live birth to 5 years) into a model life table (MLT) [5]. MLTs represent the global range of levels and age-patterns of mortality rates, and are used to generate complete life tables from such data of suboptimal quality [11]. Another source of estimates of deaths, by age and sex, from which completeness can be calculated are the Global Burden of Disease (GBD) and United Nations World Population Prospects (UNWPP), which are produced for all countries [12, 13]. Both the GBD and UNWPP also estimate  ${}_5q_0$  and  ${}_{45}q_{15}$  from a range of available data using various methods (including completeness-adjusted registration data) and input these into MLTs. A potential drawback of this general approach to adjusting  ${}_{45}q_{15}$  is that if the level of completeness at 5 years and above used to adjust deaths

at ages 15–59 years is different from the true completeness for 15–59 years, then the adjusted  ${}_{45}q_{15}$  and subsequent final life table will be inaccurate. A further limitation is that the MLT may not reflect the true age pattern of local mortality.

C-RC methods, which match individual death data in a death registration (or reporting) system to another data source, can however estimate completeness of death registration for specific age groups and use this to adjust registration data and produce complete life tables [3, 14]. Importantly, C-RC studies' estimates of age-specific completeness of registration (and hence the true age pattern of mortality) do not rely on an assumed age pattern of mortality from a MLT. Reliable application of C-RC methods requires having two independent sources of data that can be successfully linked. However, many C-RC studies are time- and resource-intensive due to the data linkage process and some studies have required a new data collection independent from the registration system to be undertaken. Hence, C-RC studies of mortality registration data are relatively rare. Recent examples of such studies include those conducted in Brazil, Thailand, Vietnam, Brazil, Oman, Kiribati and Bohol, Philippines [14–20].

In summary, there is potential error in mortality statistics generated from incomplete death registration data. However, given the advantages of the empirical completeness method in estimating all-age completeness, it is worthwhile to assess its application to generate age-specific mortality statistics. This study examines two different methods that apply the empirical completeness method to generate mortality statistics and which make different assumptions about completeness at 5 years and above:

- 1) Conventional method: Adjusting  ${}_{45}q_{15}$  using completeness at ages 5 years and above, and inputting the adjusted  ${}_{45}q_{15}$  with  ${}_5q_0$  into an MLT.
- 2) Equivalent deaths method: Integrating the empirical completeness method with an MLT to ensure that they each produce an equal estimated total number of deaths for each sex (i.e. not restricting the  ${}_{45}q_{15}$  to be adjusted based on completeness at age 5 years and above).

The study compares mortality statistics and age-specific completeness according to these methods with C-RC studies and three alternative mortality estimates from Brazilian states, and C-RC studies from Thailand, Vietnam and Oman. C-RC methods in particular are a valuable comparator because they independently measure age patterns of mortality and completeness without reliance on MLTs. The study's findings should improve the ability of analysts, particularly at the subnational

level, to estimate timely and reliable mortality indicators in their populations using incomplete death registration data.

**Methods**

**Empirical completeness method**

The empirical completeness method estimates the completeness of death registration (or completeness of any routine death reporting system such as a Health and Demographic Surveillance System (HDSS) site) for all ages from the following models [7]:

$$\text{logit}(C_{jk1}^{All}) = \beta_0 + \text{RegCDRs}q_{jk} \times \beta_1 + \text{RegCDR}_{jk} \times \beta_2 + \%65_{jk} \times \beta_3 + \ln(5q0)_{jk} \times \beta_4 + C_{jk}^{<5} \times \beta_5 + k \times \beta_6 + \gamma_j \text{ (Model 1)}$$

$$\text{logit}(C_{jk2}^{All}) = \beta_0 + \text{RegCDRs}q_{jk} \times \beta_1 + \text{RegCDR}_{jk} \times \beta_2 + \%65_{jk} \times \beta_3 + \ln(5q0)_{jk} \times \beta_4 + k \times \beta_5 + \gamma_j \text{ (Model 2)}$$

where  $C_{jk1}^{All}$  is the completeness of registration at all ages in Model 1 and  $C_{jk2}^{All}$  is the completeness of registration at all ages in Model 2,  $\text{logit}(C_{jk1}^{All})$  is  $\ln\left(\frac{C_{jk1}^{All}}{1-C_{jk1}^{All}}\right)$  and

$\text{logit}(C_{jk2}^{All})$  is  $\ln\left(\frac{C_{jk2}^{All}}{1-C_{jk2}^{All}}\right)$ ,  $\text{RegCDR}_{jk}$  is the registered crude death rate (CDR),  $\text{RegCDRs}q_{jk}$  is the square of  $\text{RegCDR}$ ,  $\%65_{jk}$  is the fraction of the population aged 65 years and over,  $\ln(5q0)_{jk}$  is the natural log of the estimate of the true under-five mortality rate (this can be obtained nationally from the GBD or International Group for Mortality Estimation (IGME), or subnationally from Demographic and Health Surveys or censuses using summary or complete birth history data),  $C_{jk}^{<5}$  is the completeness of the registered under-five mortality (estimated as the under-five mortality rate from registration data divided by the estimate of the true under-five mortality rate),  $k$  is calendar year (which captures changes in the relationship between independent and dependent variables in the model),  $\gamma$  is a country-level random effect,  $j$  is country and  $\beta_0$  to  $\beta_6$  are the coefficients. Predicted completeness is then estimated using the inverse

$\text{logit}$ , e.g. for Model 1:  $\frac{e^{\text{logit}(C_{jk1}^{All})}}{e^{\text{logit}(C_{jk1}^{All})} + 1}$  [7, 21].

**Estimating mortality indicators using the empirical completeness method and model life table with flexible standards (MLTFS)**

**MLTFS**

This study uses the empirical completeness method with an MLT—the model life table with flexible standards (MLTFS)—to estimate complete life tables and hence mortality indicators. The MLTFS is a model life table system that generates a complete life table for a population using inputs of  $5q0$ ,  $45q15$  and a standard life table, and is the model life table system

used by the GBD [22]. The MLTFS was developed based on over 7000 empirical life tables from the Human Mortality Database and vital registration data adjusted for completeness. The equation used (where there is no HIV) is as follows:

$$\begin{aligned} \text{logit}({}_nq_x^c) &= \text{logit}({}_nq_x^s) \\ &+ \beta_x^1 \left( \text{logit}({}_5q_0^c) - \text{logit}({}_5q_0^s) \right) \\ &+ \beta_x^2 \left( \text{logit}({}_{45}q_{15}^c) - \text{logit}({}_{45}q_{15}^s) \right) \end{aligned}$$

where  $c$  denotes the country that the data are from (although it can denote any population the MLTFS is used for) and  $s$  is the standard life table [22]. Coefficients for the MLTFS are presented in Additional file 1: Table A.1. The standard life table in the GBD’s application of the MLTFS is either recent life tables of that country or, if these do not exist, a regional standard life table. Without a reliable standard such as the GBD country life table, the resultant mortality estimates might well be implausible [23, 24]. The MLTFS’s flexibility however allows for departures in  $5q0$  and  $45q15$  to estimate a complete life table which may differ significantly from the standard life table [22]. The resultant  ${}_nq_x^c$  from the MLTFS values can be converted to age-specific mortality rates ( ${}_nm_x^c$ ) using conventional life table calculations. At ages 85 years and over, the  ${}_nq_x^c$  values can be estimated using a Gompertz function [25].

For the application of the MLTFS with the empirical completeness method (detailed below), we recommend use of the national GBD life table as the standard life table, consistent with the GBD’s choice of standard life table and because it preserves the age pattern of mortality that best approximates the true pattern for the population, given the extensive mortality modelling methods used by the GBD, described elsewhere [26]. For the purposes of comparison of the mortality estimates where the GBD life table is used as the standard, another reliable standard, such as the national UN WPP life table, can also be used.

Two methods to estimate mortality indicators from incomplete death registration using the empirical completeness method and MLTFS are presented in this study: (1) the conventional method and (2) the equivalent deaths method.

**Conventional method**

The conventional method requires an estimate of completeness of registration at ages 5 years and above, which may come from DDMS, the empirical completeness method, or another method. This estimate is used this to adjust the  $45q15$  upwards, with the adjusted  $45q15$  and  $5q0$  inputted into the MLTFS to produce a life table. This is an established method

to generate age-specific mortality statistics from incomplete death registration data and is the conventional method described earlier. In our study, we use the empirical completeness method to estimate completeness at all ages. With this all-age completeness estimate, completeness at ages 5 years ( $C_{jk}^{5+}$ ) and above is calculated as:

$$C_{jk}^{5+} = \frac{RegD_{jk}^{5+}}{\left(\frac{RegD_{jk}^{All}}{C_{jk}^{All}} - \frac{RegD_{jk}^{<5}}{C_{jk}^{<5}}\right)}$$

where  $RegD_{jk}^{All}$  is registered deaths at all ages,  $RegD_{jk}^{5+}$  is registered deaths at ages 5 years and above,  $RegD_{jk}^{<5}$  is registered deaths at ages less than 5 years,  $C_{jk}^{All}$  is completeness at all ages,  $C_{jk}^{<5}$  is completeness at ages less than 5 years (used as an input into the empirical completeness method, estimated as the under-five mortality rate from registration data divided by the estimate of the true under-five mortality rate),  $j$  is country and  $k$  is calendar year. The denominator is estimated deaths at ages 5 years and above.

The adjusted  ${}_{45}q_{15}$  ( $Adj45q_{15jk}$ ) is calculated as:

$$Adj45q_{15jk} = \frac{Reg45q_{15jk}}{C_{jk}^{5+}}$$

where  $Reg45q_{15jk}$  is  ${}_{45}q_{15}$  calculated using registration data.

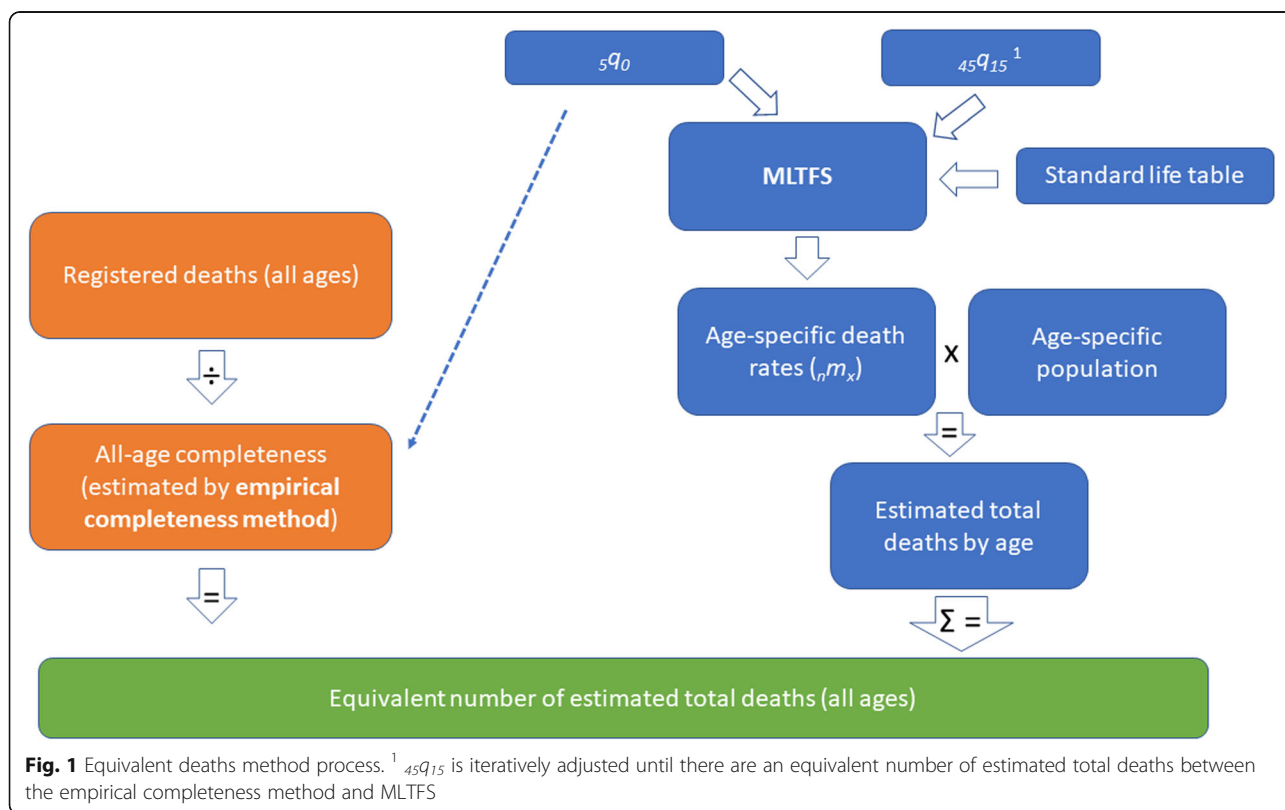
As mentioned earlier, if the level of completeness at 5 years and above is different from the true completeness for 15-59 years, then the adjusted  ${}_{45}q_{15}$  and resultant mortality statistics will be biased. Furthermore, if the MLT does not reflect the true age pattern of local mortality, this will also bias mortality statistics.

**Equivalent deaths method**

The equivalent deaths method that we propose uses the empirical completeness method and MLTFS to produce the same total number of deaths, by sex, from the following:

- Registered deaths divided by the empirical completeness method’s estimate of completeness (as a fraction).
- The MLTFS’s  ${}_n m_x$  values multiplied by the age-specific population.

If the conventional approach of adjusting  ${}_{45}q_{15}$  based on completeness at ages 5 years and above is used, it is



**Fig. 1** Equivalent deaths method process. <sup>1</sup>  ${}_{45}q_{15}$  is iteratively adjusted until there are an equivalent number of estimated total deaths between the empirical completeness method and MLTFS

highly likely the estimates of total deaths from the empirical completeness method and MLTFS would be different. The equivalent deaths method overcomes this issue—its process is shown in Fig. 1. Only limited data inputs are required:

- Estimated completeness of death registration at all ages, by sex (from the empirical completeness method)
- ${}_5q_0$ , by sex, which is the same as for the empirical completeness method
- Number of registered deaths, by age and sex (although age is not necessary for the equivalent deaths method used with model 2)
- Population by age and sex
- MLTFS model coefficients and the GBD national life table (by sex), which is used as a standard life table for national and subnational populations. As mentioned, the GBD national life table is recommended as the standard. The MLTFS coefficients are presented in Additional file 1: Table A.1.

The last remaining input into the MLTFS is the  ${}_{45}q_{15}$ , which is iteratively adjusted until the total number of deaths by sex according to the MLTFS equals the total number of deaths, by sex, according to the empirical completeness method. In combination with the  ${}_5q_0$  and other inputs, there is only one combination of  ${}_{45}q_{15}$  and estimated total deaths. The resultant outputs are  ${}_nq_x$  and  ${}_nm_x$  by sex, which can be used to produce summary mortality indicators.

A limitation of the equivalent deaths method is that both age-specific completeness and resultant mortality statistics are reliant on the accuracy of the age pattern of mortality according to the standard life table used and the resultant age pattern of mortality produced by the MLTFS (as with the conventional method). For example, a standard that, compared with the GBD life table, assumes a relatively higher level of adult compared with old-age mortality, would lead to higher estimated  ${}_{45}q_{15}$  and lower completeness of registration in ages 15–59 years when applied to the same number of registered deaths and estimated level of all-age completeness. Another limitation is that if there is population age misreporting, as is commonly found in many countries with incomplete death registration data, this will bias estimates of total deaths. Population age misreporting is most likely found at old ages, where death rates are highest. Other sources of error related to both the conventional and equivalent death methods may be from the estimate of all-age completeness (due to, for

example, uncertainty in the estimate of the true  ${}_5q_0$ ) and death registration data (reporting of age at death, date of death or place of residence of the decedent).

#### Comparison with capture-recapture studies and alternative mortality estimates from Brazilian states

We firstly compare the two approaches using results from a C-RC study in Brazil, because it provides independent and direct source information of mortality indicators and age-specific completeness that do not rely on MLTs [14]. That C-RC study matched registered deaths to deaths reported in the Ministry of Health (MoH) Mortality Information System (SIM), an independent source. In that C-RC study, total deaths, all-age completeness of registration (i.e. the civil registry) and mortality indicators were estimated by state, sex and age using generalised linear modelling with a series of covariates of demographic, geographic and socio-economic factors, as well as cause of death and place of death [14]. In this study, we compare estimated total deaths,  ${}_{45}q_{15}$  and  ${}_{25}q_{60}$  (probability of dying between age 60 and 85 years) of the conventional and equivalent death methods for all 27 states with the results in the C-RC study, as well as age-specific completeness for three states with the lowest levels of all-age completeness (Pará, Amapá and Maranhão). Our estimates are made using death registration data,  ${}_5q_0$  estimated by the GBD (which is available for Brazilian states), and the MLTFS with the GBD and, for the purposes of comparison, also the UNWPP life table for Brazil as the standard; the GBD produces life tables for Brazilian states, however, we wanted to explore the utility of the GBD and UNWPP national life tables at the subnational level for the majority of countries where the GBD does not estimate subnational life tables. The comparison of the mortality indicators for all states according to the conventional and equivalent death methods and the estimates of the C-RC study were summarised using the root mean squared difference (RMSD), relative difference (RMSD divided by the average indicator for all states), mean bias and the number of states where the indicator of each method was closer to the C-RC study than the other.

We also compare our results for 2015 to those of three alternative estimates of male  ${}_{45}q_{15}$  from Brazil reported in a recent comparative analysis of Brazilian mortality estimates, according to Queiroz et al. for 2000–2010, Schmertmann and Gonzaga for 2010, and the Instituto Brasileiro de Geografia e Estatística (IBGE—Brazil's national statistics office) for 2000–2010 [27–29]. Queiroz et al. estimated intercensal (2000–2010) completeness of death registration using three death distribution methods, and used completeness to adjust adult mortality, a method consistent with what we describe as the conventional method [28]. Schmertmann and Gonzaga's

Bayesian method combines prior estimates of completeness with a relational model life table [29, 30]. The IBGE estimates of completeness are made using the generalised growth balance method, a form of DDM [28]. Given the different time periods for which the estimates refer to, the most appropriate method of comparison is the correlation coefficient.

**Comparison with capture-recapture studies from Thailand, Vietnam and Oman**

Data are also available from C-RC studies in Thailand, Vietnam and Oman [15, 18, 20]. These studies also matched death registration to deaths reported in independent sources; a household survey (Thailand), other routine mortality reporting sources (Vietnam), and deaths reported in a census (Oman). Age-specific completeness in the Thailand, Vietnam and Oman (and sex-specific in Thailand and Vietnam) studies was calculated using the Chandrasekar-Deming method [3]. Mortality indicators were not published with these studies. For each of these studies, we compare all-age completeness and age-specific completeness with the two empirical completeness method applications. Levels of age-specific completeness can be used to generate mortality statistics. Our estimates are made using the published death registration (or reporting) data in these studies,  $5q_0$  estimated by the GBD, and the MLTFS with the GBD life table for the country as the standard.

**Results**

**Comparison with capture-recapture studies and alternative mortality estimates from Brazil**

Additional File 1: Table A.2 shows that the all-age completeness of the civil registry for Brazilian states according to the empirical completeness method and the C-RC results are very similar, being within 3 percentage points in most states and the largest difference being 7

percentage points for Acre males. In most states, the  $5q_0$  was higher according to the GBD than C-RC.

Table 1 provides summary comparison metrics for mortality indicators of the equivalent deaths and conventional methods (using GBD LT as the standard) versus the C-RC results (state-level results shown in Additional file 1: Tables A.3-A.4). Additional files 2, 3, 4 and 5 demonstrate the application of the equivalent deaths and conventional methods using both the GBD and UNWPP standard life tables. For  $45q_{15}$ , both methods provided estimates very close to C-RC, with the relative difference for the conventional method only 2% for males and 4% for females, and for the equivalent deaths method 5% for males and 7% for females, but this was inflated by the large difference in Maranhão where results were 20% different for females. The mean bias indicated no systematic bias in estimates, whilst the conventional method produced  $45q_{15}$  estimates closer to C-RC than the equivalent deaths method in 25 of the 27 states for males and 24 out of 27 for females. For  $25q_{60}$ , the relative difference for the equivalent deaths method was 3% for males and for females, and the conventional method 4% for males and for females. The equivalent deaths method provided results closer to C-RC compared with the conventional method in 21 of 27 states for males and 16 of 27 for females. Again, Maranhão was an outlier, adversely affecting the overall results for the conventional method. Of the other states with low completeness, neither the equivalent deaths method nor conventional method was clearly closer to the C-RC study for  $45q_{15}$  and  $25q_{60}$ , except that the conventional method was much closer for females  $25q_{60}$  in Amapa.

Life expectancy estimates of each method were on average within 1 year of the C-RC. The equivalent deaths method provided a smaller RMSD and relative difference than the conventional method, but the conventional method results was closer to C-RC in 15 of 27

**Table 1** Summary comparison metrics compared with C-RC study, Brazil, 2015 [14]

State	Male $45q_{15}$		Female $45q_{15}$		Male $25q_{60}$		Female $25q_{60}$	
	Equiv.	Conv.	Equiv.	Conv.	Equiv.	Conv.	Equiv.	Conv.
RMSD	0.010	0.004	0.007	0.004	0.016	0.027	0.013	0.018
Relative difference	5.2%	2.3%	7.2%	4.0%	2.5%	4.1%	2.5%	3.5%
Mean bias	+0.001	+0.001	+0.003	+0.003	+0.007	+0.007	+0.003	+0.003
States closer to C-RC	2	25	3	24	21	6	16	11
State	Male LE		Female LE		Male total deaths		Female total deaths	
	Equiv.	Conv.	Equiv.	Conv.	Equiv.	Conv.	Equiv.	Conv.
RMSD	0.51	0.61	0.41	0.61	1002	1419	834	1114
Relative difference	0.7%	0.8%	0.5%	0.8%	3.7%	8.9%	4.0%	9.9%
Mean bias	-0.18	-0.15	-0.16	-0.18	+75	-286	+122	+29
States closer to C-RC	12	15	16	11	14	13	16	11

LE life expectancy, *Equiv.* equivalent deaths method, *Conv.* conventional method

states for males and 11 states for females. The overall figures for the conventional method were affected by the large difference with the C-RC in Maranhão of 1.7 years for males and 2.8 years for females, whilst Maranhão's equivalent deaths method result was almost the same as the C-RC. The relative difference in total deaths for the equivalent deaths method (4% males and females) was less than half that of the conventional method (9% males, 10% females), but these figures were affected by the conventional method performing poorer in states with a higher number of deaths. The equivalent deaths method was closer to C-RC for slightly more states (14 out of 27 states for males, 16 states for females). Overall, the summary comparison metrics when using the UNWPP life table as the standard were very similar to when using the GBD life table (Additional file 1: Tables A.5-A.6). The most notable difference was with the life expectancy for males according to the conventional method, which exhibited a higher RMSD (0.72 UNWPP, 0.61 GBD) and larger negative mean bias (-0.59 UNWPP, -0.15 GBD) when using the UNWPP life table.

Additional file 1: Figure A.1 presents the age-specific completeness of registration according to the C-RC results, the conventional method and equivalent deaths method in three Brazilian states with the lowest levels of completeness (Pará, Amapá and Maranhão). In each state, the C-RC results show completeness was lowest at the youngest ages and also declined with older age, with these trends strongest in Maranhão, where all-age completeness is lowest. Variation in completeness by 5-year age group was greater in the conventional and equivalent deaths methods than the C-RC, exceeding 100% in some age groups. Difference in age-specific completeness between the methods was greatest in Maranhão, whilst there was very little difference between the conventional and equivalent deaths methods in Pará and Amapá. It is worth noting that for these three state differences in mortality statistics for broader age groups shown in Additional file 1: Tables A.3-A.6 are generally smaller than for completeness for 5-year age groups.

Age-specific completeness in all 11 states with all-age completeness less than 90% according to the C-RC study is shown in Additional file 1: Table A.7—as expected, age-specific completeness is more constant where all-age completeness is higher.

Comparison of the male  ${}_{45}q_{15}$  results from this study with alternative mortality estimates from Brazil show that the average correlation coefficient among our study's results (equivalent deaths method using GBD and UNWPP life table, conventional method using UNWPP life table) ranged from 0.63 to 0.66, and was more concordant than Queiroz (0.42) and IBGE (0.40) (Table 2, Additional file 1: Table A.8). Only the Bayesian estimates had a similar average correlation coefficient (0.63).

#### Comparison with capture-recapture studies from Thailand, Vietnam and Oman

In Thailand, at all ages, both methods provided completeness estimates similar to C-RC; the conventional method was closer to C-RC for males, whilst for females the two methods had a similar difference from C-RC. Completeness according to the conventional method was closer to C-RC at ages 15-59 and 60-74 years; however, the equivalent deaths method was closer at 75+ years (Table 3).

In Vietnam, the equivalent deaths method had a closer all-age completeness to C-RC (2 percentage points) than the conventional method (6 percentage points) (Table 4). Completeness according to the conventional method was however closer to C-RC at 15-59 years and, for males, 75+ years, whilst the equivalent deaths method was closer at 60-74 years for each sex. The conventional method was however substantially different to C-RC at 60-74 years, which caused the greater difference at all ages than the equivalent deaths method.

In Oman, the conventional method was closer to C-RC at all ages (2 percentage points compared with 5 percentage points for the equivalent deaths method), 15-44,

**Table 2** Correlation coefficient of male  ${}_{45}q_{15}$  according to study methods, C-RC and alternative estimates [28]

	<b>Equiv. (GBD std. LT) 2015</b>	<b>Conv. (GBD std. LT) 2015</b>	<b>Equiv. (UN std. LT) 2015</b>	<b>C-RC 2015</b>	<b>Queiroz 2000-2010</b>	<b>IBGE 2000-2010</b>	<b>Bayesian 2010</b>
Queiroz 2000-2010	0.619**	0.442*	0.629**	0.423*		0.188	0.649**
IBGE 2000-2010	0.589**	0.759**	0.585**	0.720**	0.188		0.619**
Bayesian 2010	0.689**	0.771**	0.677**	0.800**	0.649**	0.619**	
<b>Average</b>	<b>0.632</b>	<b>0.657</b>	<b>0.630</b>	<b>0.648</b>	<b>0.418</b>	<b>0.403</b>	<b>0.634</b>

*Equiv.* equivalent deaths method, *Conv.* conventional method, *Std. LT* standard life table

\* $p < 0.05$

\*\* $p < 0.01$

**Table 3** Age-specific completeness of death registration, equivalent deaths method, conventional method and C-RC, by sex, Thailand C-RC study 2005-06 and Vietnam C-RC study 2009 [18, 20]

Country, age	Males			Females		
	Equivalent deaths	Conventional	C-RC	Equivalent deaths	Conventional	C-RC
<b>Thailand</b>						
0-14 (both sexes)	85.3	84.8	59.8			
15-59	98.2	93.1	90.9	98.4	92.6	85.2
60-74	100.1	95.6	92.5	100.8	95.4	92.3
75+	84.0	81.3	94.4	87.0	83.4	93.9
All ages	94.2	90.1	91.0	93.9	89.3	91.6
<b>Vietnam</b>						
15-59	86.2	81.0	80.4	87.4	80.9	80.5
60-74	69.9	66.2	82.0	75.8	70.5	83.0
75+	88.2	84.9	81.7	83.3	78.9	81.0
All ages	79.3	75.5	81.2	79.8	75.1	81.3

Note: Results for 0-14 years not reported in Vietnam C-RC study

45-64 and 85+ years, whilst the equivalent deaths method was only closer at 65-84 years (Table 4).

## Discussion

This study has compared the application of two approaches—the conventional and equivalent deaths methods—to estimate mortality indicators from incomplete mortality data that use the empirical completeness method. The methods presented have significant appeal in enabling generation of mortality statistics from limited data, especially at the subnational level where comparator estimates from the GBD and UNWPP are mostly not available. The use of the empirical completeness method to generate mortality indicators from incomplete registration data is supported by it showing very similar levels of all-age completeness to the C-RC studies from the different countries.

The comparison was conducted with C-RC studies, the only independent estimates of age-specific

mortality and completeness in populations with incomplete mortality data which are not reliant on assumed age patterns from MLTs. Overall, the conventional and equivalent deaths methods provided estimates of mortality indicators or age-specific completeness mostly concordant with the C-RC studies, with very few exceptions. Estimates of  ${}_{45}q_{15}$  and  ${}_{25}q_{60}$  in Brazilian states were on average within 5% of the C-RC studies, except for female  ${}_{45}q_{15}$  which was significantly inflated by the outlier of Maranhão. Compared to all the C-RC studies, the conventional method overwhelmingly provided closer estimates at 15-59 years (or  ${}_{45}q_{15}$ ), the equivalent deaths method was moderately (but not overwhelmingly) closer for 60-84 years (or  ${}_{25}q_{60}$ ), whilst at all ages (or life expectancy) there was very little difference, with each method being closer for a similar number of populations. The similarity at all ages was due to the larger proportion of deaths and greater impact of life expectancy at 60-84 years than 15-59 years. The use of an alternate standard age pattern of mortality (the UN WPP national life table) made only small difference to results. Also, the male  ${}_{45}q_{15}$  findings from the equivalent deaths method (using either GBD or UN WPP standard life table) and conventional method were, on average, more concordant with alternative estimates from Brazil than those made by Queiroz et al. or the IBGE, despite the large time difference between the reference period of each study. These differences in mortality statistics are likely due to the different measurement of completeness of registration used; an advantage of the empirical completeness method is that it can estimate completeness of the

**Table 4** Age-specific completeness of death registration, equivalent deaths method, conventional method and C-RC, Oman C-RC study 2010 [15]

Age	Equivalent deaths	Conventional	C-RC
0-14	94.2	94.4	70.0
15-44	64.0	72.3	90.0
45-64	77.7	84.8	97.0
65-84	96.6	100.6	97.0
85+	92.6	91.2	91.0
All ages	86.5	91.5	89.6



most recent year of data, rather than the most recent intercensal period.

A notable characteristic of the C-RC results is that age-specific completeness, above age 5 years, is relatively constant. This may be an artefact of the all-age completeness in these C-RC study populations being over 80%, with the exception of Maranhão, and hence age-specific completeness being less likely to vary by age than if all-age completeness were lower. This characteristic explains why the conventional method provides closer estimates to C-RC studies than the equivalent deaths method at 15-59 years, because it adjusts deaths at ages 15-59 years by the level of completeness at 5 years and above. The closer estimates of the equivalent deaths method to C-RC studies at ages 60-84 years may be because both methods use the same MLT and standard life table, which at those ages may offset differences at 15-59 years for the equivalent deaths method but lead to greater differences for the conventional method. For Maranhão, where all-age completeness is below 80%, age-specific completeness declines at old ages and the national standard life table did not fit its age pattern of mortality well. However, in a setting with lower completeness of registration, poor quality age reporting for mortality and population may affect results.

In the three Brazilian states with the lowest completeness, both methods found that completeness exceeded 100% in some 5-year age groups, which shows the impact of the assumed MLT age pattern in smaller age groups. This may not necessarily be problematic for the production of mortality statistics, because age-specific completeness in excess of 100% occurred in Pará and Amapá but all their mortality statistics at broader age groups were close to that of the C-RC. However, should age-specific completeness greater than 100% occur, users should assess whether there is age misreporting in death or population data (e.g. 'heaping' of ages on digits ending in 5 or 10), whether deaths in subnational areas use consistent definitions of residence for deaths and population and whether data inputs into the empirical completeness method are appropriate; users can also potentially use an alternative age standard for the MLT.

The choice of whether to use the equivalent deaths method or conventional method may be best informed by the mortality indicators of interest. From the evidence presented of these populations with relatively high completeness, the conventional method is better at estimating  ${}_{45}q_{15}$ , the equivalent deaths method is better at estimating  ${}_{25}q_{60}$  in most populations, and there is little impact of this choice on life expectancy. In all settings, the accuracy of the MLT and standard life table used will be important in determining accuracy of mortality statistics. Another consideration is that the equivalent

deaths method provides the same number of total deaths as that estimated by the empirical completeness method, should this be deemed important; however, this may result in distorted age-specific mortality indicators if population age reporting is incorrect, a common occurrence. Unfortunately, a wider array of C-RC studies, including in lower completeness settings, is not available to further understand how age-specific completeness varies. C-RC study results are themselves subject to limitations which may bias their age-specific accuracy, especially if the assumption of the independence of sources is violated. In the Brazil study, however, the generalised linear models used included several covariates that reduced any bias from source dependency [14].

MLTs make assumptions about the age pattern of mortality that are not always accurate, even in our examples where the GBD country life table was the standard where differences in all-age completeness with the C-RC studies were small. MLTs are developed primarily from historical life tables from high-income countries, so potentially bias mortality estimates in low- and middle-income countries. MLTs other than the MLTFS could potentially be used, such as the Log-Quad and Brass logit MLTs [23, 24, 31]. However, distinct advantages of the MLTFS are its use of a more local relevant standard than the Brass logit MLT and its greater flexibility than the Log-Quad method. A final issue of our analysis is that it did not calculate uncertainty intervals around the estimates of mortality statistics, which would provide additional information with which to assess validity of individual estimates.

Improved completeness of registration and general strengthening of CRVS should increase knowledge of true age patterns of mortality, and how it varies by level of mortality and other important characteristics. Efforts to strengthen death registration systems need to focus not only on increasing the number of registered deaths but also on quality of age reporting of deaths, which is poor in some populations despite the UN Principles and Recommendations recommending the collection of date of birth and date of death, from which to calculate age at death [32]. In India, for example, 34% of registered deaths, or 2.1 million annually, do not report age in publications, thus greatly reducing their utility for guiding public policy [33]. Some countries also do not collect information on the date of birth of the deceased, relying on age reported simply in years, which causes 'age heaping', typically at ages ending in a 0 or 5, and more generally poor quality age data. An advantage of the equivalent deaths method is that it can be used, unlike the conventional death method, even without age at death data.

## Conclusions

We have assessed relatively straightforward approaches—the equivalent deaths and conventional methods—that are used with the empirical completeness method to enable analysts to generate mortality statistics from incomplete death registration data. These approaches can be used on registration data as well as other mortality data sources (e.g. Census and HDSS) and only require limited data that are readily available at the national and subnational level.

## Abbreviations

C-RC: Capture-recapture; CRVS: Civil registration and vital statistics system; DDM: Death distribution methods; GBD: Global Burden of Disease; HDSS: Health and Demographic Surveillance System; IBGE: Instituto Brasileiro de Geografia e Estatística; IGME: International Group for Mortality Estimation; MLT: Model life table; MLTFS: Model life table with flexible standards; MoH: Ministry of Health; RMSD: Root mean squared difference; SDG: Sustainable Development Goals; SIM: Mortality Information System; UNWPP: United Nations World Population Prospects

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12963-021-00262-3>.

**Additional File 1: Table A.1.** MLTFS coefficients. **Table A.2.** All-age completeness of Brazil civil registry (empirical completeness method and C-RC) and male and female  $_{5q0}$  estimates (GBD and C-RC), by sex and state, Brazil, 2015. **Table A.3.**  $_{45q15}$  and  $_{25q60}$ , equivalent deaths method (GBD std. LT), conventional method (GBD std. LT) and C-RC, by sex and state, and summary comparison metrics, Brazil, 2015. **Table A.4.** Life expectancy and total deaths, equivalent deaths method (GBD std. LT), conventional method (GBD std. LT) and C-RC, by sex and state, and summary comparison metrics, Brazil, 2015. **Table A.5.**  $_{45q15}$  and  $_{25q60}$ , equivalent deaths method (UN WPP std. LT), conventional method (UN WPP std. LT) and C-RC, by sex and state, and summary comparison metrics, Brazil, 2015. **Table A.6.** Life expectancy, equivalent deaths method (UN WPP std. LT), conventional method (UN WPP std. LT) and C-RC, by sex and state, and summary comparison metrics, Brazil, 2015. **Table A.7.** Age-specific completeness of Brazil civil registry (%), by state and sex, 2015, C-RC study. **Table A.8.** Male  $_{45q15}$  according to equivalent deaths method (GBD and UN std. LT), conventional method (GBD std. LT), C-RC, Queiroz, IBGE and Bayesian. **Figure A.1.** Age-specific completeness of Brazil civil registry (%), equivalent deaths method (GBD std. LT), conventional method (GBD std. LT) and C-RC, by sex, Pará, Amapá and Maranhão, 2015.

**Additional File 2.** Demonstration of Equivalent deaths method (GBD standard life table) in Brazilian states, 2015.

**Additional File 3.** Demonstration of Conventional deaths method (GBD standard life table) in Brazilian states, 2015.

**Additional File 4.** Demonstration of Equivalent deaths method (UN standard life table) in Brazilian states, 2015.

**Additional File 5.** Demonstration of Conventional deaths method (UN standard life table) in Brazilian states, 2015.

## Acknowledgements

Not applicable.

## Authors' contributions

TA contributed to the analytical framework, conducted the data analysis and validation and drafted earlier versions of the manuscript. ADL conceptualised the research theme, contributed to the analytical framework and edited earlier versions of the manuscript. The authors read and approved the final manuscript.

## Funding

This study was funded under an award from Bloomberg Philanthropies and the Australian Department of Foreign Affairs and Trade to the University of Melbourne to support the Data for Health Initiative. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

## Availability of data and materials

The data from the Brazil, Thailand, Oman and Vietnam studies is shown in their respective references. The demonstration of the equivalent deaths and conventional methods for Brazil is shown in Additional files 2, 3, 4 and 5.

## Declarations

### Ethics approval and consent to participate

All data used are secondary data and there was no collection of data directly from individuals, so ethics approval was not sought.

### Consent for publication

Not applicable.

### Competing interests

The authors declare they have no competing interests.

### Author details

<sup>1</sup>Melbourne School of Population and Global Health, The University of Melbourne, Level 5, Building 379, 207 Bouverie Street, Carlton, Victoria 3010, Australia. <sup>2</sup>Institute for Health Metrics and Evaluation, University of Washington, 2301 5th Ave, Seattle, WA 98121, USA.

Received: 18 June 2020 Accepted: 24 May 2021

Published online: 07 June 2021

## References

- Richards N, Sorchik R, Brolan C. Why the sustainable development goal agenda needs strong civil registration and vital statistics systems. In: CRVS Development Series. Carlton: University of Melbourne, Civil Registration and Vital Statistics Improvement, Bloomberg Philanthropies Data for Health Initiative; 2018.
- Mikkelsen L, Phillips DE, AbouZahr C, Setel PW, de Savigny D, Lozano R, et al. A global assessment of civil registration and vital statistics systems: monitoring data quality and progress. *Lancet*. 2015;386(10001):1395–406. [https://doi.org/10.1016/S0140-6736\(15\)60171-4](https://doi.org/10.1016/S0140-6736(15)60171-4).
- Chandrasekar C, Deming W. On a method of estimating birth and death rates and the extent of registration. *J Am Stat Assoc*. 1949;44:101–15.
- Dorrington R. The synthetic extinct generations method. In: Moultrie TDR, Hill A, Hill K, Timaeus I, Zaba B, editors. Tools for demographic estimation. Paris: IUSSP; 2013. p. 275–92.
- Dorrington R. The generalized growth balance method. In: Moultrie TDR, Hill A, Hill K, Timaeus I, Zaba B, editors. Tools for demographic estimation. Paris: IUSSP; 2013. p. 258–74.
- Murray CJ, Rajaratnam JK, Marcus J, Laakso T, Lopez AD. What can we conclude from death registration? Improved methods for evaluating completeness. *PLoS Med*. 2010;7:e1000262.
- Adair T, Lopez AD. Estimating the completeness of death registration: an empirical method. *PLoS One*. 2018;13(5):e0197047. <https://doi.org/10.1371/journal.pone.0197047>.
- Bennett NG, Horiuchi S. Mortality estimation from registered deaths in less developed countries. *Demography*. 1984;21(2):217–33. <https://doi.org/10.2307/2061041>.
- Dorrington R. The brass growth balance method. In: Moultrie TDR, Hill A, Hill K, Timaeus I, Zaba B, editors. Tools for demographic estimation. Paris: IUSSP; 2013. p. 196–208.
- Hill K. Estimating census and death registration completeness. *Asian Pac Popul Forum*. 1987;1:8–13 23-14.
- Moultrie T, Timaeus I. Introduction to model life tables. In: Moultrie TDR, Hill A, Hill K, Timaeus I, Zaba B, editors. Tools for demographic estimation. Paris: IUSSP; 2013. p. 341–7.
- GBD 2017 Mortality Collaborators. Global, regional, and national age-sex-specific mortality and life expectancy, 1950–2017: a

- systematic analysis for the Global Burden of Disease Study 2017. *Lancet*. 2018;392:1684–735.
13. United Nations Population Division. World population prospects: the 2017 revision. In: *World Population Prospects*. New York: United Nations; 2017.
  14. Costa LFL, Montenegro MMS, Neto DLR, de Oliveira ATR, Trindade JEO, Adair T, et al. Estimating completeness of national and subnational death reporting in Brazil: application of record linkage methods. *Popul Health Metrics*. 2020;18(1):22. <https://doi.org/10.1186/s12963-020-00223-2>.
  15. Al Muzahmi SNK. Mortality patterns in Oman: demographic and epidemiological review. PhD Thesis. Brisbane: The University of Queensland, School of Population Health; 2015.
  16. Carter KL, Baiteke T, Teea T, Tabunga T, Itienang M, Rao C, et al. Mortality and life expectancy in Kiribati based on analysis of reported deaths. *Popul Health Metrics*. 2016;14(1):3. <https://doi.org/10.1186/s12963-016-0072-6>.
  17. Carter KL, Williams G, Tallo V, Sanvictores D, Madera H, Riley I. Capture-recapture analysis of all-cause mortality data in Bohol, Philippines. *Popul Health Metrics*. 2011;9(1):9. <https://doi.org/10.1186/1478-7954-9-9>.
  18. Hoa NP, Rao C, Hoy DG, Hinh ND, Chuc NT, Ngo DA. Mortality measures from sample-based surveillance: evidence of the epidemiological transition in Viet Nam. *Bull World Health Organ*. 2012;90(10):764–72. <https://doi.org/10.2471/BLT.11.100750>.
  19. Rao C, Kelly M. Overview of the principles and international experiences in implementing record linkage mechanisms to assess completeness of death registration. vol. Technical Paper No. 2017/5. New York: Population Division, Department of Economic and Social Affairs, United Nations; 2017.
  20. Vapattanawong P, Prasartkul P. Under-registration of deaths in Thailand in 2005-2006: results of cross-matching data from two sources. *Bull World Health Organ*. 2011;89(11):806–12. <https://doi.org/10.2471/BLT.10.083931>.
  21. UN Inter-agency Group for Child Mortality Estimation (IGME). Child mortality estimates. New York: United Nations Children's Fund; 2019. <https://childmortality.org/>.
  22. Wang H, Dwyer-Lindgren L, Lofgren KT, Rajaratnam JK, Marcus JR, Levin-Rector A, et al. Age-specific and sex-specific mortality in 187 countries, 1970-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*. 2012;380(9859):2071–94. [https://doi.org/10.1016/S0140-6736\(12\)61719-X](https://doi.org/10.1016/S0140-6736(12)61719-X).
  23. Wilmoth J, Zureick S, Canudas-Romo V, Inoue M, Sawyer C. A flexible two-dimensional mortality model for use in indirect estimation. *Popul Stud*. 2012;66(1):1–28. <https://doi.org/10.1080/00324728.2011.611411>.
  24. Timaeus I. Fitting model life tables to a pair of estimates of child and adult mortality. In: Moultrie TDR, Hill A, Hill K, Timaeus I, Zaba B, editors. *Tools for demographic estimation*. Paris: IUSSP; 2013. p. 348–56.
  25. Preston S, Heuveline P, Guillot M. *Demography: measuring and modeling population processes*. Oxford: Blackwell Publishing; 2001.
  26. GBD Mortality Collaborators. Global, regional, and national under-5 mortality, adult mortality, age-specific mortality, and life expectancy, 1970-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet*. 2017;390(10100):1084–150. [https://doi.org/10.1016/S0140-6736\(17\)31833-0](https://doi.org/10.1016/S0140-6736(17)31833-0).
  27. Queiroz BFF, Gonzaga M, Lima E. Completeness of death-count coverage and adult mortality (45q15) for Brazilian states from 1980 to 2010. *Rev Bras Epidemiol (Brazilian Journal of Epidemiology)*. 2017;20 Suppl. 1:21–33.
  28. Queiroz BL, Gonzaga MR, Vasconcelos AMN, Lopes BT, Abreu DMX. Comparative analysis of completeness of death registration, adult mortality and life expectancy at birth in Brazil at the subnational level. *Popul Health Metrics*. 2020;18(S1):11. <https://doi.org/10.1186/s12963-020-00213-4>.
  29. Schmertmann CP, Gonzaga MR. Bayesian estimation of age-specific mortality and life expectancy for small areas with defective vital records. *Demography*. 2018;55(4):1363–88. <https://doi.org/10.1007/s13524-018-0695-2>.
  30. Gonzaga MR, Schmertmann CP. Estimating age- and sex-specific mortality rates for small areas with TOPALS regression: an application to Brazil in 2010. *Rev Bras Estud Popul*. 2016;33(3):629–52.
  31. Brass W. On the scale of mortality. In: Brass W, editor. *Biological aspects of demography*. London: Taylor & Francis; 1971.
  32. United Nations Statistics Division. Principles and recommendations for a vital statistics system: revision 3. In: Nations U, editor. *Statistical Papers*, Series M. New York: United Nations; 2014. <https://doi.org/10.18356/63cf3eba-en>.
  33. Office of the Registrar General India: Vital statistics of India based on the civil registration system 2015. Office of the Registrar General India ed. New Delhi: Office of the Registrar General, India; 2016.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

