

Handbook for the use of death distribution methods

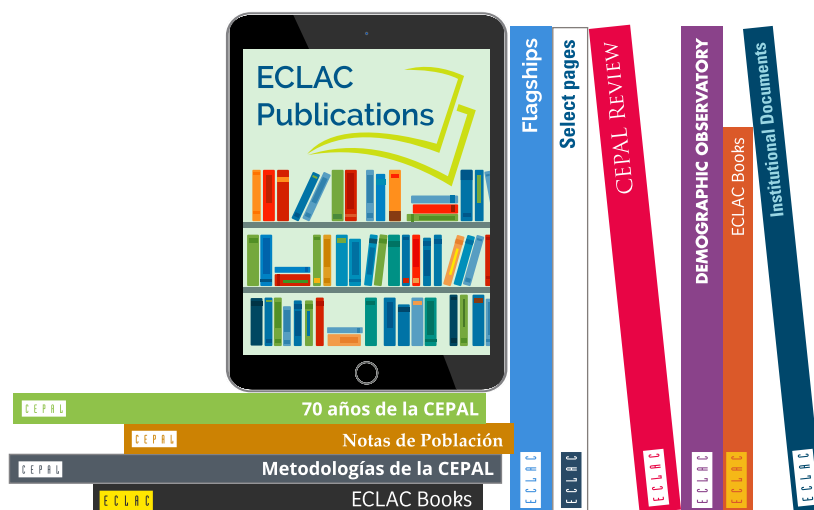
Helena Cruz Castanheira
José H. C. Monteiro da Silva



UNITED NATIONS



Thank you for your interest in this ECLAC publication



Please register if you would like to receive information on our editorial products and activities. When you register, you may specify your particular areas of interest and you will gain access to our products in other formats.

Register



www.cepal.org/en/publications



www.instagram.com/publicacionesdelacepal



www.facebook.com/publicacionesdelacepal



www.issuu.com/publicacionescepal/stacks



www.cepal.org/es/publicaciones/apps

SERIES

POPULATION AND DEVELOPMENT

137

Handbook for the use of death distribution methods

Helena Cruz Castanheira
José H. C. Monteiro da Silva



This document was prepared by Helena Cruz Castanheira and José H. C. Monteiro da Silva, of the Demography and Population Information Area of the Latin American and Caribbean Demographic Centre (CELADE)-Population Division of the Economic Commission for Latin America and the Caribbean (ECLAC), as part of the activities of the ECLAC/United Nations Population Fund (UNFPA) Regional Programme on Population and Development.

The views expressed in this document, which is a translation of an original that has been reproduced without formal editing, are those of the authors and do not necessarily reflect the views of the Organization or the countries it represents.

United Nations publication
ISSN: 1680-9009 (electronic version)
ISSN: 1680-8991 (print version)
LC/TS.2022/107/REV.1
Distribution: L
Copyright © United Nations, 2023
All rights reserved
Printed at United Nations, Santiago
S.22-00770

This publication should be cited as: H. Cruz Castanheira and J. H. C. Monteiro da Silva, "Handbook for the use of death distribution methods", *Population and Development series*, No. 137 (LC/TS.2022/107/REV.1), Santiago, Economic Commission for Latin America and the Caribbean (ECLAC), 2023.

Applications for authorization to reproduce this work in whole or in part should be sent to the Economic Commission for Latin America and the Caribbean (ECLAC), Documents and Publications Division, publicaciones.cepal@un.org. Member States and their governmental institutions may reproduce this work without prior authorization, but are requested to mention the source and to inform ECLAC of such reproduction.

Contents

Abstract.....	7
Introduction	9
I. Methodological description of DDMs.....	11
A. Growth balance methods	11
1. Brass growth balance (BGB)	11
2. General growth balance (GGB)	12
B. Synthetic extinct generations (SEG) methods	13
1. Preston-Coale method	13
2. Extinct generations (SEG) method.....	14
C. Combined GGB-SEG method (SEG-delta method).....	15
II. Data sources	17
A. Preliminary analysis: evaluating data quality	17
1. General evaluation of death series.....	17
2. Evaluation of age reporting	18
3. Counting or displacement errors in ages.....	18
4. Data evaluation for older ages.....	18
5. Evaluation of data by sex.....	19
B. Example of R application	19
1. Evaluating data quality	20
2. Data preparation for applying DDMs	25
3. Application of DDMs	26
III. Discussion and recommendations.....	33
A. Use of observed data.....	33
B. Harmonized death series.....	33
C. Information evaluation prior to method application	33
D. Choice of method.....	34
E. Caution with use in a subnational context.....	34

Bibliography	35
Annex	37
Population and Development Series: issues published.....	39

Tables

Table 1	Quality indicators for age reporting (Whipple, Myers and Coale-Li) in death data: men, Azuay, Ecuador, 2001 and 2010; and women, Arequipa, Peru, 2007 and 2017	22
Table 2	Quality indicators for age reporting (Whipple, Myers and Coale-Li) in population census data: men, Azuay, Ecuador, 2001 and 2010; and women, Arequipa, Peru, 2007 and 2017	22
Table 3	Estimates of the relative completeness of deaths, one-census BGB method: men, Azuay, Ecuador, 2001 and 2010; and women, Arequipa, Peru, 2007 and 2017	28
Table 4	Death completeness estimates, Preston-Coale one-census method: men, Azuay, Ecuador, 2001 and 2010; and women, Arequipa, Peru, 2007 and 2017	30
Table 5	Death completeness estimates, GGB, SEG and SEG-Delta two-census methods: men, Azuay, Ecuador, 2001–2010; and women, Arequipa, Peru, 2007–2017	31

Figures

Figure 1	Total deaths per year: men, Azuay, Ecuador, 2001–2010	21
Figure 2	Total deaths per year: women, Arequipa, Peru, 2007–2017	21
Figure 3	D80+ / D60+ death ratio: men, Azuay, Ecuador, and women, Arequipa, Peru, compared to men and women in Sweden	23
Figure 4	Ratio of death ages: men, Azuay, Ecuador, 2001 and 2010	23
Figure 5	Ratio of census population ages: men, Azuay, Ecuador, 2001 and 2010	24
Figure 6	Ratio of death ages: women, Arequipa, Peru, 2007–2017	24
Figure 7	Ratio of census population ages: women, Arequipa, Peru, 2007 and 2017	25
Figure 8	Diagnosis of the completeness estimation line fit using the BGB method: men, Azuay, Ecuador, 2010	27
Figure 9	Diagnosis of age ranges used for calculating completeness using the Preston-Coale method: men, Azuay, Ecuador, 2001	29
Figure 10	Diagnosis of age ranges used for calculating completeness using the Preston-Coale method: women, Arequipa, Peru, 2007	29
Figure 11	Sensitivity analysis of the selection of age ranges for DDM estimates in Azuay, Ecuador, 2001–2010	31
Figure 12	Sensitivity analysis of the selection of age ranges for DDM estimates in Arequipa, Peru, 2007–2017	32

Images

Image 1	Death data format: file <i>defunciones.csv</i>	20
Image 2	Population data format: file <i>poblacion.csv</i>	20
Image 3	Data format for applying one-census DDMs	26
Image 4	Data format for applying two-census DDMs	26
Image 5	Bgb function call for completeness estimation using one census and BGB method	27

Image 6	Bgb function call for completeness estimation using one census and BGB method for a specific age group	27
Image 7	Preston.coale function call for estimating completeness using one census and the Preston-Coale method	28
Image 8	Preston.coale function call for estimating completeness using one census and the Preston-Coale method for specific age ranges	28
Image 9	Ddm, ggb, seg and ggbseg function calls for completeness estimation using two-census methods	30

Abstract

Death distribution methods (DDMs) are demographic tools that use mathematical relationships within the population to estimate expected number of deaths by age and sex. Over time, these methods have proved to be of great use in monitoring and evaluating civil registration and vital statistics (CRVS) systems in countries that have not yet managed the complete registration of deaths, as is the case with several Latin American and Caribbean countries. Currently, some of these methods—especially the two-census models—have attained prominence and have been used in a number of national and subnational studies in the region, and packages for their use have been developed in the R programming language. However, few studies address the step-by-step application of these methods, including assessing data quality by age and estimating completeness. This document therefore aims to offer a handbook for the use of five DDMs: (1) the Brass growth balance (BGB) method, (2) the generalized growth balance (GGB) method, (3) the Preston-Coale method, (4) the synthetic extinct generations (SEG) method and (5) the combined GGB-SEG method (SEG-delta method). Use of these methods requires assessing a series of assumptions about the population's age reporting, population stability, migration and the completeness of registers by age and, for that, a series of tools and indicators for ascertaining the quality of the age data are presented. The different evaluation techniques and DDMs are then applied to two subnational examples: the Province of Azuay in Ecuador (data for men between 2001 and 2010) and the Department of Arequipa in Peru (data for women between 2007 and 2017). At the end, the sensitivity of the completeness results in relation to the different selection criteria of the age groups for estimation is discussed. It is hoped that this document and the R scripts on the website appendix will contribute to the use of these methods and techniques for assessing the quality of death and population data and to promoting the continuous monitoring of CRVS systems by the countries of Latin America and the Caribbean.

Introduction

Latin America and the Caribbean have made significant progress with their CRVS systems over the past 30 years. In addition to the initial implementation of those systems in some countries, most of the region succeeded in expanding the coverage and completeness of vital events and in implementing protocols to ensure the quality of cause-of-death data (Del Popolo and Bay, 2021). However, the CRVS systems of the region's countries remain at different stages of development. Chile, for example, has a computerized and integrated registration system for its health and civil registration statistics, with weekly data updates, whereas Haiti has published no statistics on live births and deaths for the last three decades (Del Popolo and Bay, 2021).

The coronavirus disease (COVID-19) pandemic, which struck the region during the first half of 2020, brought the debate about the quality of the number of cases and number of deaths reported into the open and showcased the importance of age- and sex-disaggregated data for the correct monitoring of epidemiological trends and disease transmission and mortality patterns (Diaz and others, 2021; Kelly, Mathenge and Rao, 2021). Across the region, that monitoring was only possible in those countries which, over recent decades, had managed to develop continuous systems. Even so, many of those countries still have problems with the coverage, completeness and quality of cause-of-death information, which poses yet another challenge for the design of epidemiological control strategies (ECLAC, 2021).

Death data collected by CRVS systems are the gold standard for estimating mortality, given that they are continuous over time, can be disaggregated to lower territorial levels, are low cost (if the institutional infrastructure is already in place in the country) and can be correlated to the population's access to rights. Because of the importance and complex nature of the information they contain, these data require the implementation of continuous monitoring and evaluation systems. Evaluation is important both to allow the estimation of demographic indicators and to identify systematic problems in data collection, with which direct interventions to improve the system can be adopted. Evaluations can be performed by analysing the completeness, coverage, accuracy, relevance, availability and timeliness of the data (United Nations, 2014 and 2021).

The technical and academic literature has developed a number of methods for evaluating the coverage and completeness of CRVS systems (Hill, 2021; Rao and others, 2020; Moultrie and others, 2013; United Nations, 2002). In general, completeness estimation methods involve the construction of expected

values for the number of deaths and births which, in turn, are compared with the observed values. In countries with more than one source, statistical matching of data entries (such as the capture-recapture method) can be used to estimate total events, as was recently implemented in Brazil (IBGE, 2018). In addition, there are demographic methods that are used, based on mathematical relations within the population (births, deaths and migrations), to calculate the expected number of events by age and sex (Hill, 2021). Among the demographic methods are death distribution methods, known by the acronym DDMs, which are used to estimate the relative completeness of death records for the adult population.

This technical note is intended to examine the DDMs available for estimating the completeness of adult mortality information and to discuss their main limitations and advantages in national and subnational applications in Latin America and the Caribbean. It presents an example of the use of the methods, based on the evaluation of data quality, using different packages developed in the R programming language: the DDM package (Riffe, Lima and Queiroz, 2017) and the DemoTools package (Riffe and others, 2019). Finally, it discusses the methods' advantages and limitations and the importance of assessing the quality of the input data prior to use. The example files and R scripts used to apply the methods are annexed to this note.

I. Methodological description of DDMs

Death distribution methods (DDMs) use age- and sex-disaggregated population and death data to estimate adult mortality. They can be divided into two groups: growth balance methods, and synthetic extinct generations methods (Hill, You and Choi, 2009).

The first group includes the **Brass growth balance (BGB) method**, which was developed by Brass (1975) and is considered the first DDM, and its extension, the **generalized growth balance (GGB) method**, proposed by Hill (1987). The next group contains the second documented DDM, developed by Preston and others (1980) and based on the extinct generations method defined by Vincent (1951), which is known as the **Preston-Coale method** and its extension, the **synthetic extinct generations (SEG) method** (Bennett and Horiuchi, 1981 and 1984). Finally, the handbook presents a third method: a hybrid version of the GGB and the SEG known as the **SEG-delta method** (Dorrington, Timaeus and Moultrie, 2008; Hill, You and Choi, 2009).

A. Growth balance methods

1. Brass growth balance (BGB)

Brass (1975) developed the first and simplest DDM, which assumes that the population under analysis is stable. In a closed population, for any open age group $a+$, the entry rate ($b(x+)$) is equal to the growth rate ($r(x+)$) plus the exit rate ($d(x+)$). In a stable population, the growth rate is constant for all age groups ($r(x+)=r$); therefore, the entry and exit rates are linearly related, as can be seen in equation 1:

$$b(x+) = r + d(x+) \quad (1)$$

Since the entry rate is calculated solely from the population's age distribution, any coverage error that does not vary with age is cancelled out; whereas the exit rate, calculated from deaths and the population by age, will be affected by any differences in coverage between population and deaths. The slope of the line relating the entry rate to the exit rate will estimate the completeness (C) of registered deaths (d^{obs}) and will provide an adjustment factor (equations 2 and 3).

$$d(x+) = \frac{d^{obs}(x+)}{C} \quad (2)$$

$$b(x+) = r + \frac{1}{C} * d^{obs}(x+) \quad (3)$$

For the adjustment of the line, at least eight points from adjacent age groups should be chosen. The selection criteria depend on the country context: for example, in cases where migration exists, the use of less affected age groups—such as those over 30 or 35 years of age—is recommended (Moultrie and others, 2013). Similarly, the use of older age groups is not recommended, due to problems with age reporting (Moultrie and others, 2013).

Inputs

- Census population by age and sex in a given period.
- Deaths by age and sex recorded in a specific period.

Outputs

- Stable population growth rate.
- Relative completeness of death registrations.

Assumptions

- Stable population.
- Population closed to migration.
- Completeness of death registrations is constant, starting at a given age (15 years, for example).

Limitations

- Stable population: problems in countries undergoing mortality transitions (Martin, 1980).
- Subjective criteria for selecting age groups leads to a great deal of variability in the results depending on the groups chosen.

2. General growth balance (GGB)

The BGB method described in the previous section can be generalized when information from two or more censuses is available, since in such cases the growth rate of each age group can be calculated from the census data, without having to comply with the stability assumption.

As noted by Hill (1987), the relationship between the exit rate (observed deaths D^{obs} over the population aged x or more $N(x+)$ calculated from the two censuses) and the entry rate (population entering at age x $N(x)$ over $N(x+)$ minus the growth rate $r(x+)$, calculated from the growth rate between the two census periods) for each age group allows the estimation of an intercept (k) that shows the variation in enumeration completeness between the two censuses and a slope ($1/C$) that estimates the completeness of the death registrations (C) relative to the average coverage of the two censuses (equation 4).

$$\frac{N(x)}{N(x+)} - r(x+) = \frac{1}{C} * \frac{D^{obs}(x)}{N(x+)} + k \quad (4)$$

As with the Brass method, for the selection of the line adjustment points, the BGB method recommends the use of at least eight points from adjacent age groups and of ages less affected by migration or age reporting problems (Hill, You and Choi, 2009; Moultrie and others, 2013).

Inputs

- Census population by age and sex over two time periods.
- Recorded deaths by age and sex over the intercensal period.

Outputs

- Relative intercensal coverage.
- Relative completeness of death registrations.

Assumptions

- Population closed to migration.
- Completeness of death registrations is constant above a certain age (15 years, for example).

Limitations

- Sensitive to age reporting problems and to intercensal coverage differentials (Dorrington, Timaeus and Moultrie, 2008; Hill, You and Choi, 2009).

B. Synthetic extinct generations (SEG) methods

1. Preston-Coale method

The basic principle of the Preston-Coale method is that, in a stable population (where mortality and birth rates are constant—and consequently so is the growth rate r —and which is closed to migration), the number of persons of age x ($N(x)$) can be estimated by examining the number of deaths starting at age x , since both population and deaths grow by the same factor r and cohort populations can be reconstructed through the deaths of their members (Preston and others, 1980) (equation 5).

$$N(x) = \sum_{a=x}^{\infty} D(a) * e^{r(a-x)} \quad (5)$$

Now, since registered deaths have a level of completeness equal to C such that $D(x)=D^{obs}(x)/C$, equation 5 becomes:

$$N(x) = \sum_{a=x}^{\infty} \frac{D^{obs}(x)}{C} * e^{r(a-x)} \quad (6)$$

As completeness C is assumed to be constant from a certain age group onwards,

$$N(x) = \frac{1}{C} \sum_{a=x}^{\infty} D^{obs}(x) * e^{r(a-x)} \quad (7)$$

Therefore, using recorded deaths, completeness can be estimated by division:

$$C = \frac{\hat{N}(x)}{N(x)} \quad (8)$$

where $\hat{N}(x)$ is the population at age x estimated from the observed deaths in the data and $\hat{N}(x)$ is the observed population at age x . From the estimates of these two elements, a series of ratios $\hat{N}(x)/N(x)$ are obtained for various age ranges. As with the BGB and GGB methods, the criteria for selecting points in the series to estimate relative completeness depend on the population context for adjusting the estimation to the method's assumptions (Moultrie and others, 2013).

One way to estimate completeness proposed by Moultrie and others (2013) is to calculate the median (C_{p50}) and the 25% (C_{p25}) and 75% (C_{p75}) percentiles within a given age range (15 to 65 years, for example) and then calculate completeness by means of equation 9.

$$C = 0.50 * C_{p50} + 0.25 * C_{p25} + 0.25 * C_{p75} \quad (9)$$

Inputs

- Census population by age and sex in a given period.
- Registered deaths by age and sex in a given period.
- Growth rate r of the stable population; for example, an estimate from two censuses or the output estimate of the Brass method can be used.

Outputs

- Completeness of death registrations.

Assumptions

- Stable population.
- Population closed to migration.
- Completeness of death registrations is constant as of a certain age (15 years, for example).

Limitations

- A stable population that is closed to migration is required.

2. Extinct generations (SEG) method

Bennett and Horiuchi (1981 and 1984) generalized the Preston-Coale method to estimate the completeness of deaths when the population is not stable. This method maintains that in a closed population, the number of people alive —of a given age at a given time— must be equal to the number of people who will die after that age, as initially proposed by Vincent (1951). The key idea behind this method is that the future flow of deaths in a cohort can be replaced by current deaths, starting at a certain age, if the effects of future population growth by age group are taken into account.

The main difference with the one-census method is that the growth rate is calculated for each age group over two time periods ($r(x)$). Thus, the increase in deaths and in the population over the intercensal period $t=t_2-t_1$ is adjusted by the factor $e^{\int_a^w r(y,t)dy}$ and no longer by the stable population constant factor e^{rt} used in Preston-Coale. In addition to this difference, in calculating the population estimated by deaths in the intercensal period (equation 10), the principle of the SEG method is the same as that of Preston-Coale, including the criteria for selecting age groups for the completeness estimation.

$$N(x) = \frac{\hat{N}(x)}{C} = \frac{1}{C} \int_x^\infty D^{obs}(a) * e^{\int_x^a r(y)dy} da \quad (10)$$

Inputs

- Census population by age and sex over two time periods.
- Recorded deaths by age and sex over the intercensal period.

Outputs

- Relative completeness of death registrations.

Assumptions

- Population closed to migration.
- Completeness of death registration is constant as of a certain age group (15 years, for example).
- The two censuses have the same population coverage.
- There are no substantive errors in the reporting of sex and age.

Limitations

- Sensitive to migration (Dorrington, Timaeus and Moultrie, 2008; Hill, You and Choi, 2009).
- Sensitive to differentials in intercensal coverage (Dorrington, Timaeus and Moultrie, 2008; Hill, You and Choi, 2009).

C. Combined GGB-SEG method (SEG-delta method)

The SEG-delta method is a two-census method and uses the same input data as the GGB and SEG methods and is based on the same assumptions. Essentially, this hybrid method applies a correction to the census enumeration completeness differential, before applying the SEG method. This correction is estimated by factor k (the ratio k^1/k^2), calculated by the line adjustment in the GGB method (Dorrington, Timaeus and Moultrie, 2008; Hill, You and Choi, 2009).

The main advantage of this hybrid model is that it allows adjustments to one of the most sensitive points of the SEG method (the coverage differential between censuses), which makes it more robust. At the same time, the elimination of the closed population assumption is still problematic, but it can be minimized by selecting an appropriate age range for the completeness estimate (Hill, You and Choi, 2009).

II. Data sources

All death distribution methods require that the data have a certain standard of quality in their age reporting. In addition, for the two-census methods (SEG, GGB and SEG-delta), the data series must be consistent and in alignment over time. Attention must therefore be paid to evaluating the quality of the age- and sex-disaggregated information on population and deaths in order to identify any possible shortcomings in the assumptions (Moultrie and others, 2013).

A. Preliminary analysis: evaluating data quality

A number of graphical data visualization tools and indexes are available to assess age misreporting in the population (Aburto and others, 2021; Agostinho, 2009; Moultrie and others, 2013). Some of these tools are described below to provide analysts with a range of options and to ensure they are aware of the main points to evaluate when using each of them.

To evaluate the misreporting of age, the examples presented in this handbook use the R package DemoTools (Riffe and others, 2019), which includes several functions for calculating most of the indicators described.

1. General evaluation of death series

The first step, particularly for death series over time, is to check death total trends by age and sex. In the absence of adverse events—such as armed conflicts, extreme weather events or geological phenomena—the expectation is that deaths will increase as the population ages. Thus, in a relatively large sample (e.g. a thousand deaths or more a year), strong variations in the data may indicate problems of quality, coverage, completeness or even the occurrence of adverse events (earthquakes, armed conflicts, epidemics or pandemics). This type of erratic behaviour in reported deaths may become more pronounced as the level of analysis narrows: from the national to the subnational level, for example, or for specific population subgroups.

2. Evaluation of age reporting

There are a number of indicators used to assess age reporting. One of the most common errors in age reporting is a preference for numbers ending in 0 and 5 (Shryock and others, 1976). This handbook will examine two of the most widely used indicators: Myers and Whipple. Others—such as the Noubissi index and the Bachi index—are available in Aburto and others (2021), but they will not be presented here as the idea is already covered by the two examples chosen. In addition, these methods correlate highly when applied to the same data and age range, making it unnecessary to use more than one or two methods to assess the quality of the databases' age reporting (see Aburto and others, 2021, for details on the application of several of these methods and indicators for assessing age reporting).

(a) Whipple's index

The purpose of Whipple's index is to measure the attraction of one or more digits among ages. Its most common use is to measure the attraction of ages with a terminal digit of 0 or 5 (Shryock and others, 1976), and it is generally calculated for adult age groups (between 20 and 70 years of age). It can be calculated in R using the DemoTools package. Its output is a value between 1 (ages are not concentrated between final digits 0 and 5) and 5 (ages are completely concentrated in final digits 0 and 5) (Aburto and others, 2021).

(b) Myers' index

Myers' index estimates the preference for each of the 10 digits in reported ages by calculating the relative proportions of each of the 10 digits in the population's declared ages. Its basic principle is that absent age misreporting, the total population with ages terminating in each digit should amount to approximately 10% of the total population (Shryock and others, 1976). The implementation of the method, using the DemoTools package, yield a result ranging from 0 (digits with perfectly uniform distribution) and 90 (all ages end in the same digit) (Aburto and others, 2021).

3. Counting or displacement errors in ages

One of the most common ways to adjust for age reporting problems—especially the preference for certain digits—is to conduct the analysis with the data aggregated into five-year groups (0 to 4, 5 to 9, 10 to 14... 80 and over), which can be done with the DDMs described in this handbook. However, problems in age reporting are often so severe that people can be displaced between these five-year groups, which can affect data in grouped ages. A tool that allows the visualization of this type of problem is the **age ratio (AR)** graph, equation 11, as proposed by Moultrie and others (2013):

$${}_5RE_x = \frac{2 * {}_5P_x}{{}_5P_{x-5} + {}_5P_{x+5}}, \text{ para } x = 5, 10, 15, \dots \quad (11)$$

Assuming that population changes are linear between age groups, the ratio between twice the population of age group x and the sum of its adjacent groups (${}_5P_{x-5} + {}_5P_{x+5}$) should be close to 1. Deviations from this value may be indicative of undercounting, data displacement errors in the age groups or phenomena such as migration and natural disasters that may affect age groups differently (Moultrie and others, 2013).

4. Data evaluation for older ages

Age misreporting or overstatement of age at death is common in older age groups and has been widely documented in the literature (Coale and Li, 1991; Moultrie and others, 2013; Preston, Elo and Stewart, 1999). The consequences of problems of this type, however, have not been closely examined in the context of applying DDMs (Hill, You and Choi, 2009). This handbook offers two different ways to evaluate such problems.

(a) Coale-Li indicator

This indicator is used to assess misreporting at older ages. In its original form, it examines the 40–90 year range for ages ending in zero. The indicator is the average given by dividing the observed frequency of people at the age containing the selected digit by the two-stage moving average using five underlying ages (Coale and Li, 1991). The two-stage moving average involves, first, estimating the moving average with five terms for all ages, after which the moving average of that first series is then estimated again. If the final index is close to one, there is no digit preference. The index produced by the DemoTools package ranges from 0 to 5, with the latter indicating that the data are completely concentrated on the digit in question (Aburto and others, 2021). In the package, the user selects the age range and the preferred digit; the defaults are an age range from 60 to the maximum age in the database and zero as the preferred digit (in general, the digits zero or five are used).

(b) Ratio

The ratio of deaths of persons aged over 80 to deaths of persons aged over 60 is used to assess the exaggeration of death ages at older ages (Agostinho, 2009). If the number of deaths of persons aged 80 and over is too high compared to the total number of deaths of persons 60 and above, this could indicate a tendency to exaggerate the age of older decedents. The value of the ratio is verified by comparing the result with a reference country that has quality death registration data; in this handbook, Sweden is used as the quality reference for this indicator.

5. Evaluation of data by sex

The evaluation of the sex ratio (or masculinity index) by age can be useful to detect possible registration differentials between sexes in civil registration and vital statistics (CRVS) systems, undercounting of more vulnerable population segments and sex preferences (Coale and Banister, 1996; Hesketh and Xing, 2006). In other cases, it also serves to identify sex selectivity trends in migratory phenomena (Leibert, 2016).

The sex ratio is the ratio of total males to total females in each age group. In general, in the region's countries, the sex ratio at birth is around 1.05.

The population's sex ratio by age depends on births, deaths by age and migration patterns within that population. Thus, it starts at a value close to 1.05 at age zero (similar to the sex ratio observed in birth records) and decreases as age advances.

For deaths, according to each country's morbidity patterns, either more or fewer male and female decedents are expected at certain ages. Up to the age of 14, mortality is generally expected to be similar between males and females. For other ages, it depends on the country's morbidity profile: for example, higher mortality among men can be expected if the country has a high homicide rate, or among women if maternal mortality is high, and so on.

B. Example of R application

This section applies the methodological approach described above for assessing data quality and estimating the completeness of death statistics in two population groups from two different Latin American countries: men in the Province of Azuay in Ecuador over the period 2001–2010 (INEC, n.d., 2010 and 2001), and women in the Department of Arequipa in Peru over the period 2007–2017 (INEI, 2007 and 2017). The R scripts for replicating this analysis are attached to this document and the description of their content is in the file *descripcion_archivos.txt*.

Images 1 and 2 show the format of the base data for the analysis. The death and population data are available in the files attached to this document (*defunciones.csv* and *poblacion.csv*) and the details of the variables are in the file *descripcion_archivos.txt*.

Image 1
Death data format: file *defunciones.csv*

	edad	pais	region	ano	sexo	defunciones
1:	0	Ecuador	Azuay	2001	f	56
2:	1	Ecuador	Azuay	2001	f	23
3:	2	Ecuador	Azuay	2001	f	13
4:	3	Ecuador	Azuay	2001	f	5
5:	4	Ecuador	Azuay	2001	f	8

7640:	86	Peru	Arequipa	2017	m	77
7641:	87	Peru	Arequipa	2017	m	81
7642:	88	Peru	Arequipa	2017	m	68
7643:	89	Peru	Arequipa	2017	m	68
7644:	90	Peru	Arequipa	2017	m	389

Source: Prepared by the authors.

Image 2
Population data format: file *poblacion.csv*

	pais	region	ano	fecha	sexo	edad	poblacion
1:	Ecuador	Azuay	2001	2001-11-25	f	0	5987
2:	Ecuador	Azuay	2001	2001-11-25	f	1	6505
3:	Ecuador	Azuay	2001	2001-11-25	f	2	6963
4:	Ecuador	Azuay	2001	2001-11-25	f	3	7116
5:	Ecuador	Azuay	2001	2001-11-25	f	4	6929

1452:	Peru	Madre de Dios	2017	2017-10-22	m	86	28
1453:	Peru	Madre de Dios	2017	2017-10-22	m	87	33
1454:	Peru	Madre de Dios	2017	2017-10-22	m	88	13
1455:	Peru	Madre de Dios	2017	2017-10-22	m	89	19
1456:	Peru	Madre de Dios	2017	2017-10-22	m	90	45

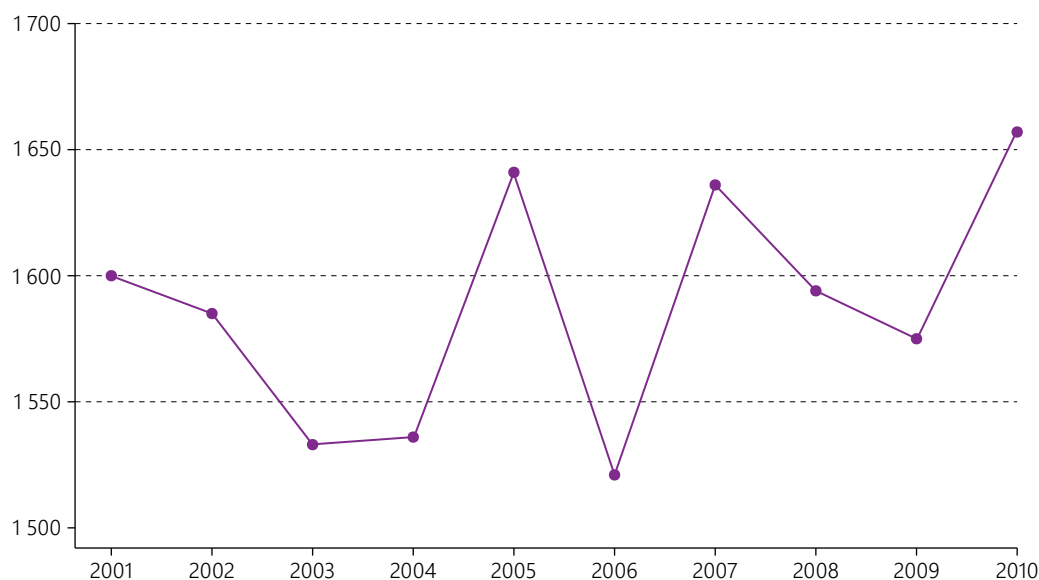
Source: Prepared by the authors.

1. Evaluating data quality

(a) Evaluating the death series

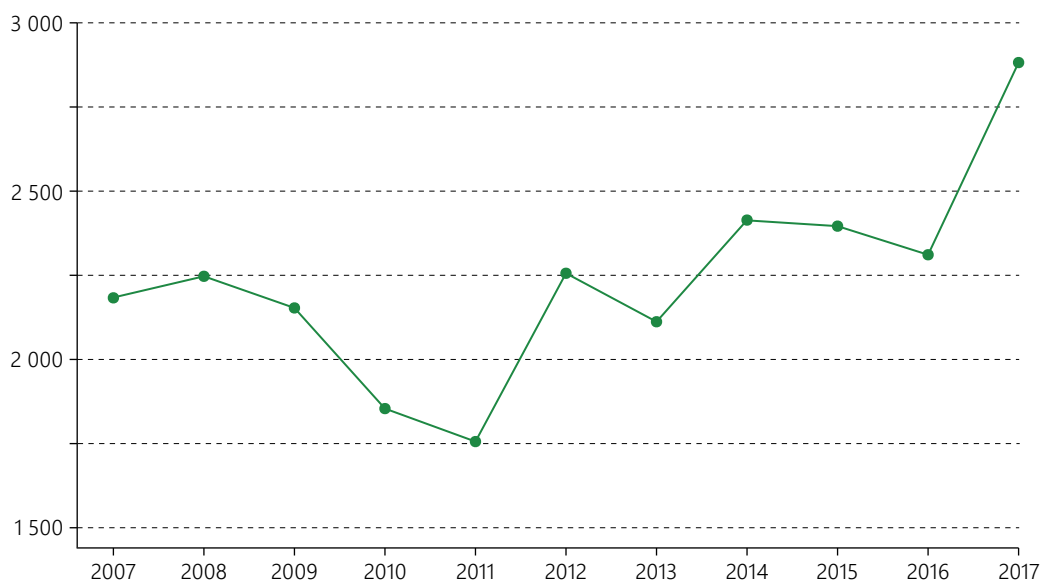
One important first step in evaluating the data is to plot the death series by year over the analysis period. Figures 1 and 2 show total deaths for each of the evaluated regions. Such visualizations are important to evaluate trends in the data, since —absent extreme events such as migratory flows, pandemics, conflicts, earthquakes and extreme climate events— it is expected that, in general, deaths will increase (or decrease) incrementally and not drastically. Strong variations at the national or subnational levels —for example, greater than 10% in samples with more than 1,000 deaths per year— may indicate registration or data coverage problems.

Figure 1
Total deaths per year: men, Azuay, Ecuador, 2001–2010



Source: National Institute of Statistics and Censuses (INEC), "Anuario de Nacimientos y Defunciones" [online] <https://www.ecuadorencifras.gob.ec/anuario-de-nacimientos-y-defunciones/>.

Figure 2
Total deaths per year: women, Arequipa, Peru, 2007–2017



Source: National Institute of Statistics and Informatics (INEI), "Bases de datos" [online] <https://www.inei.gob.pe/bases-de-datos>.

(b) Age reporting information

In order to assess digit preference in death age reporting and in the census population at adult and older ages, the Whipple, Myers and Coale-Li indexes were used. The R functions for calculating those indicators are in the DemoTools package and can be accessed using the commands *check_heaping_whipple*, *check_heaping_myers* and *check_heaping_coale_li*.

Tables 1 and 2 show the results of calculating those indexes for the selected regions of Ecuador and Peru. According to the criteria used, there are no suspicions of serious age misreporting problems in either case. In fact, the most recent census data indicate an improvement in age information in most of the indicators.

Table 1
Quality indicators for age reporting (Whipple, Myers and Coale-Li) in death data: men, Azuay, Ecuador, 2001 and 2010; and women, Arequipa, Peru, 2007 and 2017

	Whipple	Myers	Coale-Li
Azuay 2001	1.115	3.495	1.048
Azuay 2010	1.110	4.830	1.028
Arequipa 2007	1.059	4.337	1.154
Arequipa 2017	1.044	4.114	0.944

Source: National Institute of Statistics and Censuses (INEC), "Anuario de Nacimientos y Defunciones" [online] <https://www.ecuadorencifras.gob.ec/anuario-de-nacimientos-y-defunciones/>, and National Institute of Statistics and Informatics (INEI), "Bases de datos" [online] <https://www.inei.gob.pe/bases-de-datos/>.

Table 2
Quality indicators for age reporting (Whipple, Myers and Coale-Li) in population census data: men, Azuay, Ecuador, 2001 and 2010; and women, Arequipa, Peru, 2007 and 2017

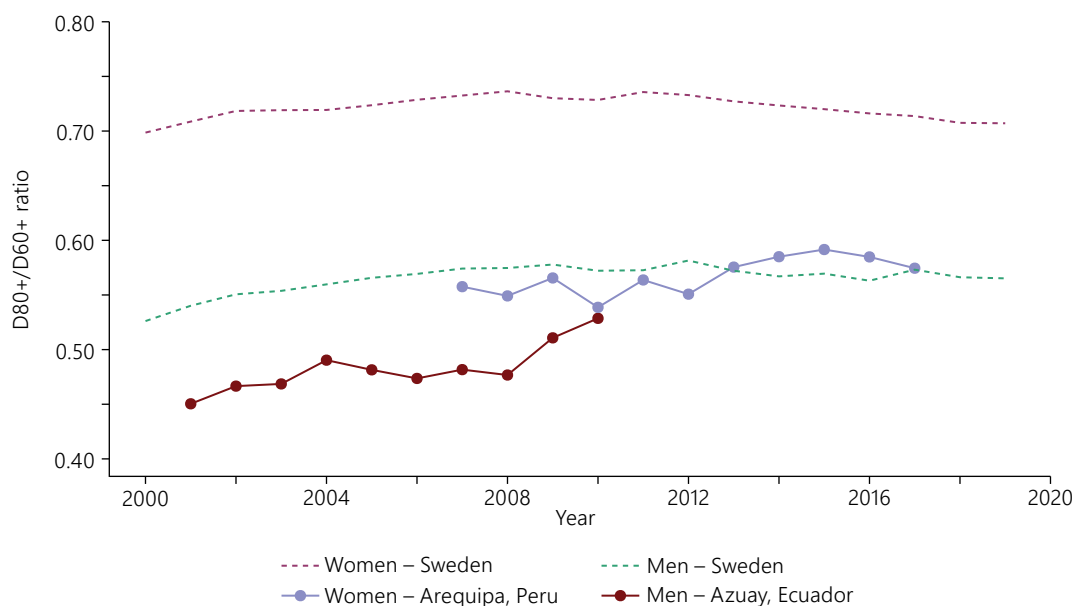
	Whipple	Myers	Coale-Li
Azuay 2001	1.156	4.027	1.411
Azuay 2010	1.025	1.102	1.167
Arequipa 2007	1.087	3.571	1.313
Arequipa 2017	1.014	0.556	0.985

Source: National Institute of Statistics and Censuses (INEC), Censo de Población y Vivienda, 2010 and 2001 [online database] <https://www.ecuadorencifras.gob.ec/base-de-datos-censo-de-poblacion-y-vivienda/>, and National Institute of Statistics and Informatics (INEI), Censos Nacionales XII de Población, VII de Vivienda, 2017 and Censos Nacionales XI de Población, VI de Vivienda, 2007 [online database] <https://www.inei.gob.pe/bases-de-datos/>.

In death data, evaluating the exaggeration of the age of the elderly is also important and this can be achieved, for example, by calculating the ratio between the total number of deaths of persons aged 80 years and over and those of persons aged 60 years and older and comparing the result with a reference country that has quality data available. Figure 3 compares the two regions under review with information from Sweden. As the calculated ratios are lower than the ratios for Sweden, there is no indication of exaggeration in the death ages of older persons. Thus, the results of the Coale-Li index for deaths and the D_{80+}/D_{60+} ratio present consistent results regarding age exaggeration at older ages.

Finally, possible problems with information displacement between age groups are evaluated by calculating populations and deaths in five-year age groups and comparing adjacent groups. The idea is that population volume changes between age groups should be smooth; in the case of deaths, however, the trend by age is only approximately linear for adult ages. Thus, the evaluation of the age ratio (AR) was restricted to ages 10 to 79 for deaths and to ages 0 to 84 for the census population. The age ratio is not expected to be far outside the range of 0.9 to 1.1; in other words, a 10% difference between age group x and the average of the adjacent age groups (Moultrie and others, 2013).

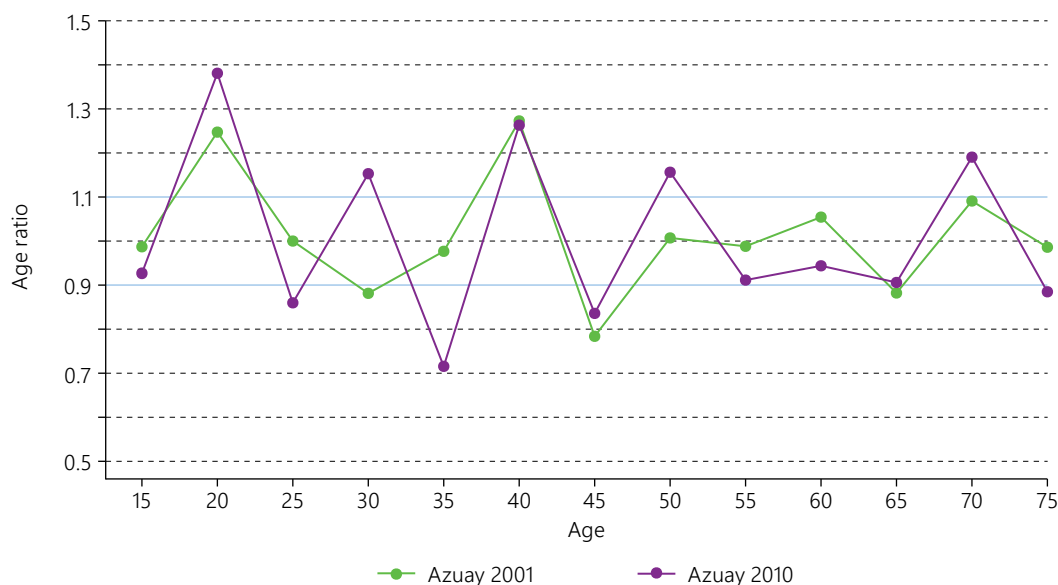
Figure 3
D80+ / D60+ death ratio: men, Azuay, Ecuador, and women, Arequipa, Peru, compared to men and women in Sweden



Source: National Institute of Statistics and Censuses (INEC), "Anuario de Nacimientos y Defunciones" [online] <https://www.ecuadrencifras.gob.ec/anuario-de-nacimientos-y-defunciones/> and National Institute of Statistics and Informatics (INEI), "Bases de datos" [online] <https://www.inei.gob.pe/bases-de-datos/>.

Figures 4 to 7 show that the death data for some age groups are slightly outside the 0.9–1.1 range. At the same time, the age ratios for the census populations are very close to 1.

Figure 4
Ratio of death ages: men, Azuay, Ecuador, 2001 and 2010



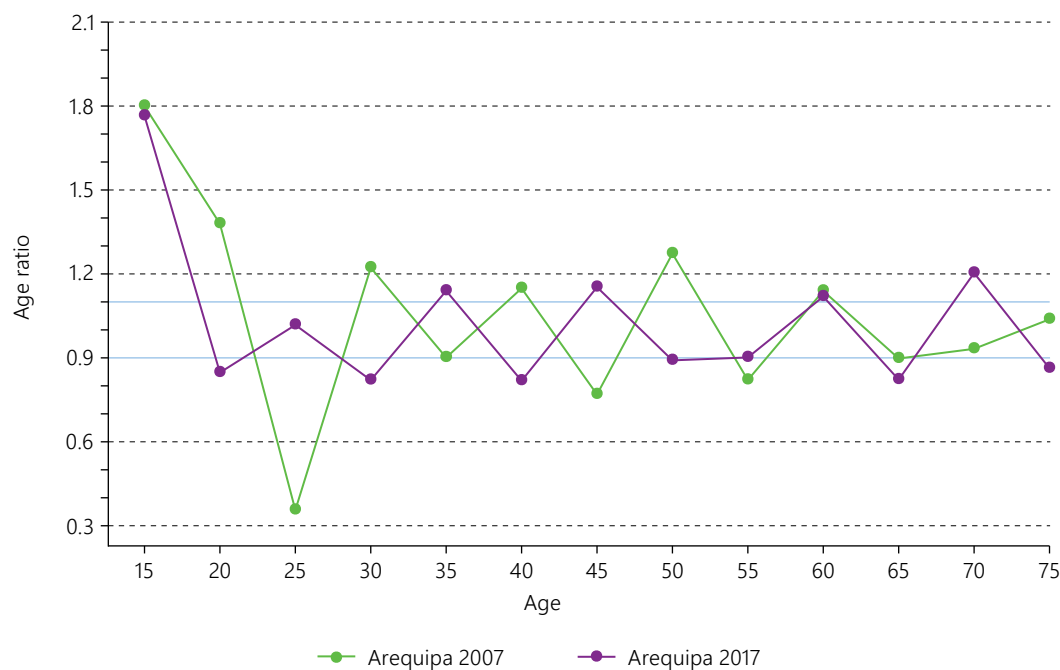
Source: National Institute of Statistics and Censuses (INEC), "Anuario de Nacimientos y Defunciones" [online] <https://www.ecuadrencifras.gob.ec/anuario-de-nacimientos-y-defunciones.15>.

Figure 5
Ratio of census population ages: men, Azuay, Ecuador, 2001 and 2010



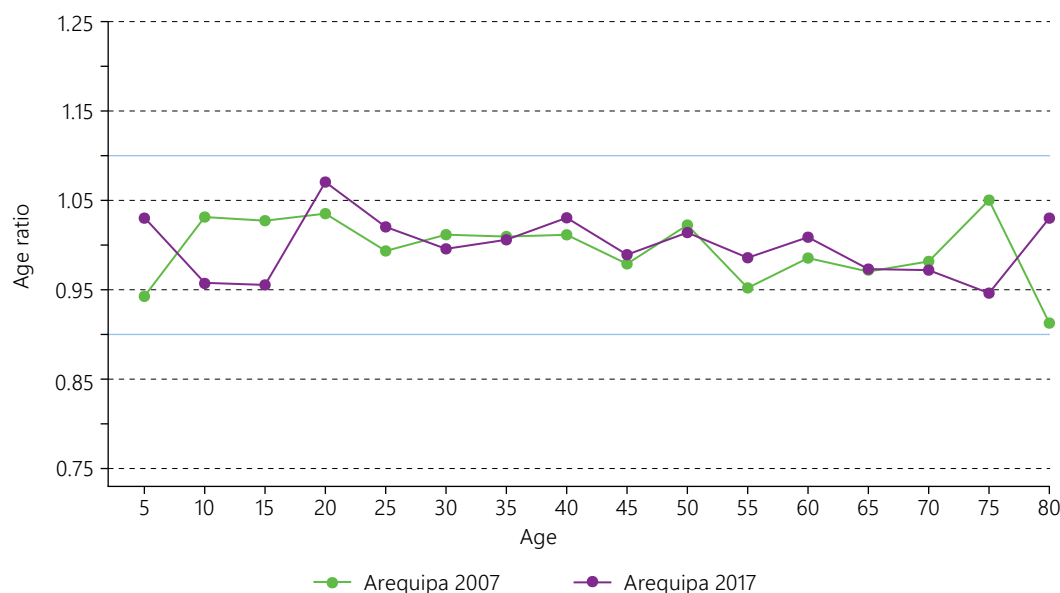
Source: National Institute of Statistics and Censuses (INEC), Censo de Población y Vivienda, 2010 and 2001 [online database] <https://www.ecuadorencifras.gob.ec/base-de-datos-censo-de-poblacion-y-vivienda/>.

Figure 6
Ratio of death ages: women, Arequipa, Peru, 2007–2017



Source: National Institute of Statistics and Informatics (INEI), "Bases de datos" [online] <https://www.inei.gob.pe/bases-de-datos/>.

Figure 7
Ratio of census population ages: women, Arequipa, Peru, 2007 and 2017



Source: National Institute of Statistics and Informatics (INEI), Censos Nacionales XII de Población, VII de Vivienda, 2017 [online database] <https://www.inei.gob.pe/bases-de-datos/> and Censos Nacionales XI de Población, VI de Vivienda, 2007 [online database] <https://www.inei.gob.pe/bases-de-datos/>.

2. Data preparation for applying DDMs

Before applying the DDM functions implemented in R, the data must be formatted correctly. Image 3 shows the format for using the one-census functions (BGB and Preston-Coale) described in this handbook (see attached files), and image 4 shows the input format for the two-census functions implemented in the DDM R package (GGB, SEG and SEG-delta).

With the two-census methods, it must be noted that the information on deaths refers either to the total deaths occurring between the census dates or to the average number of deaths occurring between those dates. For Azuay, for example, deaths occurring between 25 November 2001 and 28 November 2010 are considered. Thus, for deaths occurring in 2001 and 2010, the deaths occurring in the year are distributed proportionally for the period under study: in other words, an approximate estimate of the deaths that took place between 25 November and 31 December 2001 and those that took place between 1 January and 28 November 2010. The scripts describe these calculations in detail.

Image 3
Data format for applying one-census DDMs

```
> arequipa_2017_ddm1_input
```

	age	sex	pop	deaths	date1_deaths	date2_deaths	date_pop
1	0	f	53969	108	2017-01-01	2017-12-31	2017-10-22
2	5	f	54730	12	2017-01-01	2017-12-31	2017-10-22
3	10	f	52222	9	2017-01-01	2017-12-31	2017-10-22
4	15	f	54322	30	2017-01-01	2017-12-31	2017-10-22
5	20	f	61493	25	2017-01-01	2017-12-31	2017-10-22
6	25	f	60566	29	2017-01-01	2017-12-31	2017-10-22
7	30	f	57227	32	2017-01-01	2017-12-31	2017-10-22
8	35	f	54375	49	2017-01-01	2017-12-31	2017-10-22
9	40	f	50883	54	2017-01-01	2017-12-31	2017-10-22
10	45	f	44379	83	2017-01-01	2017-12-31	2017-10-22
11	50	f	38841	90	2017-01-01	2017-12-31	2017-10-22
12	55	f	32232	119	2017-01-01	2017-12-31	2017-10-22
13	60	f	26551	174	2017-01-01	2017-12-31	2017-10-22
14	65	f	20407	192	2017-01-01	2017-12-31	2017-10-22
15	70	f	15390	293	2017-01-01	2017-12-31	2017-10-22
16	75	f	27592	1583	2017-01-01	2017-12-31	2017-10-22

Source: Prepared by the authors.

Image 4
Data format for applying two-census DDMs

```
> azuay_ddm2_input
```

	id	sex	age	pop1	pop2	deaths	date1	date2
1:	Ecuador-Azuay-m	m	0	34057	34880	101.9	2001-11-25	2010-11-28
2:	Ecuador-Azuay-m	m	5	34953	36419	19.0	2001-11-25	2010-11-28
3:	Ecuador-Azuay-m	m	10	34940	38122	20.8	2001-11-25	2010-11-28
4:	Ecuador-Azuay-m	m	15	32462	37559	41.4	2001-11-25	2010-11-28
5:	Ecuador-Azuay-m	m	20	25800	33697	48.4	2001-11-25	2010-11-28
6:	Ecuador-Azuay-m	m	25	18723	28663	48.3	2001-11-25	2010-11-28
7:	Ecuador-Azuay-m	m	30	16295	22283	49.1	2001-11-25	2010-11-28
8:	Ecuador-Azuay-m	m	35	14379	18635	55.4	2001-11-25	2010-11-28
9:	Ecuador-Azuay-m	m	40	13012	15812	64.2	2001-11-25	2010-11-28
10:	Ecuador-Azuay-m	m	45	10787	14781	71.3	2001-11-25	2010-11-28
11:	Ecuador-Azuay-m	m	50	10117	12469	77.9	2001-11-25	2010-11-28
12:	Ecuador-Azuay-m	m	55	7804	10734	84.7	2001-11-25	2010-11-28
13:	Ecuador-Azuay-m	m	60	6939	9234	90.5	2001-11-25	2010-11-28
14:	Ecuador-Azuay-m	m	65	5743	7660	106.8	2001-11-25	2010-11-28
15:	Ecuador-Azuay-m	m	70	4668	5867	125.9	2001-11-25	2010-11-28
16:	Ecuador-Azuay-m	m	75	9113	10229	578.7	2001-11-25	2010-11-28

Source: Prepared by the authors.

3. Application of DDMs

(a) One-census methods

(i) BGB method

An example of the function call *bgb* for completeness estimation using the BGB method is presented in image 5. The standard of the function calculates estimates of growth rate *r* and completeness *c* for all possible age groups. Image 6 presents the calculation for a pre-defined selection of ages as another option.

Image 5

Bgb function call for completeness estimation using one census and BGB method

```
> bgb( azuay_2010_ddml_input, show.plot = TRUE )
```

	r	c	lower	upper	slope	intercept	RMSE
[1,]	0.029908	1.263339	15	50	0.8012766	0.02990808	0.001711
[2,]	0.028817	1.122038	15	55	0.9017813	0.02881712	0.001739
[3,]	0.027748	1.020170	15	60	0.9913952	0.02774790	0.001808
[4,]	0.026823	0.952769	15	65	1.0611277	0.02682288	0.001874
[5,]	0.026602	0.939583	15	70	1.0759232	0.02660203	0.001805
[6,]	0.029781	1.190317	20	55	0.8503879	0.02978093	0.001719
[7,]	0.028256	1.044541	20	60	0.9684648	0.02825625	0.001855
[8,]	0.027052	0.960552	20	65	1.0526286	0.02705184	0.001951
[9,]	0.026747	0.943389	20	70	1.0716454	0.02674653	0.001877
[10,]	0.027349	1.003073	25	60	1.0081293	0.02734932	0.001851
[11,]	0.026165	0.931887	25	65	1.0846152	0.02616461	0.001886
[12,]	0.026033	0.925369	25	70	1.0921952	0.02603328	0.001792
[13,]	0.024279	0.877855	30	65	1.1504866	0.02427913	0.001308
[14,]	0.024643	0.893046	30	70	1.1310846	0.02464329	0.001254
[15,]	0.023417	0.867207	35	70	1.1642029	0.02341744	0.000750

Source: Prepared by the authors.

Image 6

Bgb function call for completeness estimation using one census and BGB method for a specific age group

```
> bgb( azuay_2010_ddml_input, exact.ages = c( 20, 25, 30, 35, 40, 45, 50, 55, 60, 65 ) )
```

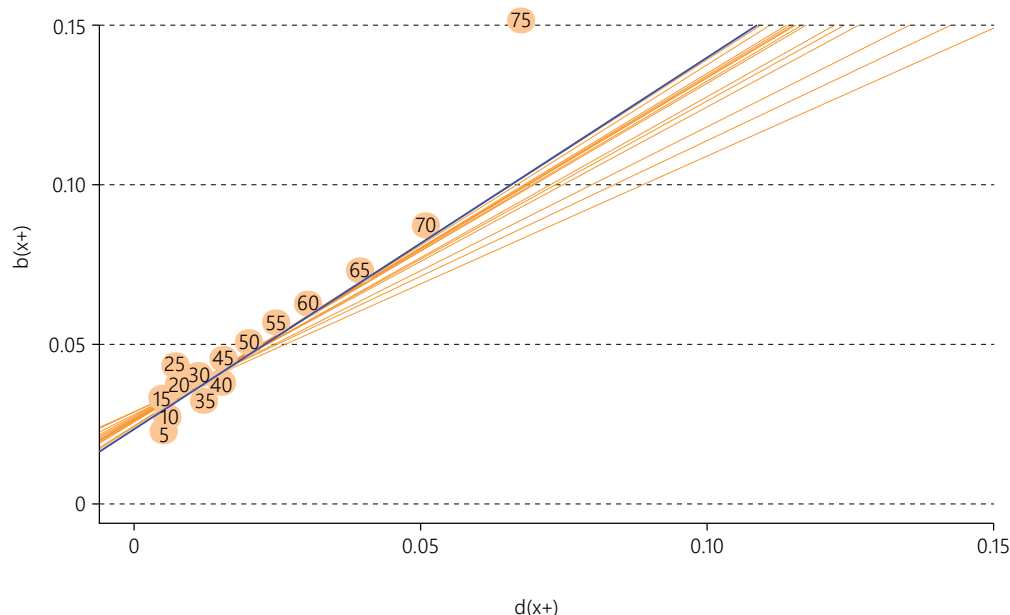
	r	c	lower	upper	slope	intercept	RMSE
[1,]	0.027052	0.960552	20	65	1.052629	0.02705184	0.001951

Source: Prepared by the authors.

The function call in image 5 also produces the diagnostic graph (figure 8) of the line fit of the equation presented above. The root mean square error (RMSE) line is shown in red on image 5, between the adjusted and observed values.

Figure 8

Diagnosis of the completeness estimation line fit using the BGB method: men, Azuay, Ecuador, 2010



Source: National Institute of Statistics and Censuses (INEC), Censo de Población y Vivienda, 2010 and 2001 [online database] <https://www.ecuadorencifras.gob.ec/base-de-datos-censo-de-poblacion-y-vivienda/> and "Anuario de Nacimientos y Defunciones" [online] <https://www.ecuadorencifras.gob.ec/anuario-de-nacimientos-y-defunciones/>.

Table 3 below presents estimates for the age ranges with the lowest error estimates in the line fit. This point will be discussed in more detail after the results of all the methods have been presented.

Table 3
Estimates of the relative completeness of deaths, one-census BGB method: men, Azuay, Ecuador, 2001 and 2010; and women, Arequipa, Peru, 2007 and 2017

	Age range	Growth rate r	Completeness–BGB
Azuay 2001	30–65	0.021	0.921
Azuay 2010	35–70	0.023	0.867
Arequipa 2007	35–70	0.035	0.901
Arequipa 2017	20–55	0.014	0.415

Source: National Institute of Statistics and Censuses (INEC), Censo de Población y Vivienda, 2010 and 2001 [online database] <https://www.ecuadorencifras.gob.ec/base-de-datos-censo-de-poblacion-y-vivienda/> and “Anuario de Nacimientos y Defunciones” [online] <https://www.ecuadorencifras.gob.ec/anuario-de-nacimientos-y-defunciones/>; National Institute of Statistics and Informatics (INEI), Censos Nacionales XII de Población, VII de Vivienda, 2017 and Censos Nacionales XI de Población, VI de Vivienda, 2007 [online database] <https://www.inei.gob.pe/bases-de-datos/>.

(ii) *Preston-Coale method*

Image 7 shows the standard call of the *preston.coale* function used to estimate the completeness of deaths using the Preston-Coale method. This standard call uses the age groups between 15 and 74 years to estimate completeness and uses the BGB function to calculate a growth rate estimate.

Image 7
Preston.coale function call for estimating completeness using one census and the Preston-Coale method

```
> preston.coale( arequipa_2017_ddm1_input, show.plot = TRUE )
      c  median    p25    p75 c15_plus lower upper
[1,] 0.421114 0.406569 0.380299 0.485063 0.419941    15    75
```

Source: Prepared by the authors.

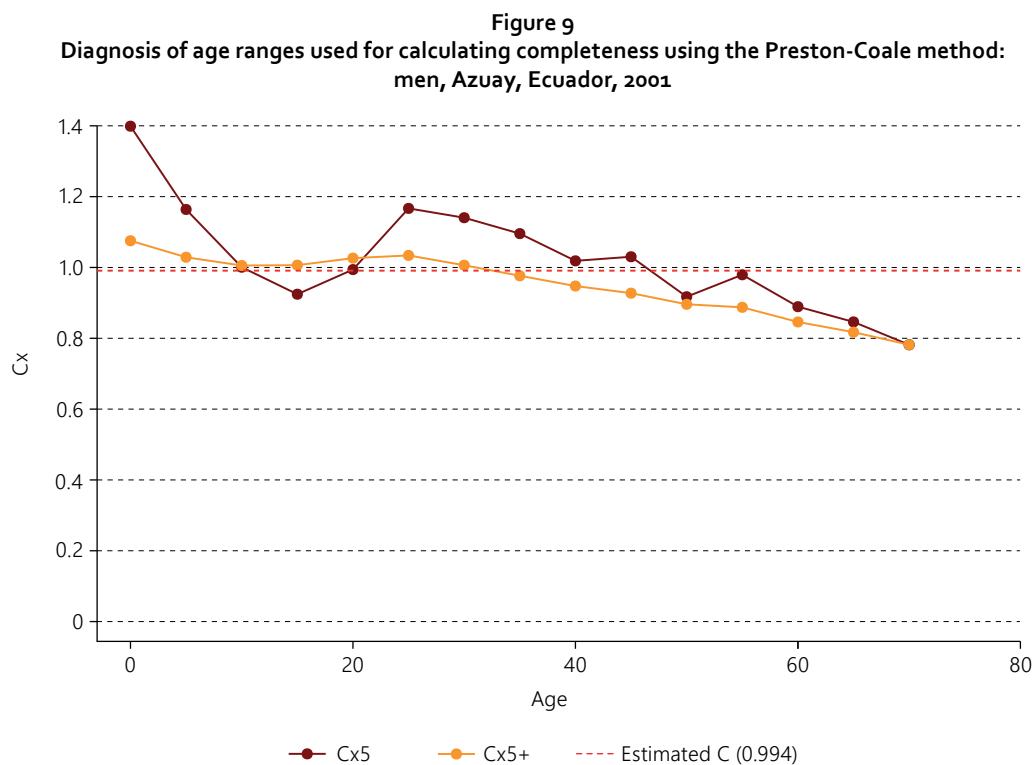
Image 8 shows an alternative mechanism with manual age group selection and the use of a user-defined growth rate r .

Image 8
Preston.coale function call for estimating completeness using one census and the Preston-Coale method for specific age ranges

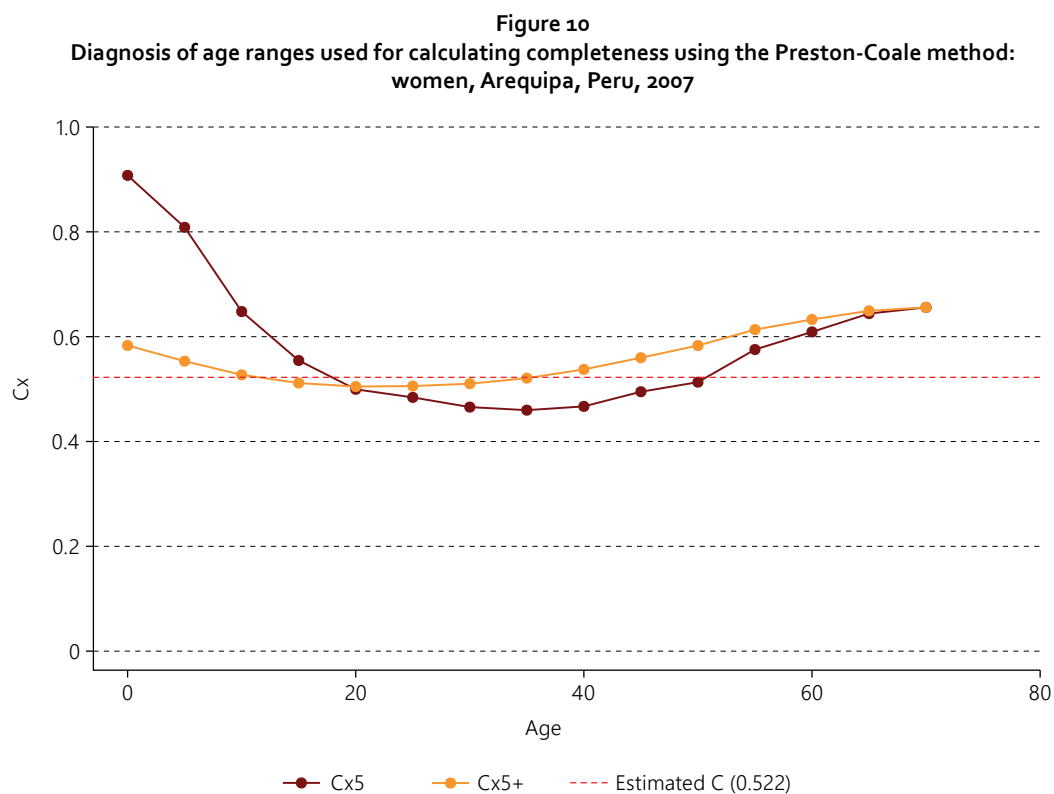
```
> preston.coale(arequipa_2017_ddm1_input, exact.ages.preston.coale = seq( 30, 70, 5 ), r = 0.025 )
      c  median    p25    p75 c15_plus lower upper
[1,] 0.426114 0.413278 0.378066 0.493806 0.419941    30    75
```

Source: Prepared by the authors.

The graphical diagnosis of the Preston-Coale method is based on plotting the C_x series versus x (age groups). Age groups with no upward or downward trends should be chosen for estimating completeness (Moultrie and others, 2013). See figures 9 and 10, below.



Source: National Institute of Statistics and Censuses (INEC), Censo de Población y Vivienda, 2010 and 2001 [online database] <https://www.ecuadorencifras.gob.ec/base-de-datos-censo-de-poblacion-y-vivienda/> and "Anuario de Nacimientos y Defunciones" [online] <https://www.ecuadorencifras.gob.ec/anuario-de-nacimientos-y-defunciones>.



Source: National Institute of Statistics and Informatics (INEI), Censos Nacionales XII de Población, VII de Vivienda, 2017, and Censos Nacionales XI de Población, VI de Vivienda, 2007 [online database] <https://www.inei.gob.pe/bases-de-datos/>.

The results of the Preston-Coale method estimates are shown on table 4. For the calculations, different growth rates were used between the initial year and the final year, close to what was observed for the period in the regions and considering that, in the first year, growth was higher (almost 2%). Age groups in which the C_x series are more constant were also chosen.

Table 4
Death completeness estimates, Preston-Coale one-census method: men, Azuay, Ecuador, 2001 and 2010;
and women, Arequipa, Peru, 2007 and 2017

	Age range	Growth rate r	Completeness–Preston-Coale
Azuay 2001	15–60	0.020	0.773
Azuay 2010	15–65	0.015	0.593
Arequipa 2007	20–55	0.020	0.475
Arequipa 2017	20–55	0.015	0.447

Source: National Institute of Statistics and Censuses (INEC), Censo de Población y Vivienda, 2010 and 2001 [online database] <https://www.ecuadorencifras.gob.ec/base-de-datos-censo-de-poblacion-y-vivienda/> and “Anuario de Nacimientos y Defunciones” [online] <https://www.ecuadorencifras.gob.ec/anuario-de-nacimientos-y-defunciones>; National Institute of Statistics and Informatics (INEI), Censos Nacionales XII de Población, VII de Vivienda, 2017 and Censos Nacionales XI de Población, VI de Vivienda, 2007 [online database] <https://www.inei.gob.pe/bases-de-datos/>.

(b) Two-census methods

The two-census methods have functions that have been implemented and tested extensively by the demography community using the DDM package in R. Image 9 shows the function calls of all those methods: (1) *ggb* function (GGB), (2) *seg* function (SEG), (3) *ggbseg* function (SEG-delta), and (4) *ddm* function (summary of all three, only with the completeness estimates of each one). The specific functions for each method contain more details of their application, such as the selected age ranges, error estimates, methods used, estimates of relative intercensal coverage (delta or k) and so on. Thus, table 5 presents the results of applying these methods with the standard age selection, selected for error minimization.

Image 9
Ddm, ggb, seg and ggbseg function calls for completeness estimation using two-census methods

```
> ddm(azuay_ddm2_input)
      id      ggb      seg      ggbseg
1 Ecuador-Azuay-m 0.6313072 0.9425945 0.6256271
> seg(azuay_ddm2_input)
      id Mxcoverage  median    l25    u25 weighted lower upper    RMSE
Ecuador-Azuay-m Ecuador-Azuay-m 0.9425945 0.9529603 0.916135 0.9691443 0.9478 15 50 0.04570368
> ggb(azuay_ddm2_input)
      id Mxcoverage lower upper    a    b    delta    k1 k2    k3
Ecuador-Azuay-m Ecuador-Azuay-m 0.6269949 35 70 -0.01125977 1.51604 0.903544 0.903544 1 0.6596133
      t1    t2    t    lm.method nx.method opt.method    r2
Ecuador-Azuay-m 2001.899 2010.907 9.008219    tukey    2    r2 0.9973524
> ggbseg(azuay_ddm2_input)
      id Mxcoverage  median    l25    u25 weighted lower.ggb upper.ggb
Ecuador-Azuay-m Ecuador-Azuay-m 0.6256271 0.6231044 0.6122837 0.6323317 0.622706 35 70
      lower.ggbseg upper.ggbseg RMSE.ggbseg
Ecuador-Azuay-m 35 70 0.01535295
```

Source: Prepared by the authors.

In general, two-census methods are more robust since they do not use the stability assumption; stability is uncommon in populations, particularly in Latin American and Caribbean countries (Moultrie and others, 2013).

Table 5
Death completeness estimates, GGB, SEG and SEG-Delta two-census methods:
men, Azuay, Ecuador, 2001–2010; and women, Arequipa, Peru, 2007–2017

	GGB ages	GGB	k (GGB)	SEG ages	SEG	SEG-delta ages	SEG-delta
Azuay 2001–2010	35–70	0.627	0.903	15–50	0.943	35–70	0.626
Arequipa 2007–2017	35–70	0.766	0.965	35–70	0.748	20–55	0.675

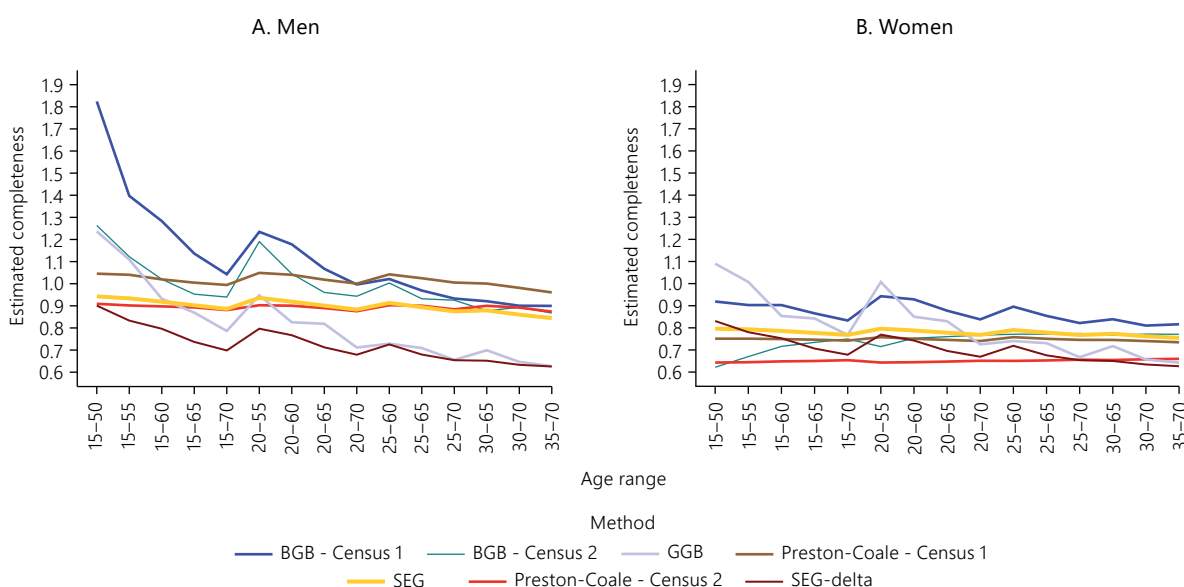
Source: National Institute of Statistics and Censuses (INEC), Censo de Población y Vivienda, 2010 and 2001 [online database] <https://www.ecuadorencifras.gob.ec/base-de-datos-censo-de-poblacion-y-vivienda/> and “Anuario de Nacimientos y Defunciones” [online] <https://www.ecuadorencifras.gob.ec/anuario-de-nacimientos-y-defunciones/>; National Institute of Statistics and Informatics (INEI), Censos Nacionales XII de Población, VII de Vivienda, 2017 and Censos Nacionales XI de Población, VI de Vivienda, 2007 [online database] <https://www.inei.gob.pe/bases-de-datos/>.

(i) Sensitivity analysis

Both one- and two-census methods are sensitive to the age groups selected for the completeness estimation, especially those methods that use the population growth rate equation (BGB and GGB). Thus, one way to plot the sensitivity of the methods is by calculating the completeness for multiples of age selection criteria and then observing the variability of the results. As shown on figures 11 and 12, the GGB method is very sensitive to the ages selected for estimation. In contrast, the SEG method is less sensitive to group selection but is more affected by the relative intercensal coverage differential. Consequently, the SEG-delta method reports greater variability in relation to the selected age groups.

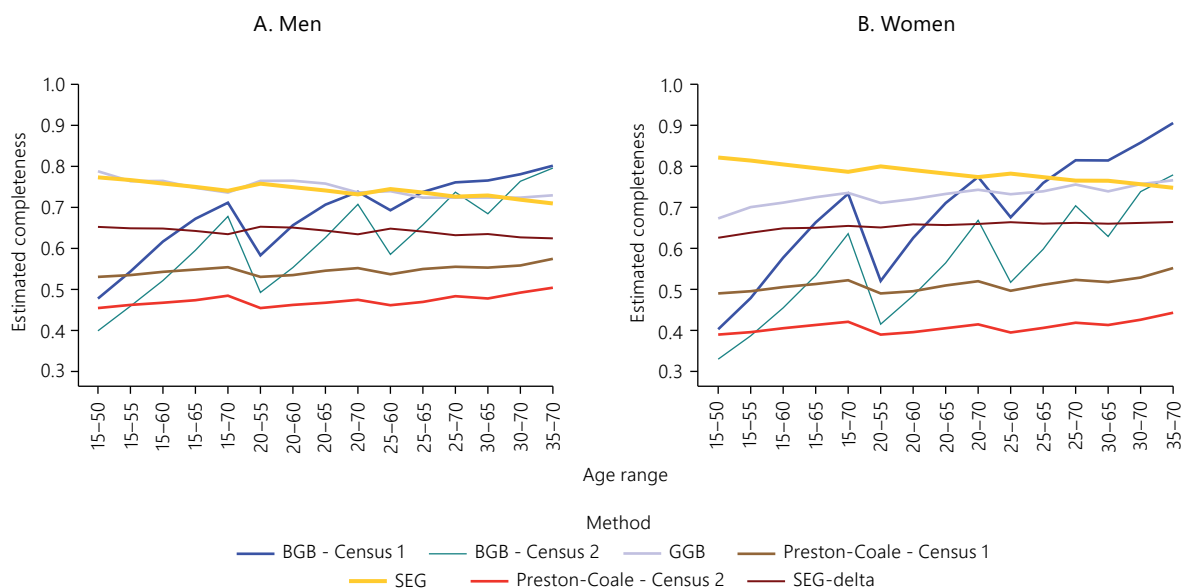
The one-census methods (BGB and Preston-Coale) yield the relative completeness of deaths during the reference period (census year). In turn, the two-census methods yield the average relative completeness of deaths over the intercensal period. Thus, the series in figures 11 and 12 with one-census and two-census methods use different reference periods and are not directly comparable. In addition, the Preston-Coale method uses a population growth rate as an input and, in this case, the growth rate resulting from the BGB method using the 15 to 59 age groups was used. Therefore, further analysis could be done by testing the sensitivity of this method for the choice of the age group for the BGB growth rate.

Figure 11
Sensitivity analysis of the selection of age ranges for DDM estimates in Azuay, Ecuador, 2001–2010



Source: National Institute of Statistics and Censuses (INEC), Censo de Población y Vivienda, 2010 and 2001 [online database] <https://www.ecuadorencifras.gob.ec/base-de-datos-censo-de-poblacion-y-vivienda/> and “Anuario de Nacimientos y Defunciones” [online] <https://www.ecuadorencifras.gob.ec/anuario-de-nacimientos-y-defunciones/>; National Institute of Statistics and Informatics (INEI), Censos Nacionales XII de Población, VII de Vivienda, 2017 and Censos Nacionales XI de Población, VI de Vivienda, 2007 [online database] <https://www.inei.gob.pe/bases-de-datos/>.

Figure 12
Sensitivity analysis of the selection of age ranges for DDM estimates in Arequipa, Peru, 2007–2017



Source: National Institute of Statistics and Censuses (INEC), Censo de Población y Vivienda, 2010 and 2001 [online database] <https://www.ecuadorencifras.gob.ec/base-de-datos-censo-de-poblacion-y-vivienda/> and “Anuario de Nacimientos y Defunciones” [online] <https://www.ecuadorencifras.gob.ec/anuario-de-nacimientos-y-defunciones/>; National Institute of Statistics and Informatics (INEI), Censos Nacionales XII de Población, VII de Vivienda, 2017 and Censos Nacionales XI de Población, VI de Vivienda, 2007 [online database] <https://www.inei.gob.pe/bases-de-datos/>.

III. Discussion and recommendations

This handbook has set out a method for using DDMs, including examples of applying the methods to a preliminary analysis of the quality of population and death data for regions of Ecuador and Peru. The following sections contain some important points and discussions to be considered in using DDMs.

A. Use of observed data

It is important to keep in mind that the use of observed data in DDMs—for both population and deaths—avoids circularity in the estimates. For example, if estimated population data are used, they already contain adjustments that could influence the estimates for completeness and intercensal coverage.

B. Harmonized death series

Death series should come from the same data source, as different sources may have different levels of completeness or coverage, leading to biased completeness estimates.

C. Information evaluation prior to method application

An analysis of the observed data is vital before the methods are applied, as they are very sensitive to breakdowns in age and migration reporting assumptions and this can be seen in the variability observed in the results (Agostinho, 2009; Dorrington, Timaeus and Moultrie, 2008; Hill, You and Choi, 2009; Moultrie and others, 2013). If necessary, the data can be adjusted using the Beers, Sprague, Spline and Kannisto methods, moving averages or other mechanisms (Riffe and others, 2019) for correcting age reporting problems in the data.

D. Choice of method

Each method has its advantages and disadvantages. The first one-census methods, BGB and Preston-Coale, have the disadvantage of assuming a stable population, which limits their usefulness in countries where the demographic transition occurred recently or remains ongoing (Hill, 2021). Among the two-census methods, GGB is highly sensitive to the assumption of a population closed to migration, while the SEG method is more sensitive to the assumption of age reporting and the intercensal coverage differentials (Dorrington, Timaeus and Moultrie, 2008; Hill, You and Choi, 2009). In general, the GGB-SEG method is a good adjustment of the SEG method, as it corrects the intercensal coverage problem to yield a method that is less sensitive to migration (Dorrington, Timaeus and Moultrie, 2008; Hill, You and Choi, 2009; Murray and others, 2010). Numerous recent studies have proposed this method as a good way to estimate adult-age completeness (Murray and others, 2010). It should nevertheless be noted that all the methods present problems and yield biased estimates when age-completeness differentials exist.

E. Caution with use in a subnational context

Using DDMs in a subnational context requires care and attention from the analyst, as the effects related to migration and data quality are magnified when dealing with smaller areas. In particular, migration should be monitored and diagnostic charts should be evaluated for choosing the most appropriate age group for the completeness estimation. In this context, the results presented in this handbook's examples must be analysed with caution.

Bibliography

- Aburto, J. M. and others (2021), "Age-heaping indices with DemoTools" [online] https://timriffe.github.io/DemoTools/articles/Age-heaping_quality_with_Demotools.html.
- Agostinho, C. S. (2009), "Estudo sobre a mortalidade adulta, para Brasil entre 1980 e 2000 e unidades da Federação em 2000: uma aplicação dos métodos de distribuição de mortes", PhD thesis, Federal University of Minas Gerais.
- Bennett, N. G. and S. Horiuchi (1984), "Mortality estimation from registered deaths in less developed countries", *Demography*, vol. 21, No. 2.
- ____ (1981), "Estimating the completeness of death registration in a closed population", *Population Index*, vol. 47, No. 2.
- Brass, W. (1975), *Methods for Estimating Fertility and Mortality from Limited and Defective Data*, Laboratories for Population Statistics.
- Coale, A. J. and J. Banister (1996), "Five decades of missing females in China", *Proceedings of the American Philosophical Society*, vol. 140, No. 4.
- Coale, A. J. and S. Li (1991), "The effect of age misreporting in China on the calculation of mortality rates at very high ages", *Demography*, vol. 28, No. 2.
- Del Popolo, F. and G. Bay (coords.) (2021), "Las estadísticas de nacimientos y defunciones en América Latina con miras al seguimiento de la Agenda 2030 para el Desarrollo Sostenible y del Consenso de Montevideo sobre Población y Desarrollo", *Population and Development series*, No. 134 (LC/TS.2021/48), Santiago, Economic Commission for Latin America and the Caribbean (ECLAC).
- Diaz, T. and others (2021), "A call for standardised age-disaggregated health data", *Lancet Healthy Longevity*, vol. 2, No. 7.
- Dorrington, R., I. Timaeus and T. Moultrie (2008), "Death distribution methods for estimating adult mortality: Sensitivity analysis with simulated data errors, revisited", document presented at the Population Association of America 2008 Annual Meeting, New Orleans, 17–19 April.
- ECLAC (Economic Commission for Latin America and the Caribbean) (2021), *Demographic Observatory, 2020* (LC/PUB.2020/20-P), Santiago.
- Hesketh, T. and Z. W. Xing (2006), "Abnormal sex ratios in human populations: causes and consequences", *Proceedings of the National Academy of Sciences*, vol. 103, No. 36.
- Hill, K. (2021), "Métodos analíticos para evaluar la completitud y la calidad del registro de las defunciones: estado actual de los conocimientos", *Population and Development series*, No. 133 (LC/TS.2021/11), Santiago, Economic Commission for Latin America and the Caribbean (ECLAC).

- _____. (1987), "Estimating census and death registration completeness", *Asian and Pacific Population Forum*, vol. 1, No. 3.
- Hill, K., D. You and Y. Choi (2009), "Death distribution methods for estimating adult mortality: sensitivity analysis with simulated data errors", *Demographic Research*, vol. 21, No. 9.
- IBGE (Brazilian Institute of Geography and Statistics) (2018), *Sistemas de Estatísticas Vitais no Brasil*, Rio de Janeiro.
- INEC (National Institute of Statistics and Censuses) (n.d.), Registro Estadístico de Defunciones Generales [online] <https://www.ecuadorencifras.gob.ec/defunciones-generales/>.
- _____. (2010), Censo de Población y Vivienda [online database] <https://www.ecuadorencifras.gob.ec/base-de-datos-censo-de-poblacion-y-vivienda/>.
- _____. (2001), Censo de Población y Vivienda [online database] <https://www.ecuadorencifras.gob.ec/base-de-datos-censo-de-poblacion-y-vivienda/>.
- INEI (National Institute of Statistics and Informatics) (2017), Censos Nacionales XII de Población, VII de Vivienda [online database] <https://www.inei.gob.pe/bases-de-datos/>.
- _____. (2007), Censos Nacionales XI de Población, VI de Vivienda [online database] <https://www.inei.gob.pe/bases-de-datos/>.
- Kelly, M., G. Mathenge and C. Rao (2021), "Lessons learnt and pathways forward for national civil registration and vital statistics systems after the COVID-19 pandemic", *Journal of Epidemiology and Global Health*, vol. 11, No. 3.
- Leibert, T. (2016), "She leaves, he stays? Sex-selective migration in rural East Germany", *Journal of Rural Studies*, vol. 43.
- Martin, L. G. (1980), "A modification for use in destabilized populations of Brass's technique for estimating completeness of death registration", *Population Studies*, vol. 34, No. 2.
- Moultrie, T. and others (2013), *Tools for Demographic Estimation*, Paris, International Union for the Scientific Study of Population (IUSSP).
- Murray, C. J. and others (2010), "What can we conclude from death registration? Improved methods for evaluating completeness", *PLoS Medicine*, vol. 7, No. 4.
- Preston, S. and others (1980), "Estimating the completeness of reporting of adult deaths in populations that are approximately stable", *Population Index*, vol. 46, No. 2.
- Preston, S. H., I. T. Elo and Q. Stewart (1999), "Effects of age misreporting on mortality estimates at older ages", *Population Studies*, vol. 53, No. 2.
- Rao, C. and others (2020), *Estimating Completeness of Birth and Death Registration: Methods and Options for Estimating Completeness of Civil Registration*, New York, Vital Strategies.
- Riffe, T. and others (2019), *DemoTools: Standardize, Evaluate, and Adjust Demographic Data*.
- Riffe, T., E. Lima and B. Queiroz (2017), *DDM: Death Registration Coverage Estimation*.
- Shryock, H. S. and others (1976), *The Methods and Materials of Demography*, Academic Press.
- United Nations (2021), "Handbook on civil registration and vital statistics systems: management, operation and maintenance. Revision 1", *Studies in Methods*, series F, No. 72, Rev. 1 (ST/ESA/STAT/SER.F/72/Rev.1), New York.
- _____. (2014), "Principles and recommendations for a vital statistics system. Revision 3", *Statistical Papers*, series M, No. 19, Rev. 3 (ST/ESA/STAT/SER.M/19/Rev.3), Nueva York.
- _____. (2002), *Methods for Estimating Adult Mortality* (ESA/P/WP.175).
- Vincent, P. (1951), "La mortalité des vieillards", *Population*, vol. 6, No. 2.

Annex

Attached to this document is a zip file, available at https://celade.cepal.org/documentos/MetodosDefunciones/anexo_metodos_defunciones.zip. The archive contains three folders with eleven files for applying the methods described in the handbook. The zip file is structured as follows:

- A. InputsDDM
 - 1. inputs_ddm_1_censo.RData
 - 2. inputs_ddm_2_censos.RData
- B. InputsManual
 - 1. defunciones.csv
 - 2. defunciones_suecia.csv
 - 3. poblacion.csv
- C. RScriptsManual
 - 1. 1_analisis_preliminar.R
 - 2. 2_preparacion_datos_ddm.R
 - 3. 3a_aplicacion_ddm_1_censo.R
 - 4. 3b_aplicacion_ddm_2_censos.R
 - 5. 4_analisis_sensibilidad_ddm.R
 - 6. funciones_ddm_1_censo.R

**Series****ECLAC****Population and Development**

Issues published

A complete list as well as pdf files are available at
www.eclac.org/publicaciones

137. Handbook for the use of death distribution methods, Helena Cruz Castanheira and José H. C. Monteiro da Silva (LC/TS.2022/107/REV.1), 2023.
136. Nuevas herramientas y análisis sobre segregación residencial socioeconómica, migración interna y movilidad cotidiana en Costa Rica, Guatemala y el Perú, 1980-2018, Jorge Rodríguez, Alexandra Martínez (LC/TS.2021/152), 2021.
135. Demographic transition and achieving the SDGs in Latin America and the Caribbean: a regional overview of the National Transfer Accounts, Marta Duda-Nyczak (LC/TS.2021/146), 2021.
134. Las estadísticas de nacimientos y defunciones en América Latina con miras al seguimiento de la Agenda 2030 para el Desarrollo Sostenible y del Consenso de Montevideo sobre Población y Desarrollo, Fabiana Del Popolo, Guiomar Bay (coords,) (LC/TS.2021/48), 2021.
133. Métodos analíticos para evaluar la completitud y la calidad del registro de las defunciones: estado actual de los conocimientos, Kenneth Hill (LC/TS.2021/11), 2021.
132. Perspectivas de la población mundial 2019: metodología de las Naciones Unidas para las estimaciones y proyecciones de población (LC/TS.2020/95), 2020.
131. Maternidad, fecundidad y paridez en la adolescencia y la juventud: continuidad y cambio en América Latina, Jorge Rodríguez Vignoli y Victoria San Juan Bernuy (LC/TS.2020/89), 2020.
130. Cuentas Nacionales de Transferencias en el Paraguay en 2012: instrumento para las políticas públicas, Verónica Serafini Geoghegan (LC/TS.2020/57), 2020.
129. Las Cuentas Nacionales de Transferencias del Perú y los desafíos para lograr los Objetivos de Desarrollo Sostenible, J. Olivera y Y. Iparraguirre (LC/TS.2019/106), 2019.
128. Transferencias entre generaciones y grupos socioeconómicos: estimaciones de las Cuentas Nacionales de Transferencias en la Argentina, Pablo Comelatto (LC/TS.2019/105), 2019.

POPULATION AND DEVELOPMENT

Issues published:

- 137 Handbook for the use of death
distribution methods
*Helena Cruz Castanheira and
José H. C. Monteiro da Silva*
- 136 Nuevas herramientas y análisis
sobre segregación residencial
socioeconómica, migración interna
y movilidad cotidiana en Costa Rica,
Guatemala y el Perú, 1980-2018
Jorge Rodríguez Vignoli y Alexandra Martínez
- 135 Demographic transition and
achieving the SDGs in Latin America
and the Caribbean
A regional overview of the National
Transfer Accounts
Marta Duda-Nyczak



Economic Commission for Latin America and the Caribbean (ECLAC)
Comisión Económica para América Latina y el Caribe (CEPAL)
www.eclac.org



LC/TS.2022/107/REV.1