The Latin American Human Mortality Database: an emerging source of mortality data for developing countries

B. Piedad Urdinola
Department of Statistics – Universidade Nacional de Colombia
bpurdinolac@unal.edu.co

Bernardo L. Queiroz
Department of Demography – UFMG - Brazil
lanza@cedeplar.ufmg.br

Abstract

The Latin America Human Mortality Database (LAHMD) is a joint project of Bernardo L. Queiroz (UFMG) and B. Piedad Urdinola (UNAL). It is inspired in the Human Mortality Database (www.mortality.org) and provide a variety of mortality data collected from various external sources for Latin America countries. In contrast to the HMD, the LAHMD comprises data considering additional dimensions not covered in the HMD (regions in the country and causes of death). The relevant documentation is also offered to the users. The HFC aims at including a wide range of countries in Latin America and covering the longest possible time periods. The data in the LAHMD are provided in the standardized format and can be easily and openly accessed. The LAHMD website is found at www.lamortalidad.org. The database contains mortality data (raw data and not adjusted for under-registration of death counts) for a number of national and subnational populations, it also includes data on causes of death. This paper describes the LAHMD, computational procedures used for data harmonization as well as its structure.

Extended Abstract Submitted to the 2014 PAA meeting

1. Introduction

The web-site aims at disseminating human mortality data and literature of human mortality in Latin America, in order to provide detailed information for researchers, students, policy makers and the general public interested in knowing trends and developments in the study of mortality in the region. The project is inspired by The Human Mortality Database (**www.mortality.org**) and is the result of joint work of Prof. B. Urdinola of the National University of Colombia, Bogota Department of Statistics and Professor Bernardo Lanza Queiroz CEDEPLAR, Belo Horizonte-Brazil, funded by the Population Association of America (**www.popassoc.org**) and the Research Direction of Universidad Nacional de Colombia-Bogotá (DIB **www.dib.unal.edu.co**), the Brazilian Science Foundation (www.cnpq.br) and the Minas Gerais Science Foundation (www.fapemig.gov.br) . At present the database contains detailed information on mortality for five countries in Latin America: Argentina, Brazil, Colombia, Mexico and Peru. All information is broken down by age, sex, region and cause of death.

2. Objectives

This paper has two main objectives. The first one is to present a detailed description of the Latin America Human Mortality Database, a description of the procedures to treat the available and provide a brief description of the potential uses of the LAHMD.

The second aim of the paper is to present a detailed evaluation of mortality data, at country level, for the five countries currently available in the database. First, we present a series of data quality checks in relation to the accuracy of reported age of death. Second, we use Death Distribution Methods (DDM) to evaluate quality of mortality data and estimate the degree of completeness of death counts registration for each country over time.

3 Data and Methods

3.1 Sources and Data Treatment (General)

The Human Mortality Database for Latin America has mortality series for total national territories by age, sex and cause of death, as well as the same categories per region within each country since 1980 to the latest available year. The countries included are: Argentina, Brazil, Colombia, Mexico and Peru. We also present population counts from censuses at the national level and, whenever possible, at the regional level too; as well as the latest population estimations for inter-census years taken from "United Nations World Population Prospects, 2010 Revision", published in 2012 by the Population Division-Population Estimates and Projections Section (http://esa.un.org/wpp/index.htm). It also provides a database with titles and authors of academic publications on mortality for the countries of interest. Finally, we provide our own estimation of adult mortality under-registration following the Hill, You and Choi (2009) combined method (see more on Methodology for Completeness of Adult Death Counts Coverage section).

Our sources, the official vital records for each country, are collected in different formats. Most of death records were provided from each national statistical authority, in other cases they were downloaded from their official websites or from a particular United Nations Office that keeps records directly sent from the official statistical offices, such as UNFPA-United Nations Fund for Population Activities or WHO-World Health Organization. In rare cases, data was digitalized from printed versions of the national statistical authority. Each data series reports the original sources for each year. In all cases the data with causes of death are defined as the underlying cause of death, coded by the relevant national authority. Underlying cause of death is defined as "the disease or injury which initiated the train of morbid events leading directly to death, or the circumstances of the accident or violence which produced the fatal injury" following the International Classification of Diseases (ICD).

Each country's particularities on the codification of causes of deaths are specified in their own sources and data treatment. Geographic definitions use what countries have already defined, from the national borders to the regional limits. Sub-national borders are defined as they currently exist. We want to point out that although all series are taken from official sources, they do not necessarily coincide. Yet, most differences are very small and are the result of our double checking process (see data treatment section) that took us to correct processing mistakes. In all cases we strongly preferred to keep fidelity to the official figures rather than making a correction. The most common mistakes were typos or additions that did not added up. For instance, if a data point was published like 5 instead of 8, it was corrected. Or if a decimal point was missing then we corrected it too (5.00 instead of 500). All sources, changes and data treatment are explained for each country, we encourage users to read the particularities of the country of your interest.

3.2 Description of the Data Tables and Data Presentation

Tables include what is described in their titles in excel format. For each excel file you will find different tabs corresponding to a particular year. For instance, death records by age, sex and province for Peru are found under the sub-title Peru in our link "Country Data". After you click in PER, Deaths by age, sex and region the table with the data will open-up and each tab is named with the information for the corresponding year.

3.3 Data Treatment

Original data was provided in electronic format or digitalized from a printed version. Electronic format data is usually provided by the official statistical offices and in rare cases is taken from datasets built up by WHO or UNFPA. In particular, data published by the WHO-Mortality Database (http://www.who.int/healthinfo/morttables/en/) and/or the United Nations Demographic Yearbook – DYB, for the corresponding year, (http://unstats.un.org/unsd/demographic/products/dyb/dyb2.htm). Once data

is at hand we processed the data to arrange each of the tables produced for each year. After that we double check for zero mortality cases due to maternal mortality causes for males, or women below age 15 or above age 50. We also double checked that there are no death records for perinatal mortality causes above age 2. In the very few cases those are found, then, those records are re-classified as unknown age and sex for the former and unknown age but kept the same sex as originally recorded for the latest.

After data is double checked by these particular causes, we prorated records with unknown age and sex, exclusively for the table that contains death records by age, sex, region and cause of death. From that table all other tables are built up until reaching the table with the least levels of disaggregation: national death counts by age and sex. Notice that unknown causes of death or unknown region are not prorated, as the use of codes for ill-defined and unknown causes of death, must be taken into account for the construction of mortality rates for specific causes, particularly if you plan to compare across countries.

Whenever the data is found in paper, rather in electronic format, the treatment is exactly the same after the data has been carefully digitalized for all files and columns exactly as they are printed. Only one additional check is done before the check on causes of death: we double check that all summations coincide with those transcribed, so we can correct for typos and processing mistakes.

3.4 Methodology for Completeness of Adult Death Counts Coverage

In developing countries, mortality estimates and knowledge of levels and trends of mortality are limited by the quality of data (Hill, 2003; Luy, 2010; United Nations 1997; 2002). The most common problems faced in these countries are incomplete coverage of vital registration systems, errors in age declaration for both population and death counts, and lack of information on causes of deaths. In recent years, collection of data for death counts has improved, but there are still limitations for studying mortality in several parts of the world. The problem is more complex when studying small areas and sub-national populations where in addition to data limitations, statistical and demographic methods also have

strong limitations. Thus, public health administrations are faced with limited information to allocate resources and it is also difficult to study the progress of public policy interventions for several countries around the world, limiting the action of government agencies in improving the quality of life of these populations.

The death distribution methods are widely used to estimate adult mortality (Timaeus, 1991). They compare the distribution of deaths by age with the age distribution of the living and provide age patterns of mortality in a defined reference period. There are two major approaches: the General Growth Balance Methods (Hill, 1987), and the Synthetic Extinct Generation method (Benneth & Horiuchi, 1981). The death distribution methods make several strong assumptions: that the population is closed to migration, that the completeness of recording of deaths is constant by age, that the completeness of recording of population is constant by age, and that ages of the living and the dead are reported without error.

The Bennett and Horiuchi (1981) method is a synthetic cohort analog of Vincent's (1951) method of extinct generations, known as Synthetic Extinct Generations (SEG) method. In this methods, age-specific growth rates are used to convert an observed distribution of deaths by age into the corresponding stationary population age distribution. Since in a stationary population the deaths above each age x are equal to the population aged x, the deaths in the stationary population above age x provide an estimate of the population of age x. The completeness of death registration relative to population is estimated by the ratio of the death-based estimate of population aged x to the observed population aged x.

The General Growth Balance (GGB) method (Hill 1987) is a generalization to all closed populations of the Brass (1975) Growth Balance method. The Demographic Balancing Equation expresses the identity that the growth rate of the population is equal to the difference between the entry rate and the exit rate. This identity holds for open-ended age segments x+. The entry rate x+ minus the

growth rate x+ provides a residual estimate of the death ratex+. We can then estimate the residual from population data from two population censuses and compared to a direct estimate using the recorded deaths (from the census or vital registration), and comparing these two records we can estimate the completeness of death recording relative to population.

Hill, You and Choi (2009) proposed that the combination of SEG and GGB might be more robust than either one individually. The combined method consists of first applying GGB to estimate any changes in census coverage (k1/k2), using the estimate to adjust one or other census to make the two consistent, and then applying SEG using the adjusted population data in place of the reported.

5- References

Condran, Gretchen A., Christine L. Himes, and Samuel H. Preston (1991). "Old-age mortality patterns in low-mortality countries: An evaluation of population and death data at advanced ages, 1950 to the present." *Population Bulletin of the United Nations*, No. 30, pp. 23-60.

Dorrington RE. 2012a. "General Growth Balance". In Moultrie TA, RE Dorrington, AG Hill, KH Hill, IM Timæus and B Zaba (eds). Tools for Demographic Estimation. Paris: International Union for the Scientific Study of Population.

Dorrington RE. 2012b. "Synthetic extinct generations". In Moultrie TA, RE Dorrington, AG Hill, KH Hill, IM Timæus and B Zaba (eds). Tools for Demographic Estimation. Paris: International Union for the Scientific Study of Population.

Hill, K.; You, D.; Choi, Y. Death Distribution Methods for Estimating Adult Mortality: sensitivity analysis with simulated data errors. Demographic Research, vol. 21 (9), pps 235-254 , 2009.

Hill, Kenneth H. (1987). "Estimating census and death registration completeness." Asian and Pacific Population Forum 1(3): 8-13+.

Methods for Estimating Adult Mortality. New York: United Nations (Sales No. E.83.XIII.2).

Preston, S. H., and N. G. Bennett (1983). "A Census-based Method for Estimating Adult Mortality." *Population Studies* 37(1): 91-104.

Preston, S., and K. Hill (1980). "Estimating the Completeness of Death Registration." *Population Studies* 34(2): 349-366.

Shryock, Henry S., and Jacob S. Siegel (1973). *The Methods and Materials of Demography*. Washington, DC: U.S. GPO. [especially pp. 114-128]

Timæus, Ian M. (1991). "Measurement of adult mortality in less developed countries: A comparative review." *Population Index* 57(4): 552-568. United Nations (2002).

6. Website

Figure 1 – Main page – Latin America Human Mortality Database (www.lamortalidad.org)

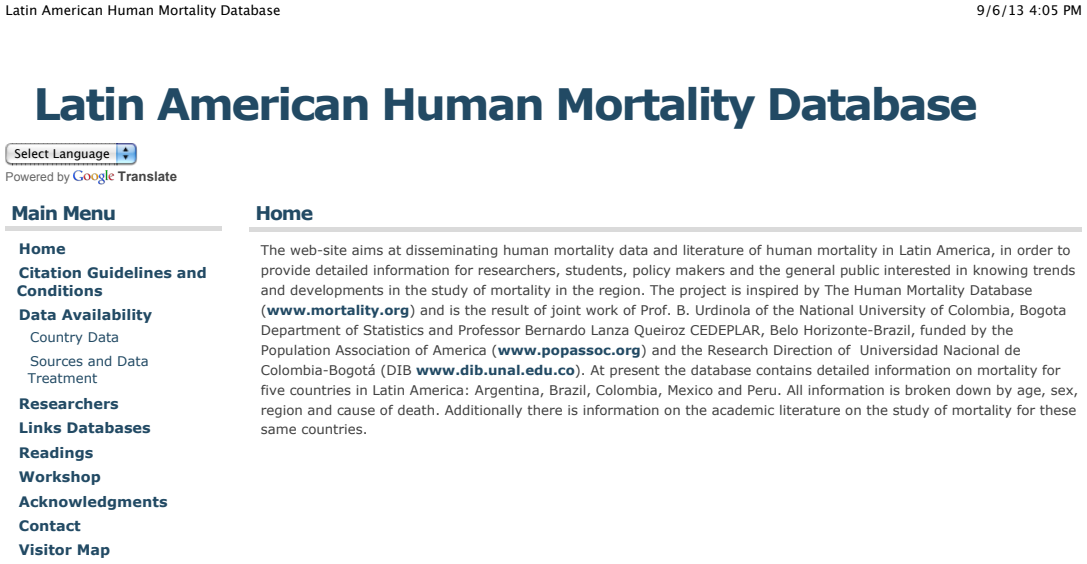

Figure 2 – Data availabilty – Latin America Human Mortality Database (www.lamortalidad.org)