# The Latin American Mortality Database (LAMBdA)

Methodological Document[1]

Version II: August 2021

LAMBdA Team[2]

[1]Website: www.ssc.wisc.edu/cdha/latinmortality2

[2]LAMBdA Team: Palloni A., Pinto G., Beltrán-Sánchez H., and Verhulst A.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

This document is a detailed description of LAMBdA's genesis, sources, and nature. It is organized in a manner that will facilitate understanding of the progression of steps taken to construct the life tables. Inevitably, there is some overlap between some chapters as each requires a foundation described in previous chapters and/or anticipates developments taken up in subsequent chapters. To facilitate reading of the material, Chapter 2 is built as a road map and is a preview that summarizes in a non-technical manner strategies to estimate life table parameters and the order in which they were used. Chapter 3 is a thorough but lengthy and somewhat technical description of methods used to estimate LAMBdA's life tables throughout the 1850-2010 period. It concentrates on estimation of adjustments to build adult mortality life tables and identifies how these adjusted adult life tables could be blended with estimates of infant and child mortality to complete the life tables. Chapter 4 is a review of multiple methods used to estimate infant and child mortality. Chapters 5 and 7 are renditions of specific procedures to handle two problems in the construction of single age life tables, splitting of raw data available in five year age groups and treatment of the open age group. In particular, Chapter 5 is an evaluation of performance of alternative methods to disaggregate raw deaths and population counts grouped in coarse categories into single year counts and Chapter 7 compares results of the method followed in LAMBdA to estimate mortality in an open age group (85+) with alternative treatments. Chapter 6 is a description of techniques used to create the module of mortality by causes of death included in LAMBdA. Chapter 8 is a detailed review of results of a full array of tests and consistency checks of the entire database and comparisons with other mortality databases. Chapter 9 is a set of country reports and, finally, Chapter 10 contains supplementary and supporting material referred to in various parts of the main document.

# Chapter 2

# LAMBdA methods and data sources: a brief summary

This first chapter of LAMBdA documentation is a road map. In it we offer a brief review of stages required to produce LAMBdA's life tables and is intended to be a summary and general overview rather than a technical compilation of methods used, the subject of subsequent chapters of the documentation. To help the reader, Figure 2.1 displays a diagram disclosing key junctures in a decision tree that leads to the construction of LAMBdA's life tables and the set of procedures associated with each resulting branch. The decision points in the figure are a function of the nature of available vital statistics and census counts for a given period and the age groups used in each case. Thus, for example, in countries with no vital statistics before 1950, mortality at ages 5 and over is estimated using variants of the generalized ogive procedure (see Chapter 3) whereas mortality under age 5 was estimated using a country-specific pool of adjusted direct and indirect estimates in combination with suitable models.

## 2.1 Definitions

In this section we clarify the meaning of a handful of terms used throughout the documentation and define precisely a few useful concepts. To facilitate integration with subsequent sections of the chapter, the order followed in the description of terms and concepts reproduces approximately the order of production of life tables. Thus, we start with the raw data and end with a single age, single calendar year (adjusted) complete life table.

1. **Raw data:** The raw data on which LAMBdA rests consists of information collected from vital statistics and census publications (or directly from vital statistics and population census officials). It includes census figures for population by age and sex as well as counts of births and deaths. For the most part the information is available in five year age groups although for years after 1970 but only for some countries we use more fine-grained data by single years of age.

   As will become clear in the rest of the documentation, there is a stark contrast between raw data available before 1950-1965 for most countries. To minimize cluttering of labels notation and unless we need precise definition of timing, we will use the expression "before 1950" to refer to periods of time that, in some of these countries, may precede years as late as 1965.

2. **Single age data:** The bulk of procedures to assess data quality and to adjust for errors are applied to population and vital events counts in single years of age, not to the raw data directly. Single age estimates were obtained from the raw data after applying standard cubic interpolation (Sprague multipliers). It was a choice dictated by pragmatic criteria and mostly to ensure we adhere to well-established practices. However, because cubic interpolation is not the only way of splitting events or counts grouped in broad age categories, we also employed alternative methods and systematically compared results. Chapter 5 of this documentation describes the outcome of these tests. By and large, they reveal that alternative procedures produce estimates that are no different from those included in LAMBdA and, when they are, differences are inconsequential.

3. **Adult pivotal life tables:** As is well known, the key, albeit not the only, contrast across systems of models of human mortality patterns is rooted in the relations between early childhood and adult mortality. LAMBdA life tables were constructed maximally avoiding assumptions about such relations. To do so, we kept separate the estimation of mortality above (adult) and below (childhood) age 5. Attaining compartmentalization of early childhood and adult mortality required to first build an incomplete, adult life table representing 5 and over mortality, generate independent estimates of infant and early child mortality and, finally, blend together these two sets of estimates into complete life tables.

   Adult *pivotal* life tables are single age life table starting at age 5, with an open age group at 85+. They are estimated under one of three sets of conditions described below.

   - *Two successive census and complete counts of intercensal deaths*: these are cases when we can avail ourselves of two consecutive censuses and counts of intercensal deaths. We adjust adult mortality rates of the reference intercensal period for (a) relative completeness of death registration and (b) adult age misreporting (see Chapter 3). The adjustment is based on census counts at both ends of the intercensal period and the entire series of intercensal deaths. The pivotal life table is computed using the adjusted average rates for the intercensal period centered in the middle of it.

   - *Single population census and death counts in neighboring years*: these are cases when the data available is reduced to one census and a limited count of deaths within a period of three to five years centered around the census (pivotal) year. We adjust adult mortality rates centered on the census year for both relative completeness and age misreporting (See Chapter 3).

- *Two successive censuses but no intercensal deaths counts*: these are cases when there is information for two consecutive censuses separated by at most 15 years but no or only incomplete information on death counts during the intercensal period. We first estimate an adult pivotal life table centered in the middle of the intercensal period using the generalized ogive procedure. We then use each census and information on intercensal rates of growth to estimate life tables for each of the two censal years.

4. **Mortality rates in the open age group:** An important part in the construction of the adult pivotal life tables is the assessment of mortality in the so-called open age group. Although, this quantity has little influence on estimates of most key life table functions of interest, it does have an impact on life table functions evaluated at older ages, particularly in populations with low levels of mortality.

   It is only after 1970 that we are able to secure data for single years of age extending beyond age 85. In all other cases the information we process ends with the age group 85+. The techniques we use to adjust adult mortality for completeness and age misreporting are applied to observed mortality rates, not to counts of deaths (or populations) and populations separately. As a consequence of this, we cannot deploy procedures, such as extinct generations, that rely on counts of deaths to fine-tune estimates of mortality rates in the open age group. Instead, we adjust the observed mortality rate in the open age group applying the same adjustments factors associated with other adult ages. To examine the sensitivity of our estimates of older ages life table functions to this treatment of the open age group, we employed alternative methods. A comparison of results across methods is discussed in Chapter 7 of this documentation. The conclusion we draw is that with no exceptions, and when considering alternative techniques that could only be employed after adjusting for completeness and age misreporting, only small differences are produced and these have trivial effects even on measures of relevance to old age mortality.

5. **Adult life tables for single calendar years:** Once the set of adult pivotal life tables is in place, we construct non-pivotal, single calendar year, life tables. LAMBdA includes two types of non-pivotal complete life tables. The first are those computed from pivotal life tables that were not estimated availing ourselves from continuous intercensal vital statistics. These pertain mostly to years before 1950 and include life tables computed using the ogive procedure or specialized adjustments. To obtain single calendar year life tables from these pivotal life tables separated by discrete and unequal time intervals we use local least squares fitting on age-specific mortality rates from the available pivotal life tables. The predicted values were then chained together to calculate life tables for single calendar years.

   The second set of non-pivotal, single calendar years life tables, built mostly in the period after 1950, were calculated using adjusted yearly vital statistics from age 5 to 85+. In all cases we first estimate yearly population counts for intercensal years using simple exponential growth functions. We then calculate yearly mortality rates with observed death counts and adjust them for completeness and age misreporting with adjustment factors identical to those applied for the construction of the pivotal life table in the corresponding intercensal period.

The above definitions apply to adult life tables only. To generate *complete* life tables, e.g. including under five mortality, we first generate independent estimates of infant and child mortality.

6. **Infant, early child and childhood mortality:** Unless otherwise noted, infant and early child mortality rates refer to mortality rates at ages 0 and in the age interval 1-4 respectively. We use the term *childhood mortality rate* to refer to the under 5 mortality rate.[1] In all cases infant and childhood mortality rates are estimated separately and quasi independently of adult mortality rates.[2] The estimation of these rates requires multiple sources of data and rests on the application of a blend of methods briefly described below and more fully explained in Chapters 3 and 4.

7. **Complete pivotal life table:** To generate a complete pivotal life table we anchored estimates of infant and childhood mortality rates to pivotal years and then chained together the adjusted mortality rates in the adult pivotal life tables and the anchored estimates of infant and childhood mortality for the pivotal year.

8. **Complete abridged pivotal life tables:** An abridged pivotal life table is a pivotal life tables in 5-year age groups (except in years 0 and 1-4). They were computed from a complete pivotal life table using routine calculations and operations described in Chapter 3. All abridged life tables end with the open age group 85+. To establish consistency between single-age and abridged life tables, we retained the functions ($l_x$, $T_x$, $E_x$) at ages 0, 1, 5, . . . , 85+ from the single year pivotal life tables.

9. **Complete single age, single calendar years life tables:** By construction, estimates of infant and child mortality are generated first for pivotal and then for single calendar years and separately and quasi independently from estimates of adult mortality. To construct a complete single age, single calendar year life table we chain together the mortality rates of the single age adult life tables and the corresponding single calendar year estimates of infant and child mortality.

10. **Cohort life tables:** Complete and right-censored birth cohort-specific life tables in single years of age were computed by chaining together age-specific mortality rates in

[1] Because under 5 mortality rate, ${}_4Q_0$ is uniquely defined by infant and early child mortality rates, ${}_1Q_0$ and ${}_4Q_1$ respectively, we only need two of these rates to determine the third. With few exceptions explicitly flagged in the documentation, we will focus on infant mortality rates and childhood mortality rates, as these will be the quantities retrieved from various sources.

[2] Mortality model 'quasi' independent estimates of infant and childhood mortality arise when estimates associated with an external model mortality pattern are not used in the construction of a life table but do contribute to a pool of estimates from which a final point estimate of infant and childhood mortality is obtained.

calendar year-specific life tables across successive calendar years. The resulting age-specific mortality rates were assumed constant within single calendar and years of age and transformed into the remaining functions of a birth cohort's life table.

## 2.2 Estimation of adult pivotal life tables

Pivotal life tables functions for ages $x \geq 5$ after 1950 were estimated using raw data from vital statistics and population censuses and the single age information derived from cubic interpolation procedures when these were needed. With a few exceptions, adult pivotal life tables before 1950 were estimated using (i) the adult age distribution from one or two population censuses, and (ii) death counts for years centered on a population census year.

### 2.2.1 Adult pivotal life tables before 1950

To estimate adult life tables functions for the period before 1950 we followed one of two procedures, a generalized version of Coale-Demeny ogive method (Coale and Demeny, 1967) and Brass intercensal growth method (Brass, 1979). We briefly review these below.[3]

1. **Generalized ogive.** Application of the generalized ogive method (see Chapter 3) rests on an *a priori* selection of a model mortality pattern. The estimation of life tables with the ogive method uses simultaneously mortality rates from the South and West Coale-Demeny life table models as well as from the Latin American model of the United Nations system (Coale et al., 1983; United Nations, 1982).[4] The method uses estimates of intercensal growth computed from two or more consecutive population censuses. These observed growth rates were then adjusted using local regression to obtain a smoothed time trend of intercensal growth rates.[5] In all cases, the generalized ogive procedure employs as inputs the smoothed (not the observed) values of the intercensal growth rates.

   Life tables computed with the generalized ogive method and each of three mortality models yield three alternative set of estimates of life expectancy at age 10 (one per age group considered including 5+, 10+, . . . , 75+). We then select the median of estimates of $e_{10}$ corresponding to age groups 5+, . . . , 60+ by sex and discard those that depend on older age groups. We then use these values as inputs in the Coale-Demeny set of equations to generate estimates of $Q_x$, $l_x$, and $L_x$ or, alternatively, as entry point in the Latin American model of the United Nations.[6]

[3]Chapter 3 contains a full description.

[4]Although the Latin American model of the United Nations life table system reproduces closely some features that characterize the South Model of the Coale-Demeny system (relations between infant and child mortality) it also resembles the North model regarding the relation between early childhood and adult (but not old age) mortality. Thus, using the Latin American model is equivalent to using a blend of the South and North model of mortality.

[5]In four countries, Argentina, Brazil, Cuba and Uruguay, all intercensal growth rates were also adjusted for migration (see Chapter 3).

[6]To generate single age life table functions for the life tables thus identified by the ogive method, we apply interpolation to the values $M_x$ from age 5 to 85+.

Because the procedure can be applied to the middle of the intercensal period as well as to the census years bracketing the intercensal period, we generate a total of three sets of candidate pivotal life tables each (one per model of mortality) by single years of age $x \geq 5$, one located in middle of the intercensal period and two anchored at the boundaries of the intercensal period.

For each pivotal year (and the associated bracketing censal years) we choose the median of the three model-specific estimates of $M_x$ and compute the full set of adult life table functions for all ages $x \geq 5$. Thus, the age pattern of adult mortality embedded in the resulting life table is a mixture of the three model patterns that served as basis of the computations.

2. **Brass's method.** Brass's method for estimating completeness of death registration is also used to generate a pivotal life tables before 1950. Unlike the generalized ogive, the method requires knowledge of age-specific death counts for three years centered in the census year but no exogenous model mortality pattern. Thus, it can only be employed in country years with available vital statistics during the period of interest. Brass' method requires estimation of a regression equation of accumulated deaths rates above age 5 on age specific 'birth rates' for populations above age 5. The constant of the model is an estimate of the rate of increase whereas the slope is an estimate of the (reciprocal) of the completeness of death registration (See Chapter 3).

The methods described above yield life tables functions for adult ages, e.g. $x \geq 5$. However, the generalized ogive method results in a byproduct, namely, three alternative estimates of infant and childhood mortality—each associated with the three models of mortality patterns. These estimates are not discarded but are instead included as elements in a country-specific pool of candidate estimates of infant and childhood mortality for the corresponding pivotal years. This pool is then used to generate the complete pivotal life table (see Chapter 4).

### 2.2.2 Adult pivotal life tables after 1950

Pivotal adult ($x \geq 5$) life tables after 1950 were calculated using single-age intercensal death rates from age 5 to 85+ adjusted for completeness (Bennett-Horiuchi) and age misreporting (see Chapter 3). Because the bulk of raw data consists of 5-year age groups specific counts, we first estimated single year age specific death and population counts using Sprague multipliers (see Chapter 5). The resulting single year of age population and death counts are the input to which we apply both Bennett-Horiuchi method to adjust for completeness of death registration and the methods to adjust for adult age overstatement.

As is the case for the period before 1950, the single-year of age life table functions computed with the above described procedure are, however, incomplete since they are only valued at ages $x \geq 5$. To construct a complete life table we add estimates of child mortality independently obtained (see below Chapter 4).

## 2.3 Estimation of infant and child mortality

A key feature of LAMBdA is that estimation of adult mortality, e.g. for ages $x \geq 5$ and infant and early childhood mortality is carried out separately and quasi independently. The aim throughout is to avoid the need to invoke compromising assumptions about the relation between childhood and adult mortality. The life tables included in LAMBdA are not completely model free, as the use of the generalized ogive methods requires at least three alternative model patterns to inform the unobserved mortality level and pattern in an intercensal period. However, even in these cases, the overall influence of model choice is less than if the estimates had been used directly as an input parameter in the life table and affects marginally the relations between childhood mortality rates and mortality rates at ages over 5 embedded in the data base. A full description of stages to estimate childhood mortality throughout the period of interest is in Chapter 4.

To estimate pivotal and yearly values of infant and early childhood mortality throughout the period under study we used an integrated procedure that includes two components. The first is a set of rules (and computations) to generate alternative estimates of infant and childhood mortality rates included in country-specific pools of feasible candidate estimates. The second is a methodology that uses the pool of estimates thus assembled to generate final point estimates for each country year. Below we describe each component in turn.

### 2.3.1 Construction of country-specific pools of candidate estimates

The pool of feasible estimates of infant and childhood mortality, e.g. for ages 0 and 1-4, for each country year consists of values retrieved from four different sources. In all cases infant and childhood mortality are treated separately.

1. *Estimates associated with model life tables*: as described before, the application of generalized ogive to estimate adult pivotal life tables yields as a byproduct three sets of alternative infant, ${}_1Q_0$, and childhood, ${}_5Q_0$ mortality. Whenever the ogive technique was used three mortality models were employed, two associated with the West and South families from the Coale-Demeny system and one associated with the Latin American families from the United Nations model system (Coale et al., 1983; United Nations, 1982). Upon choosing a single set of adult mortality rates by combining three sets of estimates, each associated with a model of mortality, we are left with three estimates of infant and child mortality rates, one per model of mortality. These three estimates were treated as feasible estimates for the country year and included in the pool of candidate estimates.

2. *Estimates from adjusted observed values of* ${}_1Q_0$ *and* ${}_4Q_1$: observed values of infant and early child mortality rates were computed from available vital statistics (observed deaths in ages 0, 1-4, and yearly births) and population census counts for age groups 0 and 1-4, separately by sex. We then generate adjusted values for the period after 1950 using (when available) either indirect estimates or birth histories from censuses and surveys covering the period 1970-2010. We compute adjustment factors or ratios of estimated to observed values for all years in which both quantities are available.

These adjustment factors produce a fairly long time trend from which we can predict yearly values after 1950 and, if needed, extrapolated values for the period before 1950.

3. *Estimates from surveys and censuses*: these were obtained for the period after 1950 using standard indirect techniques based on information on children ever born and children surviving retrieved from surveys and censuses. These estimates were combined with direct estimates computed from birth histories available in selected surveys. Although availability varies by country, in most cases the estimates comprise a large and compact mass of observable quantities that could stand on its own as a foundation to derive final point estimates.[7]

4. *Third party estimates*: these are estimates obtained from other sources, computed according to well-described methods and based on accepted and, in most cases, standard procedures. They are available for selected countries, originate in multiple sources (mostly indirect methods), and have been used frequently as accurate estimates by scholars, government organizations, and international agencies alike.

### 2.3.2 Computation of point estimates

In each country the pool of estimates was reviewed and values that appeared to be implausible were discarded. The criteria we used to discard values are the following: (a) estimate is outside a band contained within 2 standard deviations from an average of values within a five year interval centered on the year of the estimate; (b) estimate violates progression of a time trend in adult and child mortality detectable within five years after and before the year to which the estimate applies, and (c) estimate is inconsistent with adult mortality estimated for the same year, e.g. it is outside a range predicted by estimates of life expectancy above age 10 in each of the three model mortality patterns used.

The pool of heterogeneous estimates of infant and childhood mortality described above forms a time series for each country that may stretch from as far back as 1850 and reach up to 2010. However, since the number of observation points in each of the four classes of estimates described above differ by countries, the number of points of support for estimation also varies and potentially contribute to intercountry heterogeneity. The series also contain missing values, it may be irregular or noisy, particularly for periods before 1920, and more so for some countries than others. To better exploit the more robust subseries comprising the years 1950-2010, a period with less uncertainty and with less intercountry heterogeneity in missingness, we divide the task of production of point estimates into two parts, one per period.

1. **Years 1950-2010**: The country series for this period are more homogeneous and less subject to noise than the series for the period before 1950. The values included are

[7] Indirect methods of estimation based on children ever born and surviving yield estimates of mortality below ages 1, 2, 3, 5 and 10. These can be converted into estimates of the desired quantities, namely, ${}_1Q_0$ and ${}_5Q_0$ only after assuming a model of mortality pattern. In all cases we use MORTPAK to generate estimates and accepted as feasible those associated with the West and South Coale-Demeny models and the Latin American model of the United Nations. Each pair of estimates thus enters into the pool of candidate estimates.

all estimated from well established surveys, such as WFS or DHS, or from census samples and are dependent on extensively tested direct and indirect methods. But they also contain errors associated with sampling variability, disparities in regional coverage, biases due to idiosyncracies of responses or questionnaires, etc. To attenuate somewhat the influence of these factors we first estimated local regressions on the pools of country-specific observed values for the period. While local regression enables us to smooth the time series, it does not help us with missing values for years in which no estimates can be produced from the various sources. To resolve this we resorted to fitting splines with a large number of knots. The splines returned predicted values for calendar years with missing estimates.

2. **Years before 1950**: The above described point estimates populate the period 1950-2010 but are not by themselves sufficient to help us generate estimates during the period 1850-1960. To do so we add to each country's point estimates for 1950-2010 the pool of estimates for the period 1850-1950. We then model the entire series with a three parameter Gompertz function and retrieve predicted values for years before 1950 only but preserve the point estimates for the period after 1950 that contributed to the global (1850-2010) Gompertz fit.

3. **Consistency between trends in childhood and adult mortality:** One of the parameters of the Gompertz function is a ceiling parameter that identifies the upper bound of early childhood mortality associated with the entire time series. Because the parameter estimate is independent of estimates of adult mortality, the trends in early childhood mortality may be inconsistent with the one observed for adult mortality. In cases of inconsistency we opted to choose the trend embedded in the adult pivotal life table on the grounds that adult mortality was more accurately estimated. Thus, if the adult mortality trend suggests a stationary mortality regime before a particular year, say $Y$, the levels of under 5 mortality for years before $Y$ were fixed at a value identical to the point estimate from the childhood mortality trend evaluated at year $Y$. This decision will lead us astray whenever the underlying assumption that adult and under 5 mortality trends are harmonic, is incorrect. However, in the three countries where we imposed this choice (Guatemala, Honduras, Nicaragua) differences between the ceiling values adopted and those predicted by the trend were less than 4 percent. The consequence of this choice is that some countries have an invariant life table for periods preceding the one when the ceiling or threshold mortality is thought to have been reached. These life tables are meant to represent average levels during the period, not the exact magnitude of mortality in any one year.

## 2.4 Nature of the estimated life tables: the problem of model dependence

As stated before, a central preoccupation in the construction of LAMBdA is to introduce little or no mortality model dependencies. This requirement may have been violated in a handful of cases described below. We hasten to add, however, that even if this is the case, we believe we largely achieved the original goal. In fact, in a separate study (LAMBdA team, 2020) we constructed models of mortality using the LAMBdA database and estimated patterns distinct from those embedded either in the Coale-Demeny or the United Nations family. The differences are far from massive but are consistent with peculiarities of determinants and timing of mortality decline in these countries that are not embedded in other mortality patterns.

### 2.4.1 Quasi independence of early childhood and adult mortality

One of the key concerns that drove the choice of methods chosen to build the database was avoidance of assumptions about model patterns, either those linking early and adult mortality or those applicable to adult and older age mortality. By definition the original ogive method requires *ex ante* identification of a model pattern thus fixing the relation between early and adult mortality and determines the features of adult and older age mortality. In our application, however, we assume not one but three different mortality patterns which yield three candidate series of child mortality. A partial dependence on model patterns slips into the methodology because the estimates of infant and child mortality associated with the ogive procedure, though not directly used, are not discarded but instead contribute to the pool of feasible estimates of infant and child mortality that eventually generate final estimates.

The same applies to the use of alternative model mortality patterns to transform parameter estimates from indirect methods that refer to accumulated mortality below some ages (2, 3, 5, 10) into estimates of infant and childhood mortality rates. In this case we also retrieve three alternative sets of estimates but none was used directly and to the exclusion of the others and instead they each contributed to the pool of candidate estimates from which final point estimates were computed (See Chapter 4).

### 2.4.2 Adult mortality and treatment of the open age group

Additional model dependence seeps through in the estimation of mortality using the ogive procedure because all three mortality models we use as support assume a Gompertz mortality pattern applies to mortality rates before and after the open age group. Thus, by construction the corresponding pivotal life tables contain a latent pattern that connects mortality rates between ages 45 approximately and 84, on one hand, and at ages above 85. It should be noted, however, that the adult rates between ages 45 and 84 *are not* predicted by a Gompertz. It is only the magnitude of mortality rates above age 85 that are functions of the Gompertz fit.

The life table functions estimated for the period after 1950 and based on adjusted vital statistics are, in principle at least, model-free. However, they are also influenced by assumptions that restrict the range of estimates of mortality rates in the open age group. All LAMBdA life tables based on adjusted vital statistics constrain the mortality rate in the open age group to be equal to the adjusted rate, that is, the product of the observed rate in the open age group and the adjustment factor that accounts for incomplete death registration and age overstatement. This strategy is different from computations carried out in other life tables systems (Coale-Demeny, United Nations , HMD) where the mortality rate in the open age group is a function of additional fitting(s). In some cases a parametric model, for example Gompertz, is fitted to mortality rates above some arbitrary adult age and then single year mortality rates over age 85 are computed directly and accumulated to yield a new estimate of the mortality rate in the open age group. This procedure is followed both in the Coale-Demeny and the United Nations life tables. In other cases, a logistic function is fitted to rates computed over some low boundary age (usually 55 or 60). Once the parameters of the logistic function are in place one can predict mortality rates above 85 and, here again, estimate a (fitted) value of the mortality rate in the open age group. Finally, a third method consists of reconstructing the survivors column of a life table at older ages using observed deaths counts starting at some arbitrary age and then fitting a logistic (or related) function to the resulting survival function. Unlike the other two procedures, this third method is incompatible with our strategy since it requires the assumption, untenable in our case, of no distortions due to age overstatement.

To alleviate concerns about the problem of model dependence related to the treatment of the open age groups we carried a number of tests described in Chapter 7 of this documentation. There we show that the method employed in LAMBdA produces similar results to those we would have obtained had we used alternative methods consistent with the principle of preserving estimated adjustments for completeness and age overstatement.

## 2.5 Mortality by causes of death

LAMBdA contains a module on causes of deaths that consists of several components: (i) raw counts of deaths by groups of causes and age groups beginning approximately in 1950, (ii) mortality rates by causes adjusted for completeness and age misreporting, and (iii) mortality rates adjusted for completeness, age misreporting, and ill-defined (see Chapter 6 for more details).

### 2.5.1 Groups of causes of death

The data on causes of death begin in earnest around 1950 and spans a total of 5 different WHO classification systems (ICD 6th and 7th Revisions, A lists; 8th revision, A list; 9th revision B list; and 10th revision, Detailed lists and List 1). Any grouping of causes of death we decide to choose requires careful conciliation of definitions across these sometimes sharply different, classification systems. Although we also implement two alternative classifications (available on request), we chose to construct one single abbreviated classification comprising 27 groups of causes that includes a total and a group of ill-defined causes. The principles used to arrive at this classification and the rules that guided conciliation across the multiple ICD revisions are set forth in Chapter 6 of this documentation.

### 2.5.2 Adjustments: completeness, age misreporting and treatment of ill-defined

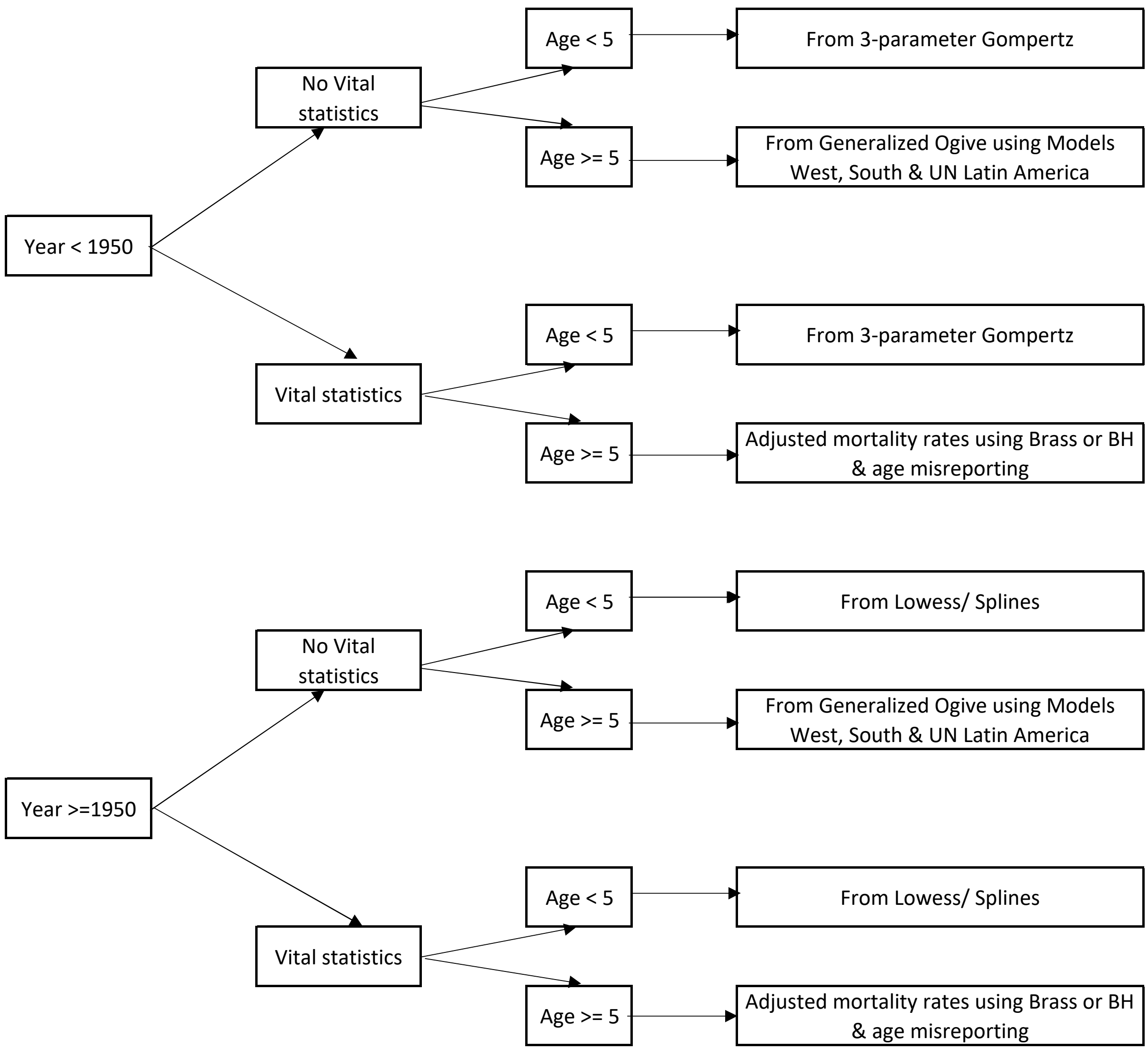
Mortality rates by groups of causes were computed from raw counts and then modified using adjustment factors employed to construct age-gender-specific mortality rates in pivotal life tables. These adjustments assume that errors of relative completeness of death registration and those associated with age misstatement are independent of causes of death.

An additional adjustment was applied to redistribute the group of causes referred to as 'senility' and 'ill-defined' in most classification systems. In some country years, but particularly so before 1970, the size of this category is substantial. Its existence and significance is attributed to variable quality of medical certification of registered deaths. In some cases treating it as a separate group is a satisfactory solution. In others, for example when one decomposes changes in life expectancy by age and cause, it is advisable to remove the ill-defined group by redistributing the causes into well-defined groups. This can be accomplished in a number of ways. LAMBdA adjustment followed a procedure that departs from the usual "proportional distribution method" whereby ill-defined causes are allocated to well-defined groups according to the observed distribution in well defined groups. Instead of this, we follow a regression-based procedure set forth in Chapter 6 of this documentation.

## 2.6 A final warning: the nature of LAMBdA and its potential uses

The above summary highlights the main limitation of LAMBdA as a mortality database. While life tables estimated for the period after 1950 can be freely used for a very broad set of purposes, including the study of yearly fluctuations, those that compose the body available for the period before 1950 should be used with more caution as they depend to a lesser extent on observed fine-grained vital events and are can only retrieve approximate mortality levels for periods of time not shorter than five years. As a consequence, life tables for the pre-mortality decline period may well be a precise portrait of broad levels, trends, and patterns, but fall short of being an accurate rendition of very short-term fluctuations, sudden shifts, and abrupt dynamics.

Figure 2.1: Graphic representation of decisions points to construct LAMBdA life tables



# Chapter 3

# Estimation of life tables 1850-2020

## 3.1 Introduction

LAMBdA is a database designed to inform mortality changes in the Latin American and Caribbean (LAC) region since 1850 approximately, a time that, for most countries in the region, is in close proximity to the end of the colonial stage, the aftermath of wars of independence from Spanish and Portuguese rule, and the establishment of nation states.[1] LAMBdA describes approximately 170 years of history albeit not always with the same level of detail nor in the same depth. There is abundant, but defective, information for the period following World War II, less so for the first half of the Twentieth Century, and rarely for the post-independence era (1830-1900). For years before 1950 we directly estimate levels and age patterns of mortality that reflect the experience of populations during periods never smaller than three years, in most cases of five to ten years and, when other options are unfeasible, over time intervals between 10 to 15 years. In contrast, for the period after 1950 there is yearly mortality data and it is possible, in principle at least, to estimate life tables reflecting mortality experiences over very short and contiguous periods of time. The last and final LAMBdA update was issued in Juy 31, 2021. This update includes country's most recent (adjusted) life tables computed with available censuses, population estimates or projections and vital statistics for the period 2010-2019. There will be no more updates as the database will be closed as of August 31st, 2021.

The construction of a life table requires information on events (deaths by age) and exposure (population by age). When accurate vital statistics and population censuses counts (or national sample surveys) are available, the requisite age specific mortality rates can be readily computed. These are then transformed into standard life tables functions, such as conditional probabilities of dying between two ages, survival probabilities from birth to any age, and residual life expectancy. The best known among these functions is the life

[1] Only two of the countries included in the database pertain to the Caribbean region, Cuba and Dominican Republic. The remaining members of our sample belong to Central-North America (Costa Rica, El Salvador, Guatemala, Honduras, Mexico, Nicaragua and Panama) and South America (Argentina, Plurinational State of Bolivia (Bolivia), Brazil, Chile, Colombia, Ecuador, Paraguay, Peru, Uruguay and Bolivarian Republic of Venezuela (Venezuela)).

expectancy at birth, the work-horse of all respectable mortality analyses.

The condition on which computation of accurate life tables depends, accurate vital statistics and census counts, is satisfied only in a fairly recent period of the history of human populations. While it is true that past mortality trends in Western Europe and some parts of North America can be studied using genealogies and population reconstruction from parish records, these are not sources readily available in LAC countries nor can they shed light on characteristics of entire nation-states. As a consequence, reconstruction of the past begins sometime after routine data collection of population and vital records, a product of modern civilization, becomes established as part of each nation-state's bureaucratic agenda. Systematic collection of relatively accurate national mortality statistics for national populations does not begin anywhere before 1750 and does not make its appearance in the LAC region until well into the Twentieth Century, too late to offer the requisite material to trace with some accuracy the history of mortality changes over a period longer than half a century. Thus, estimates of life tables for all countries in LAC and for an extended stretch of time depend on deployment of a broad array of indirect techniques, applied to different periods, with country specific variants and, oftentimes, requiring separate treatment of different age groups. A key feature of LAMBdA is that estimates of life tables are *consistent* during the period of interest and across all countries that contribute to the mortality history of the region.

This chapter begins in Section 3.2 with a description of data sources. A discussion of data flaws is in Section 3.3 whereas Section 3.4 describes methods for detection and adjustment of errors and estimation of adult life tables for the period after 1950. In this section there is an extended description of the design and results of a simulation study that identifies optimal procedures to construct adjusted life tables when suitable yearly vital statistics and population censuses are available. Section 3.5 reviews methods employed to estimate adult life tables during the period 1850-1950, when both population censuses and vital statistics are more sparse. Finally, Section 3.6 is a summary of methods to estimate infant and childhood mortality (a more detailed review of these methods is in Chapter 4 of the documentation).

## 3.2 Data sources: 1850-2010

The key ingredients for a life table are mortality rates, the ratios of deaths in one age group to the population exposed to the risk of death in that age group. An accurate death rate requires both accurate count of deaths (numerator) and of population (denominator) by age groups. Seldom are these found in the LAC region, even in the most recent period. In most cases we estimate *adjusted* life tables for the years following 1950 from two external, independent sources, population censuses and vital statistics. Estimates of mortality at ages older than 5 during 1850-1950 are derived from a combination of methods that rely on one or two censuses and vital statistics or two censuses, adjusted intercensal rates of population growth, and a carefully chosen combination of age mortality patterns. Child mortality for the same period is estimated using a blend of independently obtained adjustment factors, raw vital statistics, models of mortality, and estimated time series of adjustment factors.

### 3.2.1 Population censuses

Most LAC countries began collecting population censuses on a regular basis after 1945-50 although a handful carried out national censuses right after independence and beginning as early as 1825 (Ecuador) and Colombia (1827). In Cuba, a pioneer in the region, censuses conducted periodically, without gaping intervals, begin in 1841 as part of a centralized and systematic program of population assessments.

Census information is publicly available either directly from country national statistics offices or in summary briefs and computer files gathered by the Pan American Health Organization (PAHO), Center for Latin American Demography, CELADE, the Organization for Economic Development and Cooperation,OECD, the Population Division of the United Nations (UN), and the World Health Organization(WHO). Some of these data were made public periodically through broadly accessible sources, including multiple editions of the UN yearbook as well as in publications from PAHO. The most readily accessible census population figures are in 5-year age groups with separate information for the first two age groups, 0 and 1-4, and an open age group usually (but not always) defined at 85+, and are routinely tabulated by sex. For about one third of LAMBdA countries there is data by single year of ages up to age 100 accessible through computerized data bases starting as early as 1970 (see Tables 3.1 and 3.2). But these are the exception, not the rule.

While many countries in LAC have undertaken national population censuses, these have not always been periodic, seldom attain full population coverage prior to 1950, and the information content (age reporting) is of unequal quality. Of particular importance is the completeness of census counts, that is, the fraction of the existing total population in each age group actually identified by the census. Except for recent periods and even then irregularly, post-enumeration surveys that assess census coverage are unavailable.[2] As a consequence, assessment of the quality and degree of coverage of censuses in the region can only be undertaken by identifying scattered ancillary studies, frequently of local rather than national scope, designed to evaluate census plans, field work protocols, and ex-post judgments of the quality of the information collected. When available, we use these studies as secondary sources of quality control for estimates. However, the cornerstone of the approach used to build LAMBdA consists of evaluating the relative quality of death rates themselves rather than on making inferences about and adjusting separately for coverage and precision of population and death counts.

Table 3.1 displays a list of censuses included in LAMBdA for the period up to 2010.[3] Even if population censuses were of uniformly high quality, it is difficult to construct a continuous time series of mortality rates starting in 1850 since, with one or two exceptions, intercensal periods are irregular and oftentimes stretch over lengthy time intervals. Periodicity of census-based statistics is a recent trait, starting in most cases after 1950.

[2] Since 1960 a total of 26 post-enumeration surveys have been carried in 19 countries that fielded 101 population censuses or the equivalent of one post-enumeration survey for every 4 censuses (Borges and Sacco, 2016).

[3] In an Appendix of this chapter we describe the population census (or estimates and projections) used to construct life tables for the period 2010-20 (see Supplementary material on Chapter 10.1).

Table 3.1: Population censuses used in LAMBdA: 1850-2010.

| Country | Before 1900 | 1900-1949 | 1950 and after |
|---|---|---|---|
| Argentina | 1869, 1895 | 1914, 1947 | 1960, 1970, 1980, 1991, 2001, 2010 |
| Bolivia | | | 1950, 1976, 1982, 2001, 2012 |
| Brazil | 1872, 1890 | 1900, 1920, 1940 | 1950, 1960, 1970, 1980, 1991, 2000, 2010 |
| Chile | 1854, 1865, 1875, 1895 | 1907, 1920, 1930, 1940 | 1952, 1960, 1970, 1982, 1992, 2002, 2017 |
| Colombia | | 1905, 1912, 1918, 1928, 1938 | 1951, 1964, 1973, 1985, 1993, 2005, 2012, 2018 |
| Costa Rica | 1864, 1883, 1892 | 1927 | 1950, 1963, 1973, 1984, 2000, 2011 |
| Cuba | 1841, 1861, 1877, 1887 | 1907, 1919, 1931, 1943 | 1953, 1970, 1981, 2002, 2012 |
| Ecuador | | | 1950, 1962, 1974, 1982, 1990, 2001, 2010 |
| El Salvador | | 1930 | 1950, 1961, 1971, 1992, 2007 |
| Guatemala | 1880, 1893 | 1921, 1940 | 1950, 1964, 1973, 1981, 1994, 2002, 2011, 2018 |
| Honduras | | 1930, 1935, 1940, 1945 | 1950, 1961, 1974, 1988, 2001 |
| Mexico | 1895, 1900 | 1910, 1921, 1930, 1940 | 1950, 1960, 1970, 1980, 1990, 2000, 2010 |
| Nicaragua | | 1940 | 1950, 1963, 1971, 1995, 2005 |
| Panama | | 1911, 1920, 1930, 1940 | 1950, 1960, 1970, 1980, 1990, 2000, 2010 |
| Paraguay | | | 1950, 1962, 1972, 1982, 1992, 2002, 2012 |
| Peru | 1876 | 1940 | 1961, 1972, 1981, 1993, 2007, 2017 |
| Dominican Republic | | 1920, 1935 | 1950, 1960, 1970, 1981, 1993, 2002, 2010 |
| Uruguay | 1900 | 1908 | 1963, 1975, 1985, 1996, 2004, 2011 |
| Venezuela | | 1926, 1936, 1941 | 1950, 1961, 1971, 1981, 1990, 2001, 2011 |

### 3.2.2 Vital statistics and death counts

As early as 1880 and as part of a massive process of nation-building and state formation, LAC countries began establishing official, nationwide vital statistics systems to collect and organize information on deaths, births, and marriages. As is the case for population figures, death counts are generally available by year of death, grouped by gender and in 5 or 10-year age intervals, and include separate counts of deaths to infants (age 0) and young children (aged between 1 and 4) as well as those occurring to population older than 70, 75, 80 or 85. In most cases, death counts by age and sex are available yearly after 1950 and more erratically, if at all, during 1920-1950. When available, mortality statistics during the earlier periods are highly scattered but, in some cases, they are nicely centered around census years.

Information about death counts is as or more fragile than information about population counts. Up until recently most countries of the region either did not have established vital statistics systems in place or had one that recorded partially, imperfectly, and selectively the occurrence of vital events. Although there are other deficiencies, such as inconsistency between the timing of deaths and the timing of registration and age misreporting, the most important one is the lack of complete death registration. Before 1950 virtually no country in the region issued statistics that were anywhere near complete. After 1950 there has been considerable improvement but there are still laggards with deficient or inexistent vital statistics systems in place (Bolivia), and more than a handful of countries whose official statistics are irregularly produced and continue to be of questionable quality.

Table 3.2 summarizes the availability of death statistics in the LAC region up to 2010.[4] The table includes country-years for which the pertinent information on deaths is broken

[4]In the Appendix to this chapter we describe sources used to construct life tables for the period 2010-2019 (see Supplementary material on Chapter 10.1).

Table 3.2: Vital statistics used in LAMBdA: 1880-2010.

| Country | Years |
|---|---|
| Argentina | 1912-1915, 1944-1970, 1977-2017 |
| Bolivia | |
| Brazil | 1974-2017 |
| Chile | 1936-2017 |
| Colombia | 1926-1928, 1936-1971, 1973-1975, 1982-2018 |
| Costa Rica | 1940, 1950-2018 |
| Cuba | 1927-1936, 1959-2017 |
| Dominican Republic | 1937-1976, 1979, 1982-2018 |
| Ecuador | 1957-2017 |
| El Salvador | 1933-2014 |
| Guatemala | 1933-2017 |
| Honduras | 1933-1943, 1947-1971, 1973-1990 |
| Mexico | 1930-2017 |
| Nicaragua | 1933-1946, 1948-2017 |
| Panama | 1941-1943, 1948-2017 |
| Paraguay | 1936-1944, 1948, 1950-2018 |
| Peru | 1939-2017 |
| Uruguay | 1905-1907, 1909-1921, 1923, 1929-2017 |
| Venezuela | 1933-1945, 1947-2017 |

down by gender and is available in age categories not coarser than ten year groups.

## 3.3 Data flaws

In what follows we describe two problems that complicate the creation of a continuous, complete, and accurate time series of mortality estimates for the countries in the region. These affect the depth and accuracy of the information.

### 3.3.1 Depth: time, content, and resolution

The raw data on which LAMBdA is built contains three *depth-related* weaknesses. The first consists of shortcomings affecting *time depth* or the density of estimates per year during a period of time. Characterization of mortality levels and patterns before 1950 is mostly based on sparse vital records and population censuses that, with a few exceptions, take place at irregular and/or excessively long intervals of time. Because it is not always possible to generate a set of continuous estimates, LAMBdA's life tables document levels and patterns of mortality centered in census years and/or cover an intercensal period not exceeding 15 years. As a result of this choice, the density of estimates for the period before 1950 is on average between 0.38 and 0.89 life tables per country decade. Conditions improve substantially after 1950 when it becomes possible to compute life tables on a yearly basis if one is willing to rely on interpolated population counts. For the most part, the description we provide in this documentation is based on life tables estimated for years centered on a population census as well as on the middle of intercensal periods not exceeding 15 years. On the whole, the density of estimates for the most recent period increases to about 0.98 life tables per country-year.

Table 3.3 identifies country-years with estimated pivotal life tables.

The second weakness is one of *content depth*: we do not aspire nor are able to characterize

Table 3.3: Pivotal (adjusted) life tables centered on intercensal periods.

| Country | Before 1900 | 1900-1949 | 1950 and after |
|---|---|---|---|
| Argentina | 1882 | 1904, 1914, 1930 | 1953, 1965, 1975, 1985, 1996, 2005 |
| Bolivia | | 1925 | 1963, 1984, 1996, 2006 |
| Brazil | 1881, 1895 | 1910, 1930, 1945 | 1955, 1965, 1975, 1985, 1995, 2005 |
| Chile | 1859, 1870, 1880, 1890 | 1901, 1913, 1925, 1935, 1946 | 1956, 1965, 1976, 1987, 1997, 2006 |
| Colombia | | 1908, 1915, 1923, 1938, 1944 | 1957, 1968, 1979, 1989, 1999, 2008 |
| Costa Rica | 1873, 1887 | 1909, 1927, 1938 | 1956, 1968, 1978, 1992, 2005 |
| Cuba | 1851, 1869, 1882, 1893 | 1903, 1907, 1913, 1925, 1937, 1948 | 1961, 1975, 1991, 2006 |
| Dominican Republic | | 1927, 1942 | 1955, 1965, 1975, 1987, 1997, 2006 |
| Ecuador | | | 1956, 1968, 1978, 1986, 1995, 2005 |
| El Salvador | | 1940 | 1955, 1966, 1981, 1999, 2008 |
| Guatemala | 1886 | 1907, 1930, 1945 | 1957, 1968, 1977, 1987, 1998, 2005 |
| Honduras | | 1932, 1937, 1942, 1947 | 1955, 1961, 1967, 1974, 1981, 1988, 1994, 2001 |
| Mexico | 1897 | 1905, 1915, 1921, 1925, 1935, 1945 | 1955, 1965, 1975, 1985, 1995, 2005 |
| Nicaragua | | 1945 | 1956, 1967, 1983, 2000, 2007 |
| Panama | | 1915, 1925, 1935, 1945 | 1955, 1965, 1975, 1985, 1995, 2005 |
| Paraguay | | | 1956, 1967, 1977, 1987, 1997, 2007 |
| Peru | | 1908 | 1950, 1966, 1976, 1987, 2000, 2008 |
| Uruguay | | 1904, 1908, 1935 | 1969, 1980, 1990, 2000, 2007 |
| Venezuela | | 1931, 1938, 1945 | 1955, 1966, 1976, 1985, 1995, 2006 |

mortality at levels of aggregation lower than the nation-state. LAMBdA is a database to document national trends and is silent on regional or social class heterogeneity that may emerge with or be shaped by national trends.

The third weakness is one of *resolution depth*: the estimates in LAMBdA are oftentimes based on raw information aggregated in five year age groups and contain a category for unknown ages. The information on population and death counts is in five year age groups, starting at age 5 and ending in an open group frequently, but not always, at age 85. In the bulk of country-years of interest, the population younger than 5 is in two age groups, 0 and 1-4. However, to fully apply some of the adjustment procedures required to counterbalance errors of coverage and age statement, we generated more detail in the form of mortality rates by single years of age. Whenever the requisite data was not available, this information was obtained using different strategies. First, for the most recent period, after 1970 or so, we utilize data originally released (but not always published) in single years of age. This data requires simple corrections to minimize the impact of age heaping. Second, when the original single-year of age data were unavailable we rely on the application of standard cubic interpolation procedures (Sprague multipliers) to break five year age groups into their single year of age components. We do this only for ages older than 5 and younger than 85. Although alternative methodologies could have been used, we show elsewhere that the results would not have been significantly different (see Chapter 5 of this documentation) from those we settled on.

A different dimension of the *resolution depth* problem is the magnitude of the fraction of population and death counts of age unknown, a quantity that varies by country time-periods and becomes gradually smaller in the last two decades of the Twentieth Century. Although we tested a number of methods for redistributing unknown age counts, we finally settled for the simplest one, namely a redistribution according to the observed age distribution of counts (both deaths and populations). Other methods, based on the use of stable populations and/or approximations to known population parameters, yield results that are difficult to distinguish from those of the simple procedure but are considerably more costly and cumbersome to apply.

### 3.3.2 Accuracy: relative completeness and age reporting

The most important limitation of the raw data on which LAMBdA depends is defective coverage and age misreporting. By and large, observed death counts are a fraction of the 'true' number of deaths that take place at a particular time as they exclude events that, for a number of reasons, are never recorded. Deficiencies are worst at very young and old ages but frequently also affect population in the labor force and differentially so by gender. It is only when national vital registration systems operate efficiently and have a truly national reach, as they do in the most recent periods, that deficiencies in death counts are mostly confined to issues of consistency between timing of occurrence and recording of events.

Since population censuses are also frequently affected by coverage problems, mortality rates computed with the raw data may contain smaller *net errors* that would be expected otherwise. In general, however, the observed mortality rates underestimate mortality levels, particularly at very young and old ages. Throughout, we will refer to this as the *completeness or relative completeness problem*. We use the term *relative completeness factors* when we speak of ratios of true or error-free to observed mortality rates. Table 3.4 is a quick summary of the nature of the problem: it displays estimates of relative completeness of adult (over 5 years of age), infant (age 0) and early child (ages 1-4) death registration for a sample of LAC countries in two different periods of time. The figures in this table suggest that the quality of the information is poorer at very young ages and that, although there is a clear universal trend toward improvement, a fraction of countries still show signs of deficient registration even during the most recent periods.

Defective relative completeness is not, however, the only or even the worst problem. The accuracy of both census and death counts can be threatened by age misreporting in either source. First, age heaping is a well-known problem of population counts and can be repaired, albeit imperfectly, using simple techniques for identifying preferred digits and then redistributing population counts to follow a smoother trajectory. In most cases, these simple adjustments suffice to produce accurate summary measures and age patterns of mortality but not always with much precision at high levels of resolution.

Our general strategy for adjustments is to start out with population and death counts in five year age groups beginning at age 5 and fit polynomials (Sprague) to obtain population (death) counts by single years of age between ages 5 and 84. We then calculate single age specific accumulated counts of deaths and populations, the quantities needed to estimate adjustment factors. These quantities are minimally affected by age heaping and provide a sound empirical basis to compute the adjustment factors. In most cases the resulting rates do not contain the typical footprints of age heaping. In the few cases they do, we proceed to smooth the single age mortality rates applying a local smoother with no more than three contiguous age groups for support. Finally, we aggregate the adjusted data into five year

Table 3.4: Relative completeness of deaths registration in LAC countries: 1920-2010.

| Country | Period 1900-1949 | | | | Period 1950 + | |
|---|---|---|---|---|---|---|
| | Mid-Year | Age 0 | Age 1-4 | Age 5+ | Mid-Year | Age 5+ |
| Argentina | 1914 | 0.968 | 0.865 | 0.939 | 1953 | 0.974 |
| | | | | | 2005 | 0.995 |
| Bolivia | NA | NA | NA | NA | NA | NA |
| | | | | | NA | NA |
| Brazil | NA | NA | NA | NA | 1985 | 0.885 |
| | | | | | 2005 | 0.996 |
| Chile | 1925 | 0.867 | 0.829 | 0.852 | 1956 | 0.961 |
| | 1945 | 0.867 | 0.829 | 0.934 | 2006 | 0.980 |
| Colombia | 1944 | 0.821 | 0.815 | 0.749 | 1957 | 0.790 |
| | | | | | 2008 | 0.800 |
| Costa Rica | 1927 | 0.901 | 0.922 | 0.893 | 1956 | 0.918 |
| | 1938 | 0.901 | 0.922 | 0.893 | 2005 | 0.975 |
| Cuba | 1925 | 0.806 | 0.893 | 0.800 | 1961 | 0.890 |
| | 1948 | 0.806 | 0.893 | 0.870 | 2006 | 0.989 |
| Dominican Republic | 1942 | 0.476 | 0.451 | 0.487 | 1955 | 0.500 |
| | | | | | 2006 | 0.604 |
| Ecuador | NA | NA | NA | NA | 1956 | 0.738 |
| | | | | | 2005 | 0.805 |
| El Salvador | 1940 | 0.554 | 0.776 | 0.721 | 1955 | 0.700 |
| | | | | | 2008 | 0.714 |
| Guatemala | 1945 | 0.714 | 0.898 | 0.784 | 1957 | 0.888 |
| | | | | | 2005 | 0.940 |
| Honduras | 1942 | 0.542 | 0.551 | 0.495 | 1955 | 0.518 |
| | 1947 | 0.542 | 0.551 | 0.500 | 1989 | 0.750 |
| Mexico | 1925 | 0.843 | 0.822 | 0.752 | 1955 | 0.860 |
| | 1945 | 0.843 | 0.822 | 0.883 | 2005 | 0.959 |
| Nicaragua | 1945 | 0.526 | 0.545 | 0.498 | 1956 | 0.456 |
| | | | | | 2007 | 0.561 |
| Panama | 1945 | 0.837 | 0.757 | 0.829 | 1955 | 0.839 |
| | | | | | 2005 | 0.853 |
| Paraguay | NA | NA | NA | NA | 1956 | 0.601 |
| | | | | | 2006 | 0.681 |
| Peru | NA | NA | NA | NA | 1950 | 0.490 |
| | | | | | 2008 | 0.533 |
| Uruguay | 1908 | 0.844 | 0.822 | 0.879 | 1969 | 0.960 |
| | | | | | 2007 | 0.996 |
| Venezuela | 1938 | 0.833 | 0.857 | 0.846 | 1955 | 0.866 |
| | 1945 | 0.833 | 0.857 | 0.855 | 2006 | 0.895 |

age groups and compute abbreviated life tables.[5] In cases when the data are available in single years of ages we apply smoothing techniques to remove age heaping and then proceed as described before.

A second, more insidious,class of errors of age declaration is systematic over (under) reporting. As we show later, vital and census statistics in LAC countries are, almost without exception, affected by age overstatement, particularly at ages over 40 or 45. When the (true) age distribution of a population is roughly exponential in nature —as it always is in stable and quasi stable populations—systematic age overstatement of populations induces downward biases in mortality rates at older ages. Unfortunately, these biases are not quite fully offset when there is an equal propensity to overstate ages at death. The reason these two type of errors do not cancel each other out is that while both adult mortality rates and adult population age distributions are roughly exponential, one slopes upwards (mortality rates) whereas the other slopes downwards (population). Matters are made worse when, as is almost always the case, the rate of decrease of population with age (natural rate of increase in a stable population) is an order of magnitude lower than the rate of increase of adult mortality rates (rate of senescence or the Gompertz slope in Gompertz mortality regimes). The consequence is that unless the propensity to overestimate ages at death is much higher than the propensity to overestimate ages of population, observed mortality rates will be biased downwards. If left uncorrected, the resulting life tables will offer a misleading portrayal of the curvature of mortality at older ages, suggesting the existence of slower rates of senescence or heavy influence of selection due to changing frailty composition. As vital registration and census enumeration improve, the magnitude of these biases tends to decrease and the entire history of observed life tables will erroneously suggest inexistent trends in old age patterns of mortality and misleading relative deceleration of rates of mortality decline at older ages. Table 3.5 displays estimated biases in mortality rates at ages over 45 in a sample of country-years in LAMBdA and the corresponding errors in life expectancy at age 60.

In Section 3.4 below we describe and evaluate a battery of procedures to compute adjusted life tables that minimize errors due to imperfect relative completeness and defective age reporting. The section focuses on developments and applications to compute adjustments for observed adult mortality during 1950-2010, the period with the most complete information.[6]

[5]Because raw, five-year death and population counts, are ultimately allocated into single years of age, adjustment for heaping is redundant.

[6]A terminological clarification is in order. Although alternative definitions were considered, define as "adult" the population aged 5 and older and as "children" those younger than age 5. The weakness of this definition is that we lump together a youthful and a genuinely adult population with possibly distinct mortality experiences in one single category, the "adult population". There is no escape from this awkward labeling and classification. In fact, while it is possible and desirable to estimate mortality separately for the population younger than 5, between 5 and 19, and older than 20, there are no specially tailored procedures that we know of to adjust mortality estimates for the population aged 5-19.

Table 3.5: Estimated biases due to adult age overstatement: LAC 1950-2010.

| | | E(45) | | | E(60) | | |
|---|---|---|---|---|---|---|---|
| Country | Mid-year | Unadj (1) | Adj (2) | Difference (1) - (2) | Unadj (a) | Adj (b) | Difference (a) - (b) |
| Argentina | 1953 | 25.96 | 25.29 | 0.67 | 15.39 | 14.55 | 0.84 |
| | 2005 | 30.02 | 29.33 | 0.69 | 17.96 | 17.15 | 0.81 |
| Brazil | 1985 | 28.55 | 27.62 | 0.93 | 17.61 | 16.51 | 1.10 |
| | 2005 | 31.27 | 30.23 | 1.04 | 19.77 | 18.58 | 1.19 |
| Chile | 1956 | 24.44 | 23.72 | 0.72 | 14.57 | 13.64 | 0.93 |
| | 2006 | 33.20 | 32.16 | 1.04 | 20.45 | 19.33 | 1.12 |
| Colombia | 1957 | 27.34 | 26.46 | 0.88 | 16.68 | 15.67 | 1.01 |
| | 2008 | 35.09 | 33.86 | 1.23 | 22.29 | 20.96 | 1.33 |
| Costa Rica | 1956 | 29.08 | 28.10 | 0.98 | 17.55 | 16.46 | 1.09 |
| | 2005 | 34.96 | 33.78 | 1.18 | 22.40 | 21.13 | 1.27 |
| Cuba | 1961 | 30.13 | 29.18 | 0.95 | 18.15 | 17.08 | 1.07 |
| | 2006 | 33.46 | 32.56 | 0.90 | 20.94 | 19.95 | 0.99 |
| Dominican Republic | 1955 | 33.62 | 31.91 | 1.71 | 22.44 | 20.52 | 1.92 |
| | 2006 | 38.35 | 36.41 | 1.94 | 25.76 | 23.68 | 2.08 |
| Ecuador | 1956 | 28.75 | 27.77 | 0.98 | 17.98 | 16.83 | 1.15 |
| | 2005 | 37.42 | 35.94 | 1.48 | 25.23 | 23.62 | 1.61 |
| El Salvador | 1955 | 27.64 | 26.69 | 0.95 | 17.54 | 16.42 | 1.12 |
| | 2008 | 32.79 | 31.85 | 0.94 | 21.74 | 20.62 | 1.12 |
| Guatemala | 1957 | 24.44 | 23.68 | 0.76 | 15.06 | 14.07 | 0.99 |
| | 2005 | 31.39 | 30.42 | 0.97 | 20.22 | 19.10 | 1.12 |
| Honduras | 1955 | 30.55 | 29.14 | 1.41 | 20.37 | 18.64 | 1.73 |
| | 1989 | 37.33 | 35.61 | 1.72 | 25.06 | 23.17 | 1.89 |
| Mexico | 1955 | 26.57 | 25.80 | 0.77 | 16.69 | 15.71 | 0.98 |
| | 2005 | 33.04 | 31.97 | 1.07 | 21.13 | 19.95 | 1.18 |
| Nicaragua | 1956 | 32.09 | 30.61 | 1.48 | 21.05 | 19.37 | 1.68 |
| | 2007 | 36.23 | 34.71 | 1.52 | 24.05 | 22.41 | 1.64 |
| Panama | 1955 | 28.93 | 27.87 | 1.06 | 17.67 | 16.45 | 1.22 |
| | 2005 | 35.92 | 34.65 | 1.27 | 23.18 | 21.81 | 1.37 |
| Paraguay | 1956 | 32.97 | 31.73 | 1.24 | 20.81 | 19.44 | 1.37 |
| | 2006 | 34.84 | 33.60 | 1.24 | 22.17 | 20.84 | 1.33 |
| Peru | 1950 | 30.61 | 29.47 | 1.14 | 20.64 | 19.25 | 1.39 |
| | 2008 | 39.37 | 37.66 | 1.71 | 26.32 | 24.52 | 1.80 |
| Uruguay | 1969 | 26.72 | 26.11 | 0.61 | 15.47 | 14.69 | 0.78 |
| | 2007 | 30.35 | 29.85 | 0.50 | 18.17 | 17.57 | 0.60 |
| Venezuela | 1955 | 27.49 | 26.47 | 1.02 | 16.81 | 15.64 | 1.17 |
| | 2006 | 32.75 | 31.53 | 1.22 | 20.94 | 19.59 | 1.35 |

Unadj: unadjusted for age misreporting; Adj: Adjusted for age misreporting

## 3.4 Adjustments of adult mortality for the period 1950-2010

As should be clear from the above description, the nature of problems faced is highly heterogeneous: they vary by country, time period, age groups and, lastly, by gender. This state of affairs is complicated by the fact that there are multiple procedures, each relying on specialized assumptions, to adjust for errors in the data. To make these deficiencies tractable we proceed in three stages. In the first stage we develop an evaluation study designed to identify 'optimal' adjustment procedures for relative completeness and age misreporting. In the second stage we assess the performance of methods to adjust for relative completeness of death registration. The third stage describes a new procedure to adjust for a older age misreporting Finally, we formulate a precise integrated procedure to simultaneously adjust for both defective completeness and age misreporting. The organization of the rest of the chapter is as follows: Section 3.4.1 contains the details of the simulation and evaluation study, section 3.4.2 identifies and evaluates techniques to correct for defective population and death counts. Section 3.4.3 describes a procedure to adjust for age misreporting and Section 3.4.4 lays out an integrated adjustment procedure. Finally, Section 3.4.5 is an empirical illustration of a life table construction.

### 3.4.1 Stage I: Nature of the evaluation study

This section describes an evaluation study designed to assess the performance and establish a ranking of alternative methods to corrects for errors due to under (over) counting of population and deaths and age misreporting after 1950.[7] After 1950 all LAC countries (except Bolivia) begin to release periodic vital statistics data and official population counts. As documented in Table 3.3 we compute approximately 120 *pivotal* life tables for years comprised between 1950 and 2019.

Over the last two to three decades, but mostly in the late seventies and eighties, demographers developed a large number of techniques to adjust faulty data from censuses, vital statistics and population surveys to estimate both fertility and mortality. There are nearly 15 different, albeit not completely independent, methods to estimate relative completeness of death registration and to adjust adult mortality and associated life tables. Each of these methods has its advantages and shortcomings and they all depend on sets of non-identical but overlapping assumptions. The work by Hill and Choi partially evaluated the performance of a subset of these methods (Hill et al., 2009; Hill and Choi, 2004; Hill et al., 2005). We extend this work, as well as our previous research on the subject (Palloni and Pinto, 2004), and design a very general simulation study to evaluate a set of methods to adjust for relative completeness and age misreporting. The purpose of the simulation study is to support the choice of optimal adjustments to compute LAMBdA life tables with defective data.

Less investigated is a second problem facing the estimation of age patterns of mortal-

[7]The investigations that follow were first documented elsewhere (Palloni and Pinto, 2004).

ity, namely, age misreporting.[8] It is well-known that census and death counts by age are influenced by digit preference ('heaping') and biases due to propensity to increase (decrease) the true age. Although problematic, age heaping can be repaired because in most cases it is possible to approximately restore the original age distribution in the neighborhood of preferred digits using computations that rely on relatively safe assumptions. Systematic age misstatement is altogether different since it is harder to diagnose and its treatment requires additional knowledge of at least two functions: (a) the conditional (on age and gender) propensity of individuals to exaggerate (decrease) the true age and (b) the conditional (on age and gender) distribution of differences between the correct and declared age. To solve this problem we propose generalizations of an existing procedure to identify the presence of age misstatement, formulate a new method to estimate functions representing (a) and (b) from observable quantities, and define an integrated algorithm to adjust observed adult mortality rates for both faulty coverage and systematic age misreporting.

Neither adjustment for faulty coverage nor detection and correction of biases due to age misreporting are feasible in the absence of well-established criteria to decide which of the many (for coverage) or fewer (for age misreporting) candidate methods performs optimally. To fill this gap we carry out a systematic evaluation study of the performance of extant methods using a range of simulated conditions similar to those experienced in countries under study.

The main objective of the evaluation study is to assess the sensitivity of alternative techniques to violations of assumptions on which they are based, particularly those that are most likely to misrepresent historical conditions. To do this we first simulate populations representing different demographic profiles (stable, quasi-stable and non-stable) driven by combinations of (a) constant fertility and mortality, (b) constant fertility and declining mortality, and (c) declining fertility and declining mortality. We then combine these population profiles with different patterns of distortions due to faulty coverage of population and death counts as well as of age misreporting. A battery of techniques is deployed and in each case we compute multiple measures of performance comparing the true (target) population parameter(s) with those retrieved by each technique. We rank the performance of techniques for each combination of conditions violating assumptions on which the techniques rely. Finally, we score techniques according to their sensitivity to violation of combinations of assumptions. We then choose an optimal technique which is paired with a new procedure to adjust for age misreporting. These are then jointly used in an integrated algorithm to make final adjustments to observed adult mortality rates. A crucial issue discussed at length below is the order in which these techniques, one for adjustment of coverage of events and population counts and the other for age misreporting, must be deployed and the justification for that order.

The next subsections describe simulated population regimes, errors of coverage, and patterns of age misreporting.

[8]The simulations by Hill and colleagues also include simple forms of age misreporting. We augment and generalize this aspect to capture patterns of age misreporting that could be typical of LAC countries.

**Simulated populations**

The simulated populations depend on three sets of functions.[9] The first are colletions of demographic parameters that uniquely identify age distribution of deaths and populations, *conditional of age patterns of mortality and fertility.* The second identify the age patterns of mortality and fertility. Finally, the third set of functions define the distortions of counts and age distributions of populations and deaths. We discuss these in turn.

**Demographic parameters, initial populations and population trajectories.** Five master (female) populations were created, one stable and four non-stable populations, that represent trajectories followed during a 100 year period, from 1900 to 2000. The stable population has a $GRR = 3.03$ and $E(0) = 45$ with a natural rate of increase $r = 0.025$. This model stable population roughly corresponds to the average of LAC populations in the period 1940-60, e.g. not yet heavily perturbed by large scale net migration, as is the case in Argentina, Brazil, Cuba, and Uruguay, or fertility changes, as in the cases of Argentina and Uruguay.

We also include four non-stable populations profiles that follow (approximately) the mortality and fertility schedules for Costa Rica, Mexico, Guatemala, Argentina, and Uruguay during the period 1900-2000. The initial stable distribution for the first three non stable populations are set to be equal to the stable populations with parameters $r$ and $E(0)$ equal to those estimated around 1900 in the corresponding countries (Costa Rica, Mexico and Guatemala). In contrast, the initial age distributions corresponding to the fourth non stable profile (Argentina and Uruguay) are set equal to the observed average age distribution in population censuses within the period 1850-1900. Thus, the initial populations (and deaths) distributions roughly correspond to the actual initial starting populations in most of the LAC region.

**Age patterns of fertility and mortality.** The calculation of yearly and age specific counts of populations and deaths during the 1900-2000 periods follows standard population projection techniques and demands specification of patterns of fertility and mortality. For mortality we chose the West and South models in the Coale-Demeny family of life tables. For fertility, we adopt a unique age pattern of fertility identical to the one used in the computations of the Coale-Demeny stable population models (Coale et al., 1983). We assume throughout that each type of demographic transition in the non stable populations preserves the age patterns of mortality and fertility. Since all calculations are in single years age group both the Coale-Demeny life tables and fertility patterns were transformed into single years schedules of mortality and fertility, respectively. The transformation of the life tables functions into single years functions was carried out by strictly adhering to separator factors adopted by Coale and Demeny. The single-year fertility functions was derived using splines.

In summary, we create 10 stable and non-stable populations (five masters for the West and five masters for the South mortality models) that span a 100 year period from 1900 to

[9] The simulation and evaluation described below is an extension of a study described in Palloni and Pinto (2004). It is different from another evaluation study by Hill and colleagues (Hill et al., 2009) in that the simulated populations include consideration of systematic age misreporting that follow the standard pattern of age misreporting.

2000 and represent a very broad set of experiences, from those preserving population stability throughout, to those that remain stable up until 1950 or thereabouts, to those shifting to quasi-stability from 1930 up to 1980 and, finally, to those with little or no stability at all from the outset.[10]

**Simulated distortions I: defective census and vital statistics coverage**

Distortions due to population or death coverage were implemented in a straightforward matter. We define observed population (or death) counts by age as a fraction of the simulated (true) quantities:

$$\begin{aligned} P^o_{x,t_1} &= C_1 P^T_{x,t_1} \\ P^o_{x,t_2} &= C_2 P^T_{x,t_2};\ t_2 < t_1 \\ D^o_{x,t} &= C_3 D^T_{x,t};\ t = t_1, t_1+1, \cdots \leq t_2 \end{aligned}$$

for $x \geq 5$, where $P^o_{x,t_1}$ is the observed (distorted) population in the age group $(x, x+1]$ at time $t_1$, $P^o_{x,t_2}$ is the observed (distorted) population in the age group $(x, x+1]$ at time $t_2$, and $D^o_{x,t}$ is the observed (distorted) number of deaths in the age group $(x, x+1]$ and year $t$; $P^T_{x,t_1}$, $P^T_{x,t_2}$ and $D^T_{x,t}$ are the true (simulated) quantities and $C_1$, $C_2$ and $C_3$ are the fractions of total death counts actually observed (completeness factors). In the simulation the completeness factors for censuses were set at values in the range 0.80-1.0 in intervals of 0.5 whereas the death completeness factors vary between 0.70 and 1.0 in intervals of 0.5. Thus, altogether we produce a total of 175 patterns of distortions and each of these was combined with the 10 different population profiles producing a total of 1,750 observed demographic profiles. These profiles are sufficient to evaluate adjustment methods that require only one census and one to three years of deaths centered on the census or, alternatively, those that demand as inputs two population censuses and the yearly counts of intercensal deaths.[11]

The distortions defined above contains a strong assumption, namely, that completeness of both population and death counts is age invariant. At least in the age range within which the techniques are deployed (5-85), the assumption is unlikely to be met, particularly for population counts. To increase the representation of distortions we added two different patterns of age varying completeness of population and death counts and generated a total of 20 additional simulated populations (e.g. two for each of the 10 profiles originally defined). We show later, however, that as long as the difference between maximum and minimum completeness stays below 10 percent of the mean value of completeness, variability of completeness by age does not have a strong impact on our preferred strategy (Section 3.4.2).

[10]Full information on the five population profiles is in Section 2 of Chapter 10. Codes for the computation of the simulated populations are in https://gitlab.com/csic-echo/lambda-pop.

[11]When there is no ambiguity, we will always use the shorthand symbols $C_1$, $C_2$, and $C_3$) to refer to completeness factors. However, when more precision is needed to refer to a particular period of time, we will use $C_{t_1}$ , $C_{t_2}$ and $CD_{[t_1,t_2]}$ to make explicit the fact that we are referring to an intercensal period $[t_1, t_2]$ and that the completeness of death factor, $C_3 = CD_{[t_1,t_2]}$ refers to deaths enumerated throughout the period.

**Simulated distortions II: systematic age misreporting**

To describe the model of age misreporting we begin with a few basic definitions. Let $\theta^o_y$ be the average conditional probability that individuals aged $y$ overstate their age in a census and $\theta^u_y$ the conditional probability of understating their age. Then $(1-\theta^o_y-\theta^u_y)$ is the probability of an accurate age statement. Individuals who over(under) state their age do so by choosing, not always randomly, the age declared and observed in the census. This age could be $n>0$ years removed from the true age. As we show below, it suffices to let $n$ range between 1 and 10 since the actual frequency of distortions exceeding 10 years are exceedingly small, e.g. individuals rarely over(understate) their age by more that ten digits. Let $\rho^o_y(n)$ be the average conditional probability that individuals aged $y$ who overstate ages will do so by $n$ years with an analogous definition for the probabilities for age understatement, $\rho^u_y(n)$ and with $\sum_n \rho^u_y(n) = \sum_n \rho^o_y(n) = 1$. To compute the observed population at age $y$, $\mathrm{P}^o_y$, we consider the true (simulated) number at that age, $P^T_y$, and apply the conditional probabilities defined above:

$$P^o_y = P^T_y(1-\theta^o_y-\theta^u_y) + \sum_{n=1}^{n=10} P^T_{y-n}\rho^o_{y-n}\theta^o_{y-n} + \sum_{n=1}^{n=10} P^T_{y+n}\rho^u_{y+n}(n)\theta^u_{y+n}\,. \tag{3.4.1}$$

This expression can be generalized for all ages between 0 and 100 in compact matrix notation:

$$\Pi^o = \Theta\Pi^T \tag{3.4.2}$$

where $\Pi^o$ is the (101x1) observed population vector, $\Pi^T$ is the (101x1) true population vector and $\Theta$ is a 101x101 square matrix of transition probabilities, e.g. the probabilities of “migration” into or out of single year age-groups. In particular, the diagonal of $\Theta$ contains the probabilities of correctly declaring ages, $(1-\theta^o_y-\theta^u_y)$.[12], namely, $(1-\theta^o_y-\theta^u_y)$. Entries in the off-diagonal cells pertaining to the $yth$ row and columns $y+1, y+2, \ldots, y+10$ are the quantities $\rho^o_y(y)(1)\theta^o_y, \ldots, \rho^o_y(10)\theta^o_y$. Entries in the off-diagonal cells pertaining to the $yth$ row and columns $y-1, y-2, \ldots, j-10$ are the quantities $\rho^u_y(1)\theta^u_{y+1}, \ldots, \rho^u_y(10)\theta^u_y$. It is possible to retrieve the vector of true population counts population by pre-multiplying the previous expression by the inverse of $\Theta^{-1}$, that is

$$\Theta^{-1}\Pi^o = \Pi^T \tag{3.4.3}$$

However, this operation requires full knowledge of the matrix $\Theta$. As we show below, demographers have only superficial information about the nature of this matrix in LAC countries or anywhere else (but see Bhat (1990)). In the absence of precise knowledge of $\Theta$ one could adopt shortcuts that, as shown below, lead to identification problems that impeded specification of an invertible matrix of transition probabilities.

What do we know about age misreporting in population and death counts in LAC and in other countries? There is an extensive literature on general errors in age reporting

[12]Unless explicitly stated, we will assume a single year age distribution that includes ages $0, 1, 2, \ldots, 100+$, where the last age interval is the open age group.

(Ewbank, 1981; Chidambaram and Sathar, 1984; Kamps E., 1976; Nuñez, 1984) as well as on systematic adult age misstatement in population counts. Most of these uncover evidence of overstatement in low income countries (Mazess and Forman, 1979; Grushka, 1996; Bhat, 1987, 1990; Del Popolo, 2000; Dechter and Preston, 1991) or in US migrant groups (Hispanic or Hispanic origins)(Rosenwaike and Preston, 1984; Spencer, 1984). There is also body of literature that identifies patterns of age overstatemet in high income countries (Horiuchi and Coale, 1985; Coale and Kisker, 1986; Condran et al., 1991; Preston et al., 2003; Elo and Preston, 1994). In the US, for example, age overstatement is one of the factors that could explain the so called Black-White mortality crossover, whereby African American mortality rates dip below those of their White counterparts at very old ages (over 70). And while the conjecture of selection due to frailty has not been completely discarded, some investigations suggest that higher levels of overstatement of ages in the population (and deaths) among African American than among Whites accounts for a substantial part, but not all, of the mortality crossover (Elo and Preston, 1994). The Black-White mortality crossover is just an extreme example of the damage that age misreporting can inflict on estimates of adult mortality. As others before us have already done(Dechter and Preston, 1991; Grushka, 1996; Bhat, 1987, 1990), we will show that age overstatement is an important source of error in LAC.

Partial information on the matrix $\mathit{\Theta}$ has been obtained mostly from studies involving record linkages (Elo and Preston, 1994; Preston et al., 1996; Rosenwaike and Preston, 1984; Rosenwaike, 1987), post enumeration surveys (Ortega and Garcia, 1985) and comparisons of two independently gathered data sources that should produce the same outcomes (Bhat, 1990). In all these studies, however, the information is either aggregated in five-year age groups or applies to populations with levels of education that are much higher than those in LAC countries. Lack of age detail is problematic since computation of conditional probabilities in coarse age groups rests on approximations that, if violated, are generally harmful to the accuracy of estimates.

Using a transition matrix appropriate for a population with higher or lower levels of literacy than the target one may lead to distortions because age misreporting is associated with a population's literacy level. In the section that follows we propose an estimate of $\Theta$ suitable for LAC populations.

To estimate $\Theta$ we rely on an unusual study launched in 2002 by the Central American Center for Population at the University of Costa Rica. This study was designed to assess the quality of information of death registration and the accuracy of the 2000 census counts for an older adult population.[13] One of the components of this study was a linkage of a sample of individual census records with national voting registers, a database that contains information from birth certificates. A stratified sample of census records consisting of 7,426 individuals aged 55 and older in the census were matched to the voting registers. This represents about 0.81 of the total (observed) sample of census records originally sampled and .87 of the non-foreign born part of the sample. The final data set contains individuals

[13] We are grateful to Drs. Gilbert Brenes and Luis Rosero Bixby from the Central American Population Center at the University of Costa Rica for having provided tabulations and estimated models we use here.

classified by gender, education and other traits, and by 'true' and declared age.

To estimate the entries of matrix $\Theta$ we proceed in two steps

*i.* *Estimation of probabilities of age over and understatement, $\theta_x^o(V)$ and $\theta_x^u(V)$ where $V$ is a vector of individual characteristics, including age*:

We first estimate a logistic model to predict age misreporting, an event that affected a total of 2,894 individuals (40 percent) of whom 1,992 overreported and 902 underreported. We create a 1/0 binary variable whose value is set to 1 when there is either over or under statement and zero otherwise. To be useful in general applications, the vector $V$ only includes covariates universally available in a population census, namely, age and sex. Since the effects of sex and quadratic age had statistically insignificant effects, the final model we adopt includes true ages as the only predictor.[14] Because by design the sample only includes individuals aged 55 and above, it excludes individuals who reported ages younger than 55. As a consequence, the probabilities of age misreporting and, in particular, age understatement, will be under estimated if the true age falls (approximately) in the neighborhood of ages 55 to 59. To minimize the size of this bias we estimate models using a sample restricted to those who are 60 and older. This reduces the effective sample size from 7,246 to 6,290 of whom 1,786 overreported and 789 under reported. Table 3.6 displays estimated parameters for over and under stating ages using the weighted sample.[15]

*ii.* *Estimation of conditional probabilities of over(under) stating ages by $1 < n \leq 10$ years, $\rho_x^o(j)$ and $\rho_x^u(j)$*: We estimate a multinomial model with 9 categories that includes gender and (true) continuous age as independent variable. The resulting estimates reveal that the effects of gender are always statistically insignificant, that those of age show no clear pattern and, in addition, that their magnitude is quite small in 6 out of 8 cases for overstatement models and in 5 out of 8 contrasts for age understatement. To exploit these findings and to simplify representation of the data we estimate a null model predicting the average conditional probabilities of exaggerating (or diminishing) by $n$ years applicable to all ages older than 45 and both genders. The values of the predicted probabilities of over and understating the true age are in Table 3.7.

*iii* *Extensions*: The quantities estimated above reflect mostly errors in the population 60 and above, partially among those older than 55, and more marginally among those aged 45-49. Although empirical evidence for LAC and other populations suggests that systematic age misreporting (but not age heaping) becomes significant at ages over 55-59, we will include probabilities of misreporting (over and understatement) for ages 45-59. We estimate these using predicted values from the logistic model. This extrapolation is justified on the grounds that the fit of the model is very good and the conditional distribution of years of age misreporting is age invariant.

[14] Note that the model captures systematic age under or overstatement as well as over and understatement associated with age heaping.

[15] We also exclude foreign nationals and ambiguous records.

Although it is now possible to compute an estimator of the target mobility matrix, $\hat{\Theta}$, there is a problem of identification that cannot be resolved without additional simplifications. Suppose, for example, we seek to estimate mortality trends in a country with much lower levels of education than in Costa Rica. Replacing $\hat{\Theta}$ for the true matrix in (3.4.2), we will obtain a true distribution of ages but only under the very strong assumption that age misstatement is identical across countries. This contradicts the idea that the severity of age misstatement increases as levels of education drop or, more generally, that age misstatement is not uniform within a given population. A less constraining assumption is to argue that while the *age pattern* of age misstatement is approximately the same across subpopulations or countries, the levels (intensity) may differ. To express this idea one could shift the conditional probabilities of over and under stating ages (or a monotonic transform of these) by some constant value, say $\phi^o$ and $\phi^u$ for over and understatement respectively. While this is a reasonable strategy it generates an additional problem, namely, that a unique solution for equation (3.4.2) may no longer exist since different combinations of $\phi^o$ and $\phi^u$ embedded in the transition matrix could plausibly yield identical results. To circumvent this new difficulty we propose to use a standard pattern of *probabilities of net age overstatement* as $\varphi_x^S = \theta_x^o - \theta_x^u$ and then apply to it the conditional probabilities of overstating one's age by $n$ years (the $\rho_x^o(j)$ values defined before). Under these conditions the off-diagonal cells of the matrix defined by $\varphi_x^S$, $\hat{\Theta}^S$, simplify as all entries associated with age understatement become zeros. This reduction of the space of parameters makes the search for a unique solution more feasible.

Two conditions must be met for this standard pattern to play a helpful role. The first is that the probabilities of age overstatement always be larger than the probabilities of age understatement. The second is that the conditional distribution of $n$, the integer number of years by which individuals exaggerate (diminish) their true age, be approximately the same among those who over and understate ages. Figure 3.1 displays predicted probabilities of over and understating ages by age, $\theta_x^o, \theta_x^u$, Figure 3.2 displays the differences $\varphi_x^S = \theta_x^o - \theta_x^u$, and Figure 3.3 shows predicted conditional probabilities of over stating ages by $n$ years with $0 < n \leq 10$ or $\rho_x^o(j)$. These figures show that at least in the case of Costa Rica, the first condition is always satisfied and the second is approximately met. Differences between observed and expected quantities are of small magnitudes and concentrated at higher values of $n$, where the probabilities of over(under) stating are very small. These two items, the pair of age-specific differences between predicted probabilities of over and under statement (Table 3.6) and the conditional probabilities of overstating by $n$ years (Table 3.7), *constitute the standard pattern of age net overstatement.* The introduction of the standard simplifies the off-diagonal cells of a redefined matrix of net age overstatement, $\hat{\Theta^S}$, as all entries for age understatement become zeros. Under conditions described below, we attain identification and a unique solution for $\phi^{no}$, a parameter measuring the magnitude of net overstatement relative to the standard pattern, is possible.[16]

[16]The representation we use throughout implies that the pattern of age misreporting in any country is a multiple of the standard pattern. Although this helps the algebra and derivation of proofs, we follow a roundabout algorithm. In fact, we generate new patterns from the standard one by defining the function $\texttt{logit}(\varphi_x^i) = \alpha + \beta\texttt{logit}(\varphi_x^S)$, set the value of $\beta$ equal to 1, and then identify the level of age overstatement

Table 3.6: Estimated parameters of best logistic models for age misreporting.

| Variable | Overreporting Coeff(se) | Underreporting Coeff(se) |
|---|---|---|
| True age[1] | 0.014(.0036) | 0.002(.0040) |
| Constant | -2.127(.271) | -1.846(.297) |
| N | 6290 | 6290 |

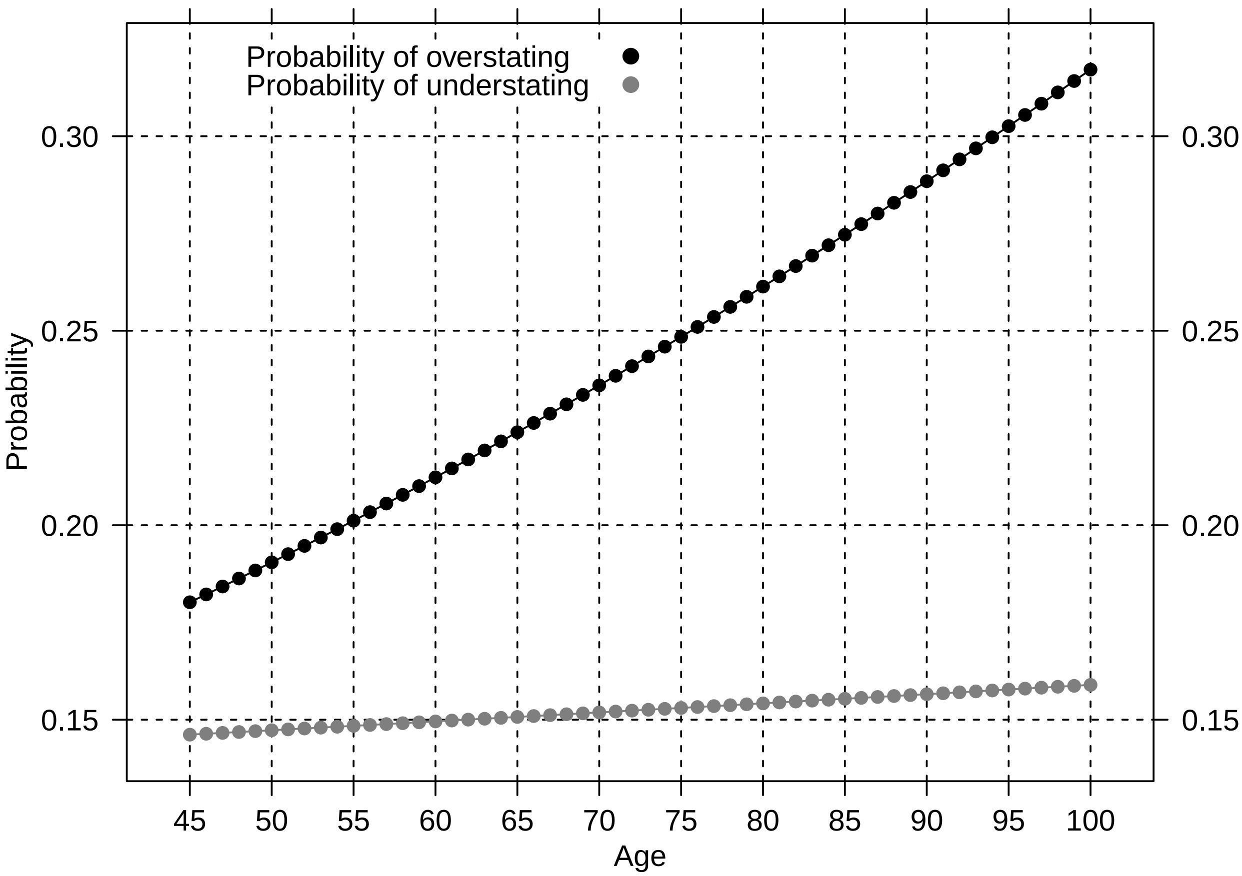
[1] Regressions estimated using sampling weights. Sample includes population with true age 40 and older and excludes ambiguous cases and foreign citizens.

Table 3.7: Average (conditional) probabilities of overreporting ages.

| | Probability[1] | |
|---|---|---|
| n | Overstating | Understating |
| 1 | 0.621 | 0.510 |
| 2 | 0.191 | 0.128 |
| 3 | 0.079 | 0.091 |
| 4 | 0.040 | 0.052 |
| 5 | 0.023 | 0.041 |
| 6 | 0.015 | 0.035 |
| 7 | 0.009 | 0.028 |
| 8 | 0.007 | 0.026 |
| 9 | 0.005 | 0.013 |
| 10+ | 0.009 | 0.060 |

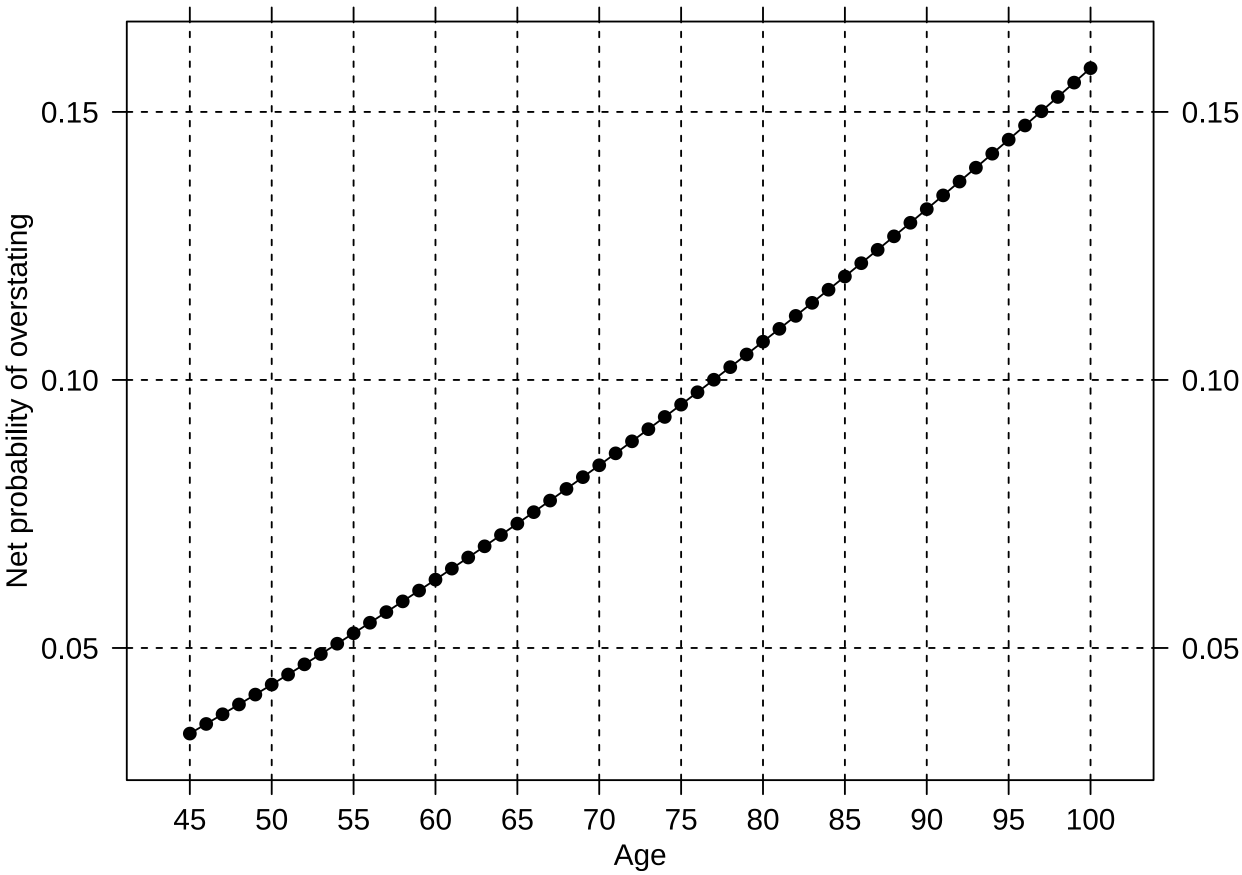
[1] Predicted values computed from a null multinomial logistic model with 10 categories, n=1786 (males and females).Estimation using sampling weights. Figures may not add up to 1 due to rounding errors.

Figure 3.1: Predicted probabilities of over(under) stating ages.



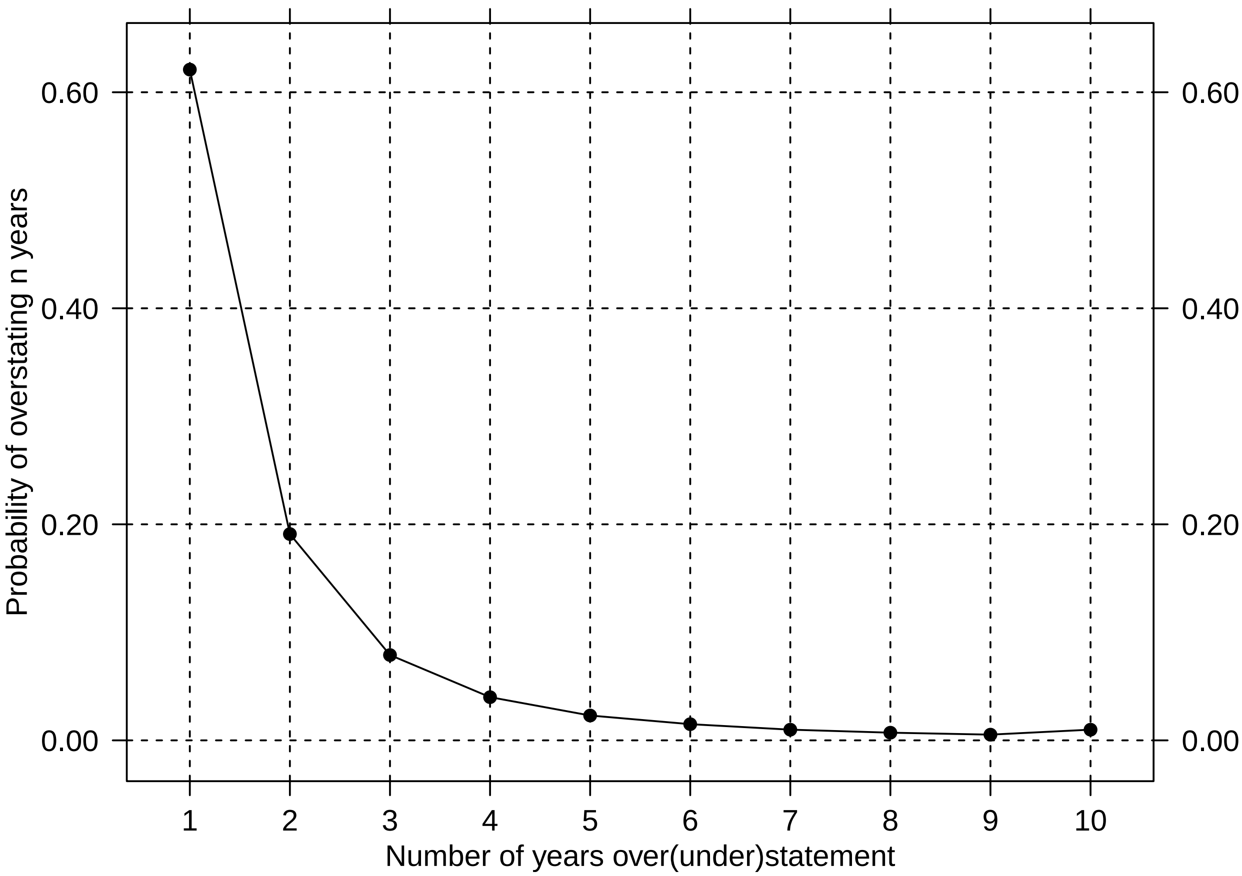

Source: Costa Rica Special study of 2000 population census.

Figure 3.2: Predicted probabilities of net overstating ages.




Source: Costa Rica Special study of 2000 population census.

Figure 3.3: Conditional probabilities of overstating age by $n$ years.




Source: Costa Rica Special study of 2000 population census.

The developments above only refer to age misreporting in population counts. However, it is known that mortality rates may also be distorted by age misreporting of ages at death (Rosenwaike, 1987). The nature of the problem in this case is somewhat different since it is not the decedent who declares the age at death but a kin or someone else unrelated to the decedent. A handful of studies based on record linkages show that there is misreporting of ages at death as well, albeit of lower magnitude than found in population counts, and that the bulk of it consists of overstatement (Rosenwaike and Preston, 1984). This is confirmed by the application of indirect techniques designed to detect age at death overstatement in a number of low and high income countries (see below). It follows that expressions analogous to (3.4.1) and (3.4.2) must be applicable for death counts as well. To make the problem tractable one needs an empirical approximation to a matrix analogous to $\Theta$ but specialized to ages at death. To our knowledge no such matrix has ever been estimated in LAC or in any other human population and we are unaware of any accessible and highly accurate national data that could be used for such purpose. In what follows we assume that the standard age pattern of age misstatement of death counts is identical to that of age misstatement of population counts but *its level may be different*. This assumption enables us to define the final model of age misreporting as a set of two equations with two unknown parameters:

$$\Pi^o = \phi^{no}\hat{\Theta}^S\Pi^T \tag{3.4.4}$$

$$\Delta^o = \lambda^{no}\hat{\Theta}^S\Delta^T \tag{3.4.5}$$

where $\Delta^T$ and $\Delta^o$ are the true and observed distributions of death counts and $\lambda^{no}$ is the magnitude of net overstatement of ages at death relative to the standard pattern. In closed populations equations (3.4.4) and (3.4.5) are naturally related (see below) and it is unlikely that there is always a unique solutions for $\phi^{no}$ and $\lambda^{no}$ unless we either fix the value of one of them or, alternatively, retrieve solely their ratio. A brief proof of lack of identification is in Section 2 of Chapter 10 and solutions for empirical estimation are in section 3.4.1.

**Simulated distortions III: combining age misreporting and defective coverage**

We now have all the ingredients to generate distorted populations using as benchmarks the five demographic profiles defined above. Since each of these is combined with two mortality patterns, there is a total of 10 profiles to consider. Letting $C_1$ and $C_2$ take on values between 0.80 and 1.0 in intervals of 0.05 whereas $C_3$ takes on values between 0.75 and 1.0 in intervals of 0.05.This yields a total of 175 different combination of defective completeness. When combined with 10 master populations, we generate a total of 1,750 populations. The unknown parameters controlling the levels of net age overstatement of population and death counts were assigned values ranging between 0 and 3 in intervals of 0.5 for a total of 36 possible patterns of age misreporting. When combined with the previous 1,750 populations they generate 63,000 populations. Finally, to represent age varying completeness of population and death registration we define two patterns, one with higher understatement at ages 45-54 and 70+ (concave upward) and another with higher understatement at ages over 70 (J-shaped). When combined with 10 master populations and 36 patterns of age misreporting

in a population $i$ by fixing $\alpha$ so that $\varphi_x^i \sim \phi^o\varphi_x^S$, where $\phi^o$ is the desired level of age over reporting.

Table 3.8: Methods to adjust for completeness of death registration[1,2,3]

| Method | Assumptions | Data | Target | Notes | Cite | Acronym |
|---|---|---|---|---|---|---|
| Brass I | 1-2-3-4-5 | B | Census counts | NA | Brass (1975) | br2Ce |
| Preston-Hill I | 1-2-3-4-5 | B | Census counts | begin at age 5 | Preston and Hill (1980) | ph5Ce |
| Preston-Hill II | 1-2-3-4-5 | B | Census counts | begin at age 10 | Preston and Hill (1980) | ph10Ce |
| Preston-Hill III | 1-2-3-4 | B | Census counts | begin at age 15 | Preston and Hill (1980) | ph15Ce |
| Martin | 1-2-3-4-6 | B | Death counts | NA | Martin (1980) | marCo |
| Bennet-Horiuchi I | 1-2-3-4 | A | Death counts | forward accum 5 | Bennett and Horiuchi (1981) | bh1Co5 |
| Bennett-Horiuchi II | 1-2-3-4 | A | Death counts | backward accum 75- | Bennett and Horiuchi (1981) | bh1Co75 |
| Bennett-Horiuchi III | 1-2-3-4 | A | Death counts | forward accumu 5+ | Bennett and Horiuchi (1981) | bh2Co5 |
| Bennett-Horiuchi IV | 1-2-3-4 | A | Death counts | backward accum 75- | Bennett and Horiuchi (1981) | bh2Co75 |
| Bennett-Horiuchi Ia | 1-2-3-4 | A | Death counts | adj growth rate | Bennett and Horiuchi (1981) | bh1Co5mix |
| Bennett-Horiuchi IIa | 1-2-3-4 | A | Death counts | adj growth rate | Bennett and Horiuchi (1981) | bh1Co75mix |
| Bennett-Horiuchi IIIa | 1-2-3-4 | A | Death counts | adj growth rate | Bennett and Horiuchi (1981) | bh2Co5mix |
| Bennett-Horiuchi IVa | 1-2-3-4 | A | Death counts | adjusted growth rate | Bennett and Horiuchi (1981) | bh2Co75mix |
| Brass I | 2-3-4 | A | Death counts | NA | Brass (1979, 1975) | br1Co |
| Brass II | 2-3-4 | A | Death counts | Variant of Br1Co | Brass (1979, 1975) | br2Co |
| Preston-Bennett | 1-2-3-4 | A | Death counts | NA | Preston and Bennett (1983) | pbCo |
| Preston-Lahiri | 1-2-3-4 | A | Death counts | begin age 10 | Preston and Lahiri (1991) | pl10Co |
| Preston-Hill Ia | 1-2-3-4-5 | A | Death counts | begin at age 5 | Preston and Hill (1980) | ph5Co |
| Preston-Hill IIa | 1-2-3-4-5 | A | Death counts | begin at age 10 | Preston and Hill (1980) | ph10Co |
| Preston-Hill IIIa | 1-2-3-4 | A | Death counts | begin at age 15 | Preston and Hill (1980) | ph15Co |

[1] See section 6 of Chapter 10 for additional definitions.
[2] Keys for assumptions : (1) Identical completeness of census counts in both census, (2) Closed to migration, (3) No age misreporting, (4) Invariant completeness by age, (5) Stability, (6) Quasi stability
[3] Keys for required data: ( A) Two censuses and intercensal deaths, (B) One census and one to three years of deaths by age

we obtain 720 additional populations. Altogether there are a total of 63,720 simulated populations observed over a period of 10 years each for a total of 6,372,000 populations in single years of age.

### Defective relative completeness: identification and adjustment

The most important techniques to detect and adjust for faulty completeness evaluated in this study are briefly summarized in Table 3.8.[17] The table highlights (a) key assumptions on which the techniques rely, and (b) information required to implement each of them. These methods share important commonalities and all but two abstain[18] from invoking stability assumptions. Yet they differ in at least one feature that, under suitable empirical conditions, grants them an advantage over competing methods. The following are key assumptions of these methods:

- Computation of rates of growth: with two exceptions all methods require computation of age specific rates of growth in an intercensal period. Because observed rates may be perturbed by differential census completeness, the estimates of the main parameter

[17] A more detailed description of each as well as citations is in Section 4 of Chapter 10 .The names we use as labels for each method are convenient tags to identify them and are not meant to reflect contributions of individual researchers. We reviewed and assessed a longer list of techniques and, with two exceptions, chose to consider only those that did not rely on the assumption of stability or quasi-stability.

[18] Names of methods are in Table 3.8.

(relative completeness of death registration) could be biased if the method is sensitive to differential census completeness. A way around this is to first adjust for relative completeness of census registration and then apply any of the relevant techniques using adjusted age specific rates of growth. This idea was first put forward by Hill and Choi (Hill and Choi, 2004; Hill et al., 2009) who suggests that one of the methods listed in the table, namely, Brass I be used to retrieve a robust estimate of the ratio of completeness of both censuses.

- Population closed to migration: none of the methods in Table 3.8 works well in the presence of significant intercensal migration. If information on net migration is available, it must be used to adjust the observed rates of intercensal growth.[19]

- Absence of age misreporting: all methods assume either no age misreporting or, alternatively, age misreporting that perturbs only trivially the figures of cumulative population above adult ages. This poses a conundrum: if, as asserted before, LAC population and mortality counts are heavily affected by age overstatement, how can one expect to obtain precise estimates of relative completeness using techniques that are highly vulnerable to age misreporting? Two conditions offer an escape from this trap. The first is that the type of age misreporting that predominates in LAC is net age overstatement. When using cumulative populations over some age $x$ the damage done to the target quantity by age misreporting only depends on population flows across age $x$. It is insensitive to transfers of population above age $x$. Furthermore, the relative volume of flows, e.g. the relative error of the target quantity, is generally light for late adulthood (less that 65 or 70) though it begins to mount after age 75 or so. Since in all cases computations only require to employ ages up to ages 70 or 75, the impact of age overstatement will be minor.[20] The second favorable condition that circumvents the problem is that the optimal method (variants of Bennett-Horiuchi I) is also the least sensitive to age misreporting of the type encountered in LAC (see below).

- Age invariant relative completeness of death registration: all techniques rely on the assumption that relative completeness of death registration is age invariant and all of them are moderately sensitive to departures from it. As we show later, however, our choice of optimal method (Bennett-Horiuchi I) is least vulnerable to violations of the assumption.

- Estimation of life expectancy at older ages: all methods adopt *ad hoc* procedures to handle the open age group. These procedures rely on exogenous computations of parameters relating the quantity of interest, life expectancy at age 75 or 70 and selected

[19]Hill and colleagues also investigated the effects of intercensal migration (Hill et al., 2009). In the simulations performed here we do not include consideration of migration but its effects are partially captured via differential censuses completeness.

[20]This is because even when there is heavy age overstatement the population at any particular age $y < x$, where $x <= 65$, is a small fraction of the population above age $x$. These ratios increase as $x$ increases due to approximately exponential decrease of population at older ages.

observed quantities in the data at hand. The relations are estimated using model life tables, stable population expressions, numerical approximations, or a combinations of all these. In the applications implemented here we follow the methods suggested by the authors in each case. Thus, some of the variability in performance, albeit a small part, is due to heterogeneous strategies to handle the open age group.

### Defective age misreporting: identification

A key component of our analysis is the detection and identification of patterns of age misstatement in the population and death counts. As shown in a previous section, the distortions associated with age misreporting in population and death counts are more complex than those involving only faulty completeness. Detection of the problem is difficult since its manifestations are quite subtle and, in the absence of overt and striking phenomena such as the US Black-White cross over, is likely to remain concealed and undetected. There are two well-tested methods to identify the existence of age over(under) statement in either population or death counts. The first method requires an external data source with correct dates of birth (or ages) in a population at a particular time that can be compared to age-specific census counts at approximately the same time. An example of this is the utilization of Medicare data in the US, a source of information that, as a rule, contains both population exposed and mortality data. Because Medicare data are linked to Social Security records and these are known to register age with high precision, mortality rates computed from Medicare data are a gold standard against which conventional mortality rates could be contrasted and their quality evaluated (Elo and Preston, 1994). If one ignores the existence of a population not covered by Medicare records, it is also feasible to link individual census records to Medicare records and investigate more precisely the nature of patterns of age misreporting in census counts. If, in addition, Medicare records are linked to the US National Death Index (NDI), it is then possible to repeat the same operations and assess the quality of reporting of age at deaths. In all cases one must assume that the coverage of population in both sources is complete or, if incomplete, identical in both.[21] Record linkage from multiple sources such as those illustrated above has rarely been used for it is costly and involves resolution of complicated confidentiality issues.

A second method is less data demanding, considerably less expensive, and simple to apply but can only *reveal* the existence of age misreporting in one of the two sources and provides few clues about its nature. The procedure was proposed by Preston and colleagues (Rosenwaike and Preston, 1984; Elo and Preston, 1994; Bhat, 1990; Grushka, 1996) and has been applied in countries of North America, Western Europe and in Latin America (Condran et al., 1991; Grushka, 1996; Dechter and Preston, 1991; Palloni and Pinto, 2004; Del Popolo, 2000). In a nutshell, the method consists of comparing cumulative population counts in a census in year $t_1$ to the expected cumulative population counts in a second population census in year $t_2$. The computation of expected quantities requires both an initial census opening

[21] The assumption is more restrictive that we made it sound: if population coverage is not complete in either source, then the subpopulations missed in each census must be random relative to their true and reported age.

the intercensal interval, a second census at time $t_2$ closing the intercensal interval, and age specific deaths counts in the intercensal period spanning an interval of $k = (t_2 - t_1 + 1)$ years. The ratio of observed to expected population is an indicator of age misstatement:

$$cmR^o_{x,[t_1,t_2]} = \frac{cmP^o_{x+k,t_2}/cmP^o_{x,t_1}}{1 - (cmD^o_{x,[t_1,t_2]}/cmP^o_{x,t_1})} \tag{3.4.6}$$

where $cmP^o_{x,t_1}$ and $cmP^o{}_{x,t_2}$ are cumulative populations over ages $x$ and $x+k$ in the first and second census respectively and $cmD^o_{x,[t_1,t_2]}$ is the cumulative number of deaths after age $x$ during the intercensal period. This expression is a simple contrast between two different estimates of the same population parameter, namely, the cumulative survival ratio: the denominator uses the complement of the observed ratio of (cumulative) intercensal deaths to (cumulative) population in the first census whereas the numerator expresses it as the survival ratio computed from the cumulative counts in two successive population censuses. The behavior of this index can be summarized as follows:[22]

1. When there are no errors, the values of the two estimates of the cumulative survival ratios will be identical and the index will be exactly 1;

2. When there is systematic age overstatement of population counts ONLY, the index will be less than 1 and will slope downward with age;

3. When there is systematic age overstatement of death counts ONLY, the index will be larger than 1 and will slope upwards with age;

4. When there is systematic age overstatement of BOTH population and death counts, the index will be generally larger than 1 and, with some exceptions, will slope upwards with age (but much less so than in case (3) above).

These expected impacts of age misreporting on the index are consistent both with results from previous simulation studies (Condran et al., 1991; Palloni and Pinto, 2004; Grushka, 1996) and with our own simulation.

The above suggests that the observed sequence of values $cmR_{x,[t_1,t_2]}$ provides partial indication, albeit not completely unambiguous, about the nature and levels of systematic age misreporting in any particular case. Before venturing too far, however, three notes of caution are needed. First, empirical patterns of age overstatement of deaths and populations could offset each other and produce ratios close to 1. That is, it is possible (but unlikely) that in scenario (4) the ratios $cmR_{x,[t_1,t_2]}$ are 1 at all ages even though there is net age overstatement in population and death counts. Because of this possibility, diagnostics based on the observed value of $cmR_{x,[t_1,t_2]}$ alone can only detect consistency (including error consistency) rather than accuracy of age declaration in population and death counts (Dechter and Preston, 1991).

[22] An informal algebraic justification of the expected behavior of the index is in Section 4 of Chapter 10.

Second, throughout we assumed that there is perfect coverage of both population and death counts and that the sequence $cmR_{x,[t_1,t_2]}$ could only be distorted by age misreporting. This is an unrealistic assumption, at least in LAC countries. In Section 4 Chapter 10 we show that, under conditions of defective census and death registration coverage, the values of the sequence $cmR_{x,[t_1,t_2]}$ will also depend on $C_{t_1}$, $C_{t_2}$ and $CD_{[t_1,t_2]}$, the completeness of the first and second census, and the average completeness of intercensal death registration, respectively.[23] As shown in Section 4 of Chapter 10, the intrusion of $C_{t_1}$, $C_{t_2}$ and $CD_{[t_1,t_2]}$ in the expression for $cmR_{x,[t_1,t_2]}$ makes it impossible to separate the influence of age overstatement and of defective completeness. Lack of completeness will generate values of the index that are far away from 1 *even if there is no age misreporting at all.* As a consequence, the observed values of $cmR_{x,[t_1,t_2]}$ cannot be used to infer patterns of age misreporting *unless population and death counts are first suitably adjusted for defective completeness.*

Third, like disparities in defective completeness, intercensal migration will distort the sequence of values $cmR_{x,[t_1,t_2]}$ even in the absence of errors in population and death counts or age distributions. If migration is known to have taken place, the observed ratios must be adjusted for age specific migration counts.

**Defective age misreporting: adjustment**

From the above, it is plain that one cannot learn much about patterns of age misreporting unless population census and death counts are first adjusted for completeness. Any method to adjust for age misreporting should only be applied if the observed data passes two basic checks. The first is that the sequence of values $cmR_{x,[t_1,t_2]}$ must be free of errors associated with defective completeness. If this is not the case, the observed sequence must be adjusted for defective completeness. The adjusted values of the sequence are computed as follows

$$(ADJcmR_{x,[t_1,t_2]}) = \frac{(C_{t_2}/C_{t_1})*(cmP^o_{x+k,t_2}/cmP^o_{x,t_1})}{1-(.5*(C_{t_1}+C_{t_2})/CD_{[t_1,t_2]})*(1/((C_{t_1}/C_{t_2})+1))*(cmD^o_{x,[t_1,t_2]}/cmP^o_{x,t_1})} \quad (3.4.7)$$

an expression that includes the observed data and the estimated adjustment factors for completeness.

The second check must ensure that the sequence of *adjusted* values is well-behaved. By this we mean that it must contain only positive values and there should be no sharp discontinuities. Negative values and/or sharp discontinuities are caused by inappropriate adjustments for relative completeness of death registration and/or violation of some or all of the assumptions made about age misreporting. But they can also be a consequence of erratic behavior of very low counts of population and deaths at extreme ages. In this case, it is advisable to trim the age groups to be considered.

Once the ratios are adjusted and their regular behavior ascertained, there remains the task of retrieving estimates of the *magnitude* of net adult age overstatement of population

[23] The quantities $C_{t_1}$, $C_{t_2}$ and $CD_{[t_1,t_2]}$ are the ratio of the observed to the true counts in the first and second census and death respectively. Following standard practice, we assume that completeness of population and death registration are age invariant. In the evaluation study described later we identify procedures that are robust to violations of this assumption. This result justifies following standard practice throughout.

and deaths. The model of age misreporting developed before rests on a known standard of age net overreporting and includes two unknown parameters, $\lambda^{no}$ and $\phi^{no}$, measuring the magnitude of net age overstatement in population and death counts respectively.

Before describing methods to retrieve parameters of age misreporting using the adjusted ratios, $ADJcmR_{x,[t_1,t_2]}$, we summarize an entirely unexpected property of the simulated populations that proves to be very helpful for identification of the unknown parameters.

**An important regularity in the simulated populations** Given its nature, it should be intuitively clear that the adjusted (for completeness and migration) sequence of values $cmR_{x,[t_1,t_2]}$ must be closely related to age and to the magnitude of net age overstatement, namely, $\lambda^{no}$ and $\phi^{no}$. Less intuitive is the nature of such a relation. It came as a surprise to us that a very simple linear model captures the relation in the simulated population. The model is as follows:

$$(cmR_{ix,[t_1,t_2]})^{-1} = \alpha_{0x} + \alpha_{1x}\lambda_i^{no} + \alpha_{2x}\phi_i^{no} \tag{3.4.8}$$

where $i$ is an index for the *simulated population*, $x \geq 45$ refers to age and, importantly, the values of $cmR_{ix,[t_1,t_2]}$ are distorted *only* by age misreporting, not by defective completeness.

In this model the independent "variables" are the values of the levels of age misreporting $\lambda_i^{no}$ and $\phi_i^{no}$ in the *ith* population whereas $\alpha_{0x}$, $\alpha_{1x}$ and $\alpha_{2x}$ are parameters estimable from the simulated data. Table 3.9 displays estimates of coefficients for the independent variables $\lambda_i^{no}$ and $\phi_i^{no}$ from these simulated population. The table shows that the fit of the model is very good and, importantly, that the estimated values of the constant of the model is always close to 1, as it should be when the parameters $\lambda_i^{no}$ and $\phi_i^{no}$ drift to 0.[24]

How can this finding help us to estimate the unknown parameters $\lambda^{no}$ and $\phi^{no}$? If the population observed by the investigator is a member of the simulated set, the observed sequence of values $(cmR_{ix})^{-1}$ must obey equation (3.4.8). Thus, knowing what the values of $\alpha_{0x}$, $\alpha_{1x}$, and $\alpha_{2x}$ are in the simulated populations suffices to identify the unknown parameters that generate the sequence $cmR_{x,[t_1,t_2]}$. This requires to simply "invert" the relation represented by (3.4.8) as follows: for any observed population we define the vector of values $[cmR_{x,[t_1,t_2]}]^{-1}$ for all $x \geqslant 45$ as the 'dependent variable' and the corresponding vectors containing the values of the coefficients for ages $x \geqslant 45$ in Table 3.9 as the "independent variables". We then estimate a regression equation using as many observed values of $[cmR_{x,[t_1,t_2]}]$ as there are single year age groups in the observed data. The estimated regression coefficients should be unbiased estimates of the pair of unknown parameters ($\lambda^{no}$, $\phi^{no}$).

Table 3.10 displays results of the inverse procedure applied to the simulated populations with a limited combination of values of the unknown parameters. The first two columns of the table display estimates of the parameters $\lambda^{no}$, $\phi^{no}$ whereas the third and fourth columns display the actual values of these parameters in the simulated data. The last column of the table displays the values of $R^2$. The table shows that, given the vector of values $\{cmR_{x=45,\ldots,100}\}$ from the simulated populations, the vectors of parameters $\{\alpha_{1x=45,\ldots,100}\}$ and $\{\alpha_{2x=45,\ldots,100}\}$ extracted from Table 3.9 used as independent variables, there is a best

[24]Recall that the range of feasible values of the parameters of interest are in the closed interval $\sim [0, 3]$. Values outside this range produce implausible death and population age distributions.

Table 3.9: Estimated regression models relating index of age misstatement and parameters of age misreporting[1]

| Age | $\alpha_0$ | $\alpha_1$ | $\alpha_2$ | $R^2$ |
|---|---|---|---|---|
| 45 | 1.000 | -0.027 | -0.004 | 1.000 |
| 46 | 1.000 | -0.012 | -0.005 | 1.000 |
| 47 | 1.000 | -0.006 | -0.005 | 1.000 |
| 48 | 1.000 | -0.003 | -0.006 | 1.000 |
| 49 | 1.000 | 0.000 | -0.007 | 1.000 |
| 50 | 1.000 | 0.002 | -0.008 | 1.000 |
| 51 | 1.000 | 0.003 | -0.009 | 1.000 |
| 52 | 1.000 | 0.005 | -0.010 | 1.000 |
| 53 | 1.000 | 0.006 | -0.011 | 1.000 |
| 54 | 1.000 | 0.008 | -0.013 | 1.000 |
| 55 | 1.000 | 0.010 | -0.014 | 1.000 |
| 56 | 1.000 | 0.012 | -0.016 | 0.999 |
| 57 | 0.999 | 0.014 | -0.019 | 0.999 |
| 58 | 0.999 | 0.017 | -0.022 | 0.999 |
| 59 | 0.999 | 0.020 | -0.025 | 0.999 |
| 60 | 0.999 | 0.024 | -0.030 | 0.999 |
| 61 | 0.999 | 0.029 | -0.035 | 0.999 |
| 62 | 0.999 | 0.035 | -0.041 | 0.999 |
| 63 | 0.998 | 0.042 | -0.048 | 0.999 |
| 64 | 0.998 | 0.051 | -0.057 | 0.998 |
| 65 | 0.997 | 0.062 | -0.069 | 0.998 |
| 66 | 0.996 | 0.076 | -0.082 | 0.998 |
| 67 | 0.995 | 0.094 | -0.099 | 0.997 |
| 68 | 0.994 | 0.116 | -0.121 | 0.997 |
| 69 | 0.992 | 0.145 | -0.148 | 0.996 |
| 70 | 0.990 | 0.183 | -0.183 | 0.995 |
| 71 | 0.986 | 0.231 | -0.228 | 0.995 |
| 72 | 0.982 | 0.295 | -0.285 | 0.994 |
| 73 | 0.975 | 0.378 | -0.360 | 0.992 |
| 74 | 0.966 | 0.490 | -0.458 | 0.991 |
| 75 | 0.952 | 0.638 | -0.586 | 0.989 |

[1] Data from simulated populations

Table 3.10: Estimates and true values of parameters of net age overstatement from inverse method.[1]

| run | $\phi^{no}$ | $\hat{\phi}^{no}$ | $\lambda^{no}$ | $\hat{\lambda}^{no}$ | $R^2$ |
|---|---|---|---|---|---|
| 1 | 0.000 | 0.061 | 0.350 | 0.370 | 1.000 |
| 2 | 0.000 | 0.002 | 0.700 | 0.685 | 1.000 |
| 3 | 0.000 | -0.059 | 1.050 | 0.999 | 1.000 |
| 4 | 0.000 | -0.118 | 1.400 | 1.313 | 1.000 |
| 5 | 0.000 | -0.178 | 1.750 | 1.628 | 1.000 |
| 6 | 0.000 | -0.238 | 2.100 | 1.942 | 1.000 |
| 7 | 0.000 | -0.298 | 2.450 | 2.256 | 1.000 |
| 8 | 0.000 | -0.358 | 2.800 | 2.571 | 1.000 |
| 9 | 0.350 | 0.393 | 0.700 | 0.727 | 1.000 |
| 10 | 0.350 | 0.392 | 1.050 | 1.078 | 1.000 |
| 11 | 0.350 | 0.391 | 1.400 | 1.429 | 1.000 |
| 12 | 0.350 | 0.390 | 1.750 | 1.780 | 1.000 |
| 13 | 0.350 | 0.388 | 2.100 | 2.130 | 1.000 |
| 14 | 0.350 | 0.387 | 2.450 | 2.481 | 1.000 |
| 15 | 0.350 | 0.386 | 2.800 | 2.832 | 1.000 |
| 16 | 0.700 | 0.710 | 1.050 | 1.067 | 1.000 |
| 17 | 0.700 | 0.755 | 1.400 | 1.445 | 1.000 |
| 18 | 0.700 | 0.801 | 1.750 | 1.823 | 1.000 |
| 19 | 0.700 | 0.846 | 2.100 | 2.201 | 1.000 |
| 20 | 0.700 | 0.892 | 2.450 | 2.579 | 1.000 |
| 21 | 0.700 | 0.938 | 2.800 | 2.957 | 1.000 |
| 22 | 1.050 | 1.013 | 1.400 | 1.393 | 1.000 |
| 23 | 1.050 | 1.096 | 1.750 | 1.791 | 1.000 |
| 24 | 1.050 | 1.179 | 2.100 | 2.189 | 1.000 |
| 25 | 1.050 | 1.262 | 2.450 | 2.587 | 1.000 |
| 26 | 1.050 | 1.345 | 2.800 | 2.985 | 1.000 |
| 27 | 1.400 | 1.303 | 1.750 | 1.704 | 1.000 |
| 28 | 1.400 | 1.416 | 2.100 | 2.117 | 1.000 |
| 29 | 1.400 | 1.530 | 2.450 | 2.530 | 1.000 |
| 30 | 1.400 | 1.643 | 2.800 | 2.943 | 1.000 |
| 31 | 1.750 | 1.582 | 2.100 | 2.004 | 0.999 |
| 32 | 1.750 | 1.720 | 2.450 | 2.427 | 1.000 |
| 33 | 1.750 | 1.859 | 2.800 | 2.851 | 1.000 |
| 34 | 2.100 | 1.851 | 2.450 | 2.292 | 0.999 |
| 35 | 2.100 | 2.009 | 2.800 | 2.723 | 1.000 |
| 36 | 2.450 | 2.110 | 2.800 | 2.569 | 0.998 |

[1] Data from simulated populations

(in mean squared error sense) solution for the unknown parameters of model (3.4.8).[25] A comparison of 'true' (first and third columns) and estimated parameters (second and fourth columns) reveals satisfactory concordance.When one of the unknown parameters is close to 0, e.g. the simulated data contains no age overreporting, the inverse technique could produce a negative estimate for that parameter. But even so, it will always generate an accurate estimate for the other parameter as long as this is different from zero. A negative estimate is thus a tell-tale sign that the unknown parameter is too close to its lowest boundary and that any adjustment should only be a function of the other unknown parameter.

[25] The model (3.4.8) is 'best' in the sense that interaction terms or higher order moments of the independent variables do not reduce the mean squared error by a statistically significant amount.

**Alternative methods to retrieve parameters of age misreporting** There are three different methods to identify estimates of $\lambda^{no}$, $\phi^{no}$. Each of these relies on assumptions and computations that are somewhat different and their sensitivity to errors in the data or to violation of assumptions are dissimilar.

1. *Constrained regression.* This is simply the inverse procedure described above with one added feature, namely, that the linear regression equation is estimated constraining the parameter space for $\lambda^{no}$, $\phi^{no}$ to be in the closed interval $\sim [0, 3]$. It is, of course, important to verify that the regression fits the data well. In addition, the constrained estimates should be approximately equal to the unconstrained ones. Significant differences may be an indication of violation of some of the assumptions.

2. *Search for optimal pair of parameters* We could define a countable set of possible combinations of values of the unknown parameters $\lambda^{no}$, $\phi^{no}$ within the permissible range. Each pair will generate a set of predicted values for the elements of the vector $cmR_{x,[t_1,t_2]}$. The mean (median) absolute difference between these predicted vectors and a vector of 1's is a measure of errors associated with the combination of parameters $\lambda^{no}$, $\phi^{no}$ that generated the predicted vector. One could then choose the combination that minimizes the mean (median) absolute difference between the two vectors. It may be the case that there are multiple pairs of estimates that perform well and distinguishing among them could be difficult. If so, a plausible strategy is to construct adjusted life tables with each of the competing pairs of estimates. In a subsequent step, one reevaluates their performance in light of the consistency of the adjusted life tables with other known life tables for the same population in different periods.

iii *Parametric method* The third method seeks to reproduce $[cmR_{x=45,\ldots,100}]$ as a function of age and then map parameters of the function onto the pairs ($\lambda^{no}$, $\phi^{no}$) that generated the data. It consists of fitting a hyperbola to a range of values of $cmR_x$

$$cmR_x = \beta_1/(\varsigma - age)^{\beta_2} \tag{3.4.9}$$

where $\varsigma$ is set equal to 76 [26] We then use the estimated parameters of function (3.4.9) to predict the pair of values $(\lambda^{no}, \phi^{no})$. Although the fit of the hyperbolic function to the simulated data distorted by age misstatement is very tight, the retrieval of the hidden parameters governing net age overstatement is generally poor. This is due to under-identification: if one uses the entire range of values attainable by $\lambda^{no}$ and $\phi^{no}$ , the function $cmR_{x=45,\ldots,100}$ can be mapped onto multiple pairs ($\lambda^{no}$, $\phi^{no}$). The procedure works best when the pair of values ($\lambda^{no}$, $\phi^{no}$) is within a limited range (approximately

[26] In cases when the values of the magnitude of age overstatement approaches the largest values allowed in the simulation (close to 3), the function $cmR_x$ attains a point of discontinuity where the derivatives with respect to age do not exist. In order to avoid such cases we used trial values for the parameter $\varsigma$ and find that, in the space of simulated populations, $\varsigma = 76$ is optimal as it always avoids points of discontinuity. This is equivalent to saying that one cannot reproduce the function for ages above 76, a trait that is partially responsible for under identification.

[0.10-2.8/3.0]). One could use all three methods and check consistency of results. If method (iii) departs from the other two, even if the parameters $\lambda^{no}$ and $\phi^{no}$ within the permissible range, methods (i) and (ii) should be preferred.

### 3.4.2 Stage II: Adjustments for defective completeness

In this section we assess the performance of techniques to adjust for defective relative completeness of censuses and death counts. The evaluation is based on results from the simulated populations described before, a space of fictitious populations and deaths counts generated by five different demographic profile, two mortality patterns and a broad range of error patterns. We evaluate the techniques' effectiveness to retrieve population parameters in two situations. In the first of these, estimation is carried out completely ignoring error patterns embedded in the space of simulated populations. In the second, we apply the techniques to selected subsets of populations with known conditions that violate a subset of assumptions.

#### Assessment of methods to adjust for defective completeness: results

The set of techniques to detect and adjust for faulty completeness evaluated in this study are summarized in Table 3.8.[27] The table identifies techniques, highlights key assumptions on which they rely and the information required to implement each of them. They all share important commonalities and only two of them (Brass I and Preston-Hill I -III) invoke the assumption of stability. They differ in at least one feature that, under suitable empirical conditions, potentially grants them a competitive advantage over other methods.

The combination of highly heterogeneous demographic conditions, diversity in the weaknesses of national vital statistics and population censuses counts, and variability of adjustment techniques, each relying on specialized assumptions, makes the choice of adjustments for any particular case a non-trivial matter. Ideally, one would like to be able to choose a very small set of techniques that, under given empirical conditions, produce optimal estimates. To support this endeavor our evaluation study assesses the performance of candidate techniques by applying them to the 63,720 thousands simulated. We then compute multiple error measures under the simulated set of conditions that violate (or not) assumptions on which the techniques rest. Although others could have been chosen, the results of the evaluation we describe here are based on one error measure, namely, the mean absolute value of the proportionate error, MAPE. For any given technique we observe a distribution of MAPE that corresponds to well defined conditions (e.g. violation (or not) of assumptions). For example, suppose we use a technique $T$ in all populations that do not violate any of the assumptions on which the technique relies. We would not expect the numerical value of the parameter estimated by $T$ to always be identical to the population parameter as computations rely on a number of approximations whose impact may vary depending on the nature of the population being examined (stable versus non stable, under model West or under model South, etc.). Thus MAPE are truly random and should have a mean equal to 0. Assume now that the technique is applied in populations that violate a subset of its

[27] We considered a longer list of methods but, with two exceptions, chose to test only those that *did not* rely on the assumption of stability or quasi-stability.

assumptions. In this case, MAPE will have both a random and a systematic component and will be distributed with a non-zero mean identical to the expected value of the bias associated with the technique given conditions that violate assumptions. More generally, we can compute not just the expected value (and bias) but also the entire distribution of MAPE for each technique and for each set of conditions that violate assumptions we care to specify. In particular, we can calculate medians, quartiles and the probability that MAPE is less than 0.05.

Table 3.11 displays statistics for the quantity MAPE associated with each technique under three different scenarios.[28] Panel A is for scenarios that include population either stable or non-stable and where the completeness of the two successive census may be different but there is no age misreporting. Panels B and C are for scenarios where the populations may experience higher age misreporting in death counts than in population counts (age misreporting 1 and Panel 2) or higher age misreporting in population counts than in death counts (age misreporting 2).[29]

To describe more precisely the techniques' performance, it is useful to keep the following rules in mind:

i. In the absence of any knowledge whatsoever about errors or deviations from stability, a search for the best method should be concentrated in panel A of Table 3.11.

ii. When exogenous information suggests stability and not much else, the search should focus on the subset of stable populations or panel B of Table 3.11. Instead, when there is prior empirical data confirming violation of stability, for example past shifts in fertility regime, but one can be agnostic about completeness and age misreporting, the search of optimal method should concentrate on panel C.

iii. When, in addition to lack of stability, there is evidence of defective coverage of population and death counts but no hint of significant net age overstatement of adult ages, the search should shift to the subset in panel D.

iv. When a scenario as in (iii) above *and* age misreporting is suspected, identification of optimal method should be done using panel E.

v. When scenario (iv) is most reasonable and is likely that completeness of two censuses is defective but equally so in both censuses, identification of the optimal choice ought to be carried out with panel F.

vi. When there is evidence suggesting that census completeness is age-dependent (as defined before) and identical in both censuses, panel G is most relevant.

[28] To simplify the table we display statistics for only a subset of all the techniques we studied. These are identified by acronyms in Table 3.8.

[29] This is one of many tables we were able to assemble that targeted different subsets of assumptions and display errors associated with violations in those assumptions. Other tables that display errors when different classes of assumptions are violated could be built.

The results in the table highlight a number of salient characteristics. First, as already noted before, Brass I method to estimate relative completeness of two consecutive censuses is uniformly good, regardless of population subset (first row of the Table). Even under the worst conditions (Panels B and C) the method delivers an optimal performance for estimation of relative completeness of two censuses. Note that with probability 1 it will produce an estimate that is within 5 percent of the true value of the parameter.[30]

Second, the magnitude of errors are larger when census coverage is defective and completeness is *not* the same in both censuses. This is because all methods except Brass I and Preston-Hill I-III rely on direct computations of age specific growth rates from the observed data, a quantity that will be in error when there is different coverage errors in two successive censuses. Indeed, the performance of these methods improves substantially when there is accurate census coverage, when one adjusts rates of growth for deficient census coverage (Bennet-Horiuchi Ia-IVa) or when coverage is defective but the same in both censuses (see Table 3.11 panel D)

Third, age misreporting affects the accuracy of all estimates but more so in some cases (Brass I and the second variant of Preston-Hill than in others (Bennett-Horiuchi, all variants).

Fourth, the magnitude of errors that obtain when relative completeness is age dependent (panel G) varies sharply by technique but, in general, are lowest for Bennett-Horiuchi's variants.

Thus, excluding population with defective census completeness, the optimal choice is always the 2-stage variant of the Bennett-Horiuchi (Bennet-Horiuchi Ia and IIa) followed very closely Brass I and Preston-Hill I and II irrespective of violations of assumptions about age misreporting.

Second, estimates from the multiple variants of Bennett-Horiuchi technique perform quite badly, even in the absence of age misreporting (Panel A) and so do all the other methods *except variants of Bennett-Horiuchi with adjusted rates of growth* (Bennett-Horiuchi Ia-IVa). This is a consequence of the fact that changing completeness of the censuses bounding an intercensal interval, biases the age specific rates of growth. This is a problem to which the adjusted Bennett-Horiuchi technique is much less sensitive to. The third finding is that, under the most general and worst conditions (Panels B), even the optimal method (adjusted Bennett-Horiuchi technique) does not have an impeccable record, as one would not expect its estimates to be with 5 percent of the true value in less than 30 percent of the cases. A similar result obtains in Panel C. In both cases though the performance of the method is satisfactory as the median error is less than 7 percent.

---

[30] To move beyond verification based purely on simulations only, Section 4 of Chapter 10 compares our estimates of ratios of relative completeness in two successive censuses computed using Brass I with ratios computed by CELADE's. Although the agreement of both sets of figures is quite close, caution should be used in the interpretation of the table. First, agreement in ratios of completeness may also be produced with incorrect census-specific estimates of completeness. Second, CELADE's figures are based on estimates and projection of populations and do not always rely on post-enumeration assessments and or indirect techniques. Thus, the agreement we ascertain can be an agreement of incorrect figures.

### Sensitivity of adjustments for completeness to age misreporting

If we exclude populations with defective census completeness, the *optimal* choice of techniques to adjust for relative completeness of death registration is always one of the variants of Bennett-Horiuchi method. Importantly, however, the Bennett-Horiuchi technique does not perform well unless a correction is introduced to adjust for different completeness of population counts in the first and second census. This could, of course, be a serious limitation were it not for a second result of our evaluation study, namely, that the modified Brass I technique to estimate relative completeness of death registration also produces a robust estimate of relative censuses completeness, namely, of the ratio $C_{t_1}/C_{t_2}$ (Hill et al., 2009). In their original study, Hill and colleagues included a limited set of distortions due to age misreporting. The same finding is replicated in our study based on simulations of a much larger array of distortions due to age misreporting. It follows that estimates from Brass I and Bennett-Horiuchi are sufficient to correct the observed values of the ratios $cmR_x$.[31]

Overall, these are remarkably fortunate results for they suggest that it is possible to adjust the sequence $cmR_x$ for defective completeness of population and death registration *even if the observed data are contaminated by age misreporting.*[32] If this were not the case, a quest to correct the data for systematic age misreporting would be futile unless coverage of census and death counts are perfect (or equally bad).

[31] The estimates Brass I and all variants of Bennett-Horiuchi techniques are *mean optimal*, in the following sense: the average error of the estimates they produce are lower than those of other techniques under all conditions spawned by the simulated populations. It does not mean that, once these techniques are applied to observed data, the adjusted mortality rates (and derived functions of the life table) will also be best estimates. This is because the sensitivity to violations of assumptions of techniques we include in the evaluation study varies depending on the *particular subset of assumptions that are violated.* Therefore, the Brass I and Bennett-Horiuchi estimates are, on average best, but they may not be the best choice were we to restrict examination to populations where a few assumptions are not met (age misreporting) but others are (censuses differential completeness). Another way of saying this is that the strategy we propose based on these results can only aspire to identify a mean global optimal, rather than a mean local optimal, candidate technique among alternative possible ones.

[32] "Fortunate results" may be an overstatement. Insensitivity of some techniques that adjust for defective death and population to errors of age misreporting is more or less expected due to the utilization of cumulative rather than age-specific counts of population and deaths.

Table 3.11: Evaluation of performance of techniques to adjust for relative completeness of population and death counts.[1]

| Method | Panel A: Stable/non-stable-time varying completeness | | | | Panel B: Stable/non-stable-time varying completeness Age misreporting (1) | | | | Panel C: Stable/non-stable-time varying completeness Age misreporting (2) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Median** | **IR** | **psCV** | $P(e \leq .05)$ | **Median** | **IR** | **psCV** | $P(e \leq .05)$ | **Median** | **IR** | **psCV** | $P(e \leq .05)$ |
| br2Ce | 0.00090 | 0.00111 | 1.22865 | 1 | 0.00471 | 0.00268 | 0.57023 | 1 | 0.00313 | 0.00197 | 0.62894 | 1 |
| ph5Ce | 0.00058 | 0.00082 | 1.41164 | 1 | 0.00698 | 0.00169 | 0.24210 | 1 | 0.00066 | 0.00079 | 1.19297 | 1 |
| ph10Ce | 0.00074 | 0.00068 | 0.91928 | 1 | 0.00751 | 0.00184 | 0.24544 | 1 | 0.00066 | 0.00097 | 1.46997 | 1 |
| ph15Ce | 0.00110 | 0.00083 | 0.75621 | 1 | 0.00802 | 0.00193 | 0.24106 | 1 | 0.00136 | 0.00103 | 0.75301 | 1 |
| bh1Co5 | 0.29675 | 0.25504 | 0.85943 | 0 | 0.28939 | 0.26050 | 0.90019 | 0 | 0.29929 | 0.30292 | 1.01213 | 0 |
| bh1Co75 | 0.29798 | 0.25597 | 0.85901 | 0 | 0.29036 | 0.26676 | 0.91872 | 0 | 0.30097 | 0.26768 | 0.88938 | 0 |
| bh2Co5 | 0.29711 | 0.25514 | 0.85874 | 0 | 0.28918 | 0.25797 | 0.89209 | 0 | 0.29969 | 0.29992 | 1.00076 | 0 |
| bh2Co75 | 0.29798 | 0.25597 | 0.85901 | 0 | 0.29036 | 0.26676 | 0.91872 | 0 | 0.30097 | 0.26768 | 0.88938 | 0 |
| bh1Co5_mix | 0.05409 | 0.06675 | 1.23406 | 0.95 | 0.05729 | 0.07375 | 1.28721 | 0.27 | 0.05550 | 0.06880 | 1.23974 | 0.5 |
| bh1Co75_mix | 0.05523 | 0.06775 | 1.22679 | 0.95 | 0.05729 | 0.07553 | 1.31845 | 0.27 | 0.05382 | 0.06782 | 1.26013 | 0.43 |
| bh2Co5_mix | 0.05428 | 0.06691 | 1.23282 | 0.95 | 0.05744 | 0.07418 | 1.29146 | 0.27 | 0.05561 | 0.06904 | 1.24147 | 0.5 |
| bh2Co75_mix | 0.05512 | 0.06828 | 1.23884 | 0.95 | 0.05747 | 0.07549 | 1.31346 | 0.26 | 0.05385 | 0.06808 | 1.26435 | 0.43 |
| br1Co | 0.02735 | 0.06853 | 2.50537 | 0.65 | 0.06797 | 0.08792 | 1.29356 | 0.34 | 0.08056 | 0.11322 | 1.40545 | 0.4 |
| br2Co | 0.00549 | 0.00560 | 1.01895 | 1 | 0.04759 | 0.02030 | 0.42655 | 0.57 | 0.04974 | 0.02592 | 0.52109 | 0.5 |
| pbCo | 0.68908 | 0.86442 | 1.25445 | 0 | 0.78071 | 0.81245 | 1.04065 | 0.005 | 0.78071 | 0.75151 | 0.96259 | 0 |
| plCo5 | 0.48008 | 0.90139 | 1.87757 | 0.0014 | 0.48294 | 0.91790 | 1.90068 | 0.0007 | 0.48536 | 0.95293 | 1.96337 | 0.0007 |
| ph5Co | 0.05374 | 0.04835 | 0.89970 | 0.44 | 0.06458 | 0.08109 | 1.25567 | 0.46 | 0.05334 | 0.04700 | 0.88100 | 0.45 |
| ph10Co | 0.05315 | 0.04662 | 0.87716 | 0.45 | 0.06515 | 0.08431 | 1.29406 | 0.46 | 0.05369 | 0.04706 | 0.87666 | 0.45 |
| ph15Co | 0.05252 | 0.04556 | 0.86758 | 0.47 | 0.06559 | 0.08679 | 1.32327 | 0.46 | 0.05331 | 0.04874 | 0.91427 | 0.46 |

[1] See Table 3.8 for key features of each method.

List of acronyms:

IR=Interquartile range;

psCV=Interquartile range/Median;

$P(e \leq .05)$=Probability of error less than 0.05

**Adjustment for completeness: summary of results**

These results suggests application of the two-step strategy adopted in the construction of all those LAMBdA life tables that rely on annual vital statistics (mostly after 1950)[33]:

i. In the absence of exogenous information about the difference in completeness between the two census and if the assumption of age invariant completeness holds, use Brass I method to estimate the relative completeness of two consecutive censuses;

ii. Adjust the observed rate of intercensal growth to account for defective relative completeness of censuses use the two-stage procedure (Bennett-Horiuchi Ia-IVa) (see Section 7 of Chapter 10)[34]

The evaluation study identifies a handful of optimal procedures to adjust for relative completeness of death registration. The problem that remains unsolved is whether or not one should also adjust for net adult age overstatement and, if so, whether the techniques derived before can lead to an optimal procedure as well. We pursue this below.

### 3.4.3 Stage III: Adjustments for defective age reporting

Do the procedures to identify and then adjust for age misreporting formulated before produce robust estimates of the true population parameters? [35] To answer this question we select the subset of simulated populations with age misreporting and defective completeness, we adjust for completeness following the two-step strategy described above, compute the indicator of age misreporting that identifies the existence of age misreporting, and then correct using the first methods defined in section 3.4.1). The main results are in Tables 3.9 and 3.10 display the main results.

Table 3.9 contains parameters associated with expression 3.4.8. These parameters were estimated from the subset of simulated population that contained age misreporting but no census coverage errors. For each age group there is a linear relation between the age misreporting index and the values of the two parameters that define the magnitude of age overstatement of population and deaths. The fit is nearly perfect and the estimated constant is close to one everywhere, irrespective of age, as it should be. These regression equations imply that the age-specific index of age overstatement, $cmR_x$ can be predicted with high accuracy using the parameters of the linear relation and the knowledge of the two parameters of age overstatement, $\lambda^{no}$ and $\phi^{no}$. This holds for all ages of interest and in all populations

[33] A highly conservative strategy, which we did not pursue, is to simply average out the results of the four best methods.

[34] A caveat is in order. Note that when relative completeness is age dependent, Bennett-Horiuchi (or any of the other three best performers) is *mean optimal*, in the sense that the weighted average of relative death completeness of observed data will be best estimated by Bennett-Horiuchi methods (or some of the other best performers). It does not mean that, once applied, the adjusted mortality rates (and derived functions of the life table) will also be best estimates of the corresponding true functions. None of the methods we include in our evaluation can escape from the assumption of constant relative completeness and, therefore, we can only aspire to find a mean optimal candidate (s).

[35] This section summarizes and generalizes results described in Section 3.4.1.

with either no errors of census coverage, with coverage errors are equal in both censuses and, finally, when the quantity $cmR_x$ is adjusted for unequal census coverage.

These relations are seemingly useless since what we seek to identify are, after all,the values of the unknown parameters,$\lambda^{no}$ and $\phi^{no}$. The question we might ask is the following: is there a way of using the known adjusted (for census coverage) values of the age-specific indeces of age overstatement, $cmR_x$, for all ages above age 45, to retrieve the two unknown parameters of age overstatement?. The answer to this question is affirmative if *we condition on knowledge of the parameters in Table 3.9* . Indeed, if those parameters are known we can use them jointly with the observed (and adjusted for completeness) age-specific values of the index $cmR_x$ to infer values for $\lambda^{no}$ and $\phi^{no}$. This is tantamount to "inverting" the model in table 3.9 , that is, to predict $cmR_x$ using *known values of the parameters in the table as the independent variables* . The estimated parameters of these regression will be the the unknown quantities $\lambda^{no}$ and $\phi^{no}$. Said otherwise, we estimate a linear regression in which the dependent variable is the set of age specific adjusted indeces $cmR_x$ and the *parameters of the regression models* in Table 3.9 as independent variables. This regression should yield slopes that are robust estimates of the two unknowns $\lambda^{no}$ and $\phi^{no}$ for the population. To show that this is indeed the case we focus on the sub-space of simulated populations with no defective coverage and with age distortions induced by combinations of the unknown parameters of age misstatement. We then choose the subset of all simulated populations characterized by *identical pairs of age overstament parameters* and estimate regressions with the index of age misstatement as dependent variables and the parameters of the relation estimated in Table 3.9 as independent variables. Because in the simulated data there are 36 possible pairs of unknown parameters, we estimate 36 different regression equations. Table 3.10 shows the results of this "inverted" procedure when applied to the simulated populations. The regression of $cmR_x$ on the known vectors of estimates of the parameters $\alpha_{1x=45,...,100}$ and $\alpha_{2x=45,...,100}$ has a tight fit and yields estimates of parameters of net age overstatement that are minimally affected by errors.[36]

The above results suggest a straightforward procedure to estimate the values of $\lambda^{no}$ and $\phi^{no}$ in any concrete case: regress the age-specific values $cmR_x$ (adjusted for census coverage) on the corresponding age specific parameters $\alpha_{1x=45,...,100}$ and $\alpha_{2x=45,...,100}$ in Table 3.9. Estimates of the resulting slopes are optimal estimates of the unknown parameters. The key requirement for the inverted procedure to work is that it is legitimate to *condition on the known values of* $\alpha_{1x=45,...,100}$ *and* $\alpha_{2x=45,...,100}$, *that is, to assume that the observed population belongs to the space of simulated populations.*

**Sensitivity assessment: general considerations**

How sensitive is the procedure to adjust for age misreporting? In this section we summarize key assumptions and review results of a study to assess sensitivity of adjustments to departures from the recommended standard of age misreporting. Recall that if errors are detected, the strategy to adjust for age misreporting is fairly straightforward but rests on the validity of a handful of assumptions. We claim that only one of these is fragile and that

[36]This is method (i) described in Section 3.4.1.

departures from it are more likely to induce errors than departures from the others. Below we begin by setting up an ideal model for estimating age misreporting from observable data. We believe the model to be sensible and realistic enough to capture most cases of empirical age misreporting. Although ideal, the model is not estimable without simplifications and constraints. We argue that the most relevant among these is the assumption involving the age pattern of net age misreporting. The model is as complex as we can think of and can be summarized as follows:

1. there is an age pattern of population age underreporting and a possibly different age pattern of population age overreporting. These patterns are embedded in a set of age-specific probabilities of over and under reporting age $x$, say $PU(x)$ and $PO(x)$;[37]

2. different populations experience the same age patterns of under and over age of population misreporting but could experience different intensities or magnitudes reflected in two age invariant parameters, $\theta_u$ and $\theta_o$;

3. there is an age pattern of age at death underreporting and a possibly different age pattern of age at death overreporting. These patterns are embedded in a set of age-specific probabilities of over and under reporting population ages, say $DU(x)$ and $DO(x)$;

4. different populations experience the same patterns of age at death misreporting but could experience different intensities reflected in two parameters, $\gamma_u$ and $\gamma_o$.

Note that the above scenario is exactly analogous to those invented by demographers to render tractable the empirical heterogeneity of human mortality, fertility and migration using model mortality age patterns. There is no discordance between these and the treatment of age misreporting suggested here. The only objection could be that whereas there are fixed and identifiable "biological constraints" that confine the shape of the force of mortality, and fertility and migration age patterns to well-known but small range, none really exists that could unequivocally apply to human age misreporting.

Suppose the only observed data are two population censuses, intercensal deaths counts, and mortality rates unaffected by defective coverage of events or populations. An observed mortality rate at age $y$, $M(y)$, can then be written as a function of the ratio of two quantities: first, the true number of deaths at various ages, $D(x)$, transformed by functions of $\theta_u * DU(x)$ and $\theta_o * DO(x)$ that distort the age distribution of deaths and , second, the true population at various ages, $P(x)$, also transformed by functions of $\gamma_u * PU(x)$ and $\gamma_o * PO(x)$. The range attained by $x$ will depend on the ages at which misreporting becomes significant and, in addition, on how far from the true value $y$ are located the ages from (into) which counts at ages y are transferred to (from).

[37] To avoid cluttering, we simplify. More realistically, we should assume patterns of unconditional probabilities that a true age x (in a population or death count) is recorded as age $x \pm k$ where $k$ can attain integer values values $1, 2, \ldots$

The foregoing model is as complicated as can get. It depends on the validity of the following assumptions: (a) there are two known age patterns of misreporting, one for population and the other for death counts and (b) observed age misreporting can be represented by simple shifts of the age patterns dependence controlled by 4 'shift' parameters. If neither of these assumptions is accepted, the situation is hopeless unless information other than mortality rates and counts becomes available.

A solution to the problem posed by the model above requires observable quantities, such as $cmR_x$ and estimates of 4 parameters. Unless the standard functions are very peculiar, identification of these parameters is impossible. Indeed, the following situation can take place: a true mortality rate could be distorted by two pairs parameters ($\theta_o$ and $\theta_u$) and ($\gamma_o$ and $\gamma_u$) that yield the same observed mortality rate than an alternative two pairs of parameters that distort the true quantity in opposite ways, for example, implying less excess deaths in the numerator but additional excess populations in the denominator. Under these conditions the observable quantities are not sufficient to identify the four unknown parameters.

To circumvent the problem we need additional constraints. One possibility is to assume that age overstatement overwhelms understatement and that observable counts can be reproduced using two age patterns of net overstatement (the result of subtracting the age pattern of overstatement from the age pattern of understatement) and two parameters representing net overstatement of death and of populations. Although this constraint is helpful, it does not entirely solve the identification problem unless we also constrain the functional shape of the net age pattern of age misreporting. At this point previous research may come to the rescue. Indeed, we know, for example, that the propensity to overstate ages of population increases with age. It may not be a stretch ( and perhaps one could verify this with specialized data) to assume that the propensity to overstate ages at death is also an increasing function of age. Under these conditions the identification problem vanishes. And, as our simulation study demonstrates, when both age patterns of errors are the same there is no identification problem whatsoever.

In summary: the complexity of the problem makes identification of desired parameters intractable. The only solution is to introduce constraints, the most important of which regards the identity of age patterns of age (net over) statement. If only one of them is known the only escape is to assume that the unknown age pattern of errors is similar or identical to the known one.

At this point the problem becomes one of judging the worth of the adjustment: are the corrected estimates derived from the simplified model better estimates than the uncorrected ones, e.g., computed ignoring the problem altogether, as has been the norm so far? There is plenty of empirical evidence demonstrating that in most known populations there is a tendency to overstate ages of population. If this were the only error in counts, all adult mortality rates would be underestimated and increasingly so with age.[38] If ages at death are underreported, errors associated with population age overreporting could be reduced,

[38]This applies universally to populations that have been growing in the recent past so that their age distributions slope downward exponentially (approximately).

completely offset, or even reversed. But, this is an unlikely scenario. More likely is that ages at death are overestimated and, if so, mortality rates at older ages will also be underestimated and increasingly so with age. Consequently, adjustments that shift upwards the observed age specific rates without distorting the age pattern of mortality, are likely to produce estimates that are closer to the true values.

The arguments made above suggest that the key issue is whether or not misidentification of the age pattern of age misreporting leads to estimates with proportionate absolute errors larger than the unadjusted figures? A closely related issue is whether the identity assumption, e.g. that age patterns of errors in population and death counts are identical. To investigate these questions we undertook a sensitivity study and simulated data with alternative patterns of age misreporting. We then compute adjustments erroneously assuming that the Costa Rican pattern of age misreporting prevails. The results are summarized in the following section.

**Sensitivity assessment: alternative standards of age misreporting**

How sensitive is the age misreporting adjustment method adopted in LAMBdA to violations of the two key underlying assumptions, namely, the identity assumption and the one regarding the standard age pattern of errors?

In what follows we assess the sensitivity of adjusted estimates of population and death counts to violations of the assumptions regarding the pattern of age misreporting. To do so, we use a subset of simulated populations consisting of an initially stable population under Model South distorted by 175 patterns of defective completeness. We then apply a new standard of age misreporting and introduce, as we did in the original simulation, 36 combinations of values of levels of age misreporting, 6 for death and 6 for population counts. This leads to a total of 6,300 simulated populations characterized by patterns of age misreporting different from the Costa Rican one. We then apply the proposed adjustment procedure (which requires to invoke the assumption of a Costa Rican pattern of age misreporting) and retrieve an adjusted sequence of values $cmR_{x,[t_1,t_2]}$, estimates of unknown parameters for levels of age misreporting, and adjusted life tables. We then compare selected statistics of the adjusted life tables with the life table that generated the data. The differences between the two are a measure of the errors associated with misidentification of the age pattern of age misreporting.

**A single alternative pattern of age misreporting.** Without additional constraints, the number of potential candidates to become alternative standard for net age overstatement is infinite. To narrow down the set of plausible candidates we modify separately the probabilities of net overstatement and the conditional probabilities of overstating by $n$ years.

First, we choose a standard for the probabilities of net overstatement that satisfies two conditions:

1. Condition 1: it has approximately the same probabilities of net overreporting at ages 45 and and 100 as the Costa Rican standard. This condition constrains the level parameters to be within the same range or parameter space as those compatible with the Costa Rican standard, e.g. $(0-3)$.

2. Condition 2: the new standard probabilities increase more rapidly with age than in the Costa Rican standard. This will reflect situations were the standard pattern producing the data imply worse age misreporting than is embedded in the Costa Rican standard.

The function that defines the probabilities of net overstatement is $P(x) = 0.18 * (1 - S(x)) + 0.15$ where $S(x)$ is a Gompertz survival function with level parameter $\alpha = 0.030$ and slope parameter $\beta = 0.80$. It attains a value equal to 1 at age 45 and a median value at age 58. Other transformations of the function $S(x)$ are of course possible. The function we use here distorts in significant ways the shape of the Costa Rican standard (from linear to logistic). It also maximizes differences in probabilities between ages 45 and 90 while simultaneously allowing room for level parameters to increase (decreases) these probabilities to the same maximum and minimum levels allowed by the Costa Rican standard.

Second, the conditional probabilities of misreporting age by $n$ years follows a nearly symmetrical binomial distribution with binomial probability $p = 0.50$. This is in stark contrast with the (approximately) negative binomial distribution embedded in the Costa Rican standard.

Figure 3.4 displays the unconditional and conditional probabilities embedded in the alternative standard.

**Effects of using an incorrect standard of age misreporting.** Results of the sensitivity exercise are in Figure 3.5. The figure displays the cumulative distribution of relative errors in estimates of life expectancy at age 45 (top panel) and 60 (bottom panel). These figures reveal two properties of the resulting estimates. First, the bulk of errors (over 95 percent) are positive, namely, the estimated values of life expectancy are higher than the true ones. This is consistent with the fact that the standard that generated the simulated populations has significantly higher probabilities of net overstatement than the Costa Rican standard used to retrieve estimates of parameters. Thus, the outcome of using a standard probabilities of overstatement than rise much slower with age than the one that generates the data will be to under adjust mortality rates and overestimate life expectancy at adult ages. Second, the distribution of errors for life expectancy at age 45 is quite benign as they are less than 5 percent in about 80 percent of cases. In contrast, the errors are larger for estimates of life expectancy at age 60 as only in 35 percent of cases are they below 5 percent.

Two final caveats. First, although the alternative pattern of age misreporting used in this sensitivity exercise departs significantly from the Costa Rican standard, it is still based on the assumption of net overstatement. But this may not be a universal feature of age misreporting. In their work on age misreporting, Preston and colleagues find that net understatement is not uncommon among US African Americans and has been found elsewhere (Preston et al., 1996)(Preston personal communication). Even though net understatement, like net overstatement, *must* lead to underestimates of old age mortality, its presence in observed data invalidates the use of a pattern of age misreporting based on net overstatement.

Second, it is also possible that overstating ages by more than 10 years is a frequent occurrence rather than a rare event. If so, the conditional distribution of $n$ assumed throughout will depart in significant ways from the true distribution as this must have a much thicker right tail. In these cases the investigator should estimate separately the density of the random variable $n$ and redefine the standard of age misreporting accordingly.

**Failure of the identity assumption: deaths and population age misreporting follow different patterns**

Because there are no known data on which to base the construction of a standard pattern of age misreporting of death counts, we assumed throughout that this was identical to the standard pattern of age misreporting of population counts. The only defense against potential problems caused by violation of the assumption is to examine the behavior of selected indicators. First, as is the case when there is misidentification of the standard pattern of population age misreporting (see above), the quantities in error will be estimates of the unknown level parameters $\lambda^{no}$, $\phi^{no}$. If departures from the identity assumption are significant, estimates of the level parameters will be implausible, e.g. they will fall outside the range contained in the simulated population set and/or the fit of sequence $cmR_{x,[t_1,t_2]}$ to the data will be deficient (even if estimates are within the legitimate range).

Of course, if the investigator suspects or has ancillary evidence that misreporting of age in death counts is light, the parameter $\phi^{no}$ could be set to zero, only parameter must be estimated, and the identity assumption is unnecessary.

The takeaway message from this exercise is that the sensitivity to departures from the standard age pattern of misreporting is not insignificant and should be carefully considered before adjusting the data. However, even in cases of extreme departures, such as the one we consider here, it is undoubtedly better to err by misidentifying the standard than by ignoring the problem altogether.

### 3.4.4 An integrated procedure to remove distortions due to defective completeness and age misreporting

The evaluation study suggests the following strategy to compute final adjustments for age misreporting:

i. In the absence of exogenous information about the difference in completeness between the two census, obtain estimates of *relative completeness* of the two census enumerations (use Brass I);

ii. Use the estimate of relative completeness obtained in the first step to correct the rates of intercensal growth and then apply one of the variants of Bennett and Horiuchi (1981) technique to estimate relative completeness of death registration;

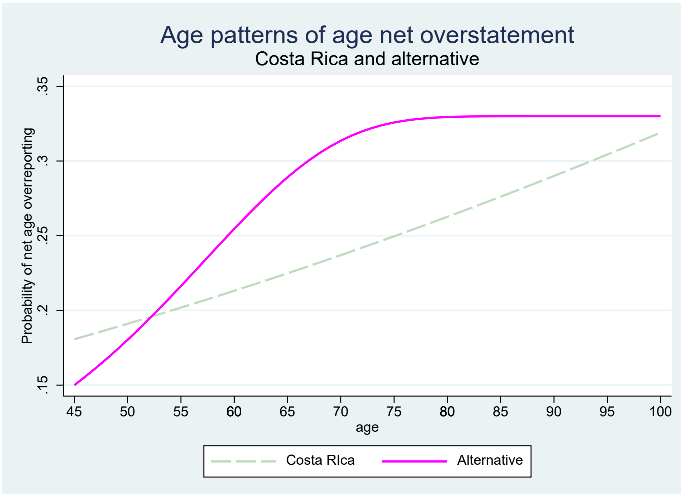
iii. Use the estimate of relative census and death completeness obtained in the first two steps to adjust the sequence of values $cmR_{x,[t_1,t_2]}$

At this point one can choose one of the following two options (or both)

iii.a Use the adjusted values of the function $cmR_{x,[t_1,t_2]}$ and apply the inverse technique in one of its three variants, unconstrained, constrained and optimal, to retrieve estimates of the unknown parameters $\lambda^{no}$, $\phi^{no}$;

or, alternatively,

Figure 3.4: Alternative standard of age misreporting: unconditional and conditional probabilities of age misreporting



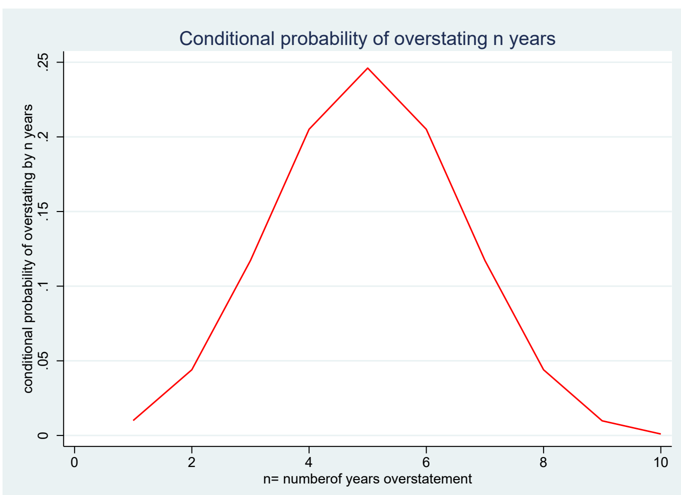
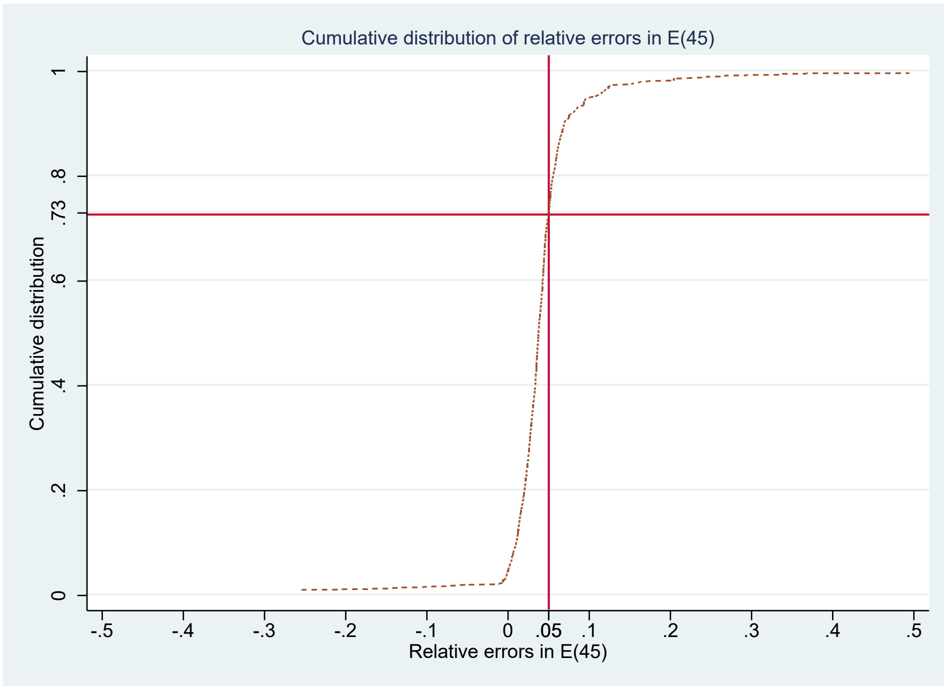
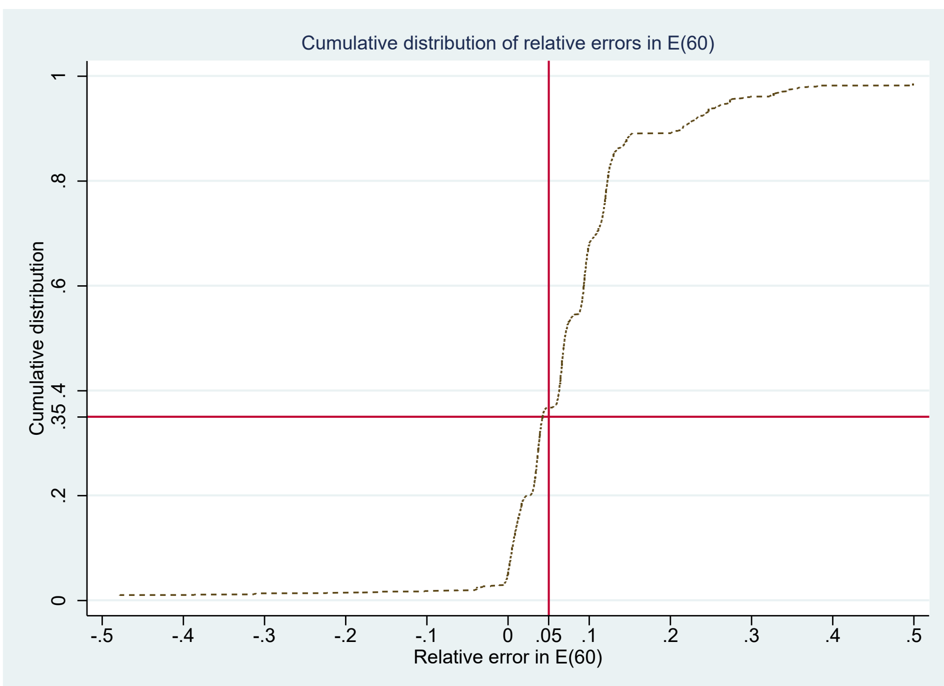



Figure 3.5: Cumulative distribution of errors in estimates of E(45) and E(60) (sensitivity to misidentification of standard schedule of age misreporting).



iii.b Use the adjusted values of the function $cmR_{x,[t_1,t_2]}$ and apply the regression-free approach to obtain best estimates of the unknown parameters of age misreporting;

iv. Compute the matrix of age misreporting, $\widehat{\Theta}^S$ and its inverse. Use these matrices and the estimates of levels of age misreporting for the unknown parameters, $\lambda^{no}$ and $\phi^{no}$, obtained in previous steps. Finally, compute adjusted age-specific population and deaths counts;

v Calculate mid-intercensal period mortality rates and adjust then for defective completeness;

vi. Compute an intercensal life table, centered in mid-period, with the adjusted intercensal mortality rates.

### 3.4.5 An illustration: the case of Guatemala

We apply the integrated procedure to data for Guatemala in the intercensal period 1981-1994. Despite much recent progress, the country's death registration and census counts are still defective and offer good testing grounds for the technique.[39]

Figure 3.6 displays plots of deviations of observed, partially, and fully adjusted sequences $cmR_{x,[1981,1994]}$. The observed sequences reflect the impact of both defective completeness and age misreporting. They exhibit the expected upward slope caused by systematic age overstatement of both population and deaths.[40] The median values of absolute deviations are 0.30 for females and 0.21 for males.

To remove errors due to defective completeness we estimate completeness of the first census relative to the second, $C_{1981}/C_{1994}$, 1.03 for males and .98 for females, and relative completeness of death registration, $CD_{[1981,1994]}/(.5*(C_{1981}+C_{1994})$, 0.90 and 0.89 for males and females. We multiply the observed values of the function $cmR_{x,[1981,1994]}$ by the adjustment factor see equation 3.4.7) and obtain the sequence of partially adjusted values plotted in the figure. As shown in the figure, adjustment for defective completeness significantly improves the behavior of the sequences, particularly among males but less so among females. The maximum deviation for males drops from about 2 to 0.5. Among females the reduction is from 7.5 to about 3.5. The median values are 0.27 and 0.056 for females and males respectively.

To adjust for age misreporting we choose the regression-free method. We identify the pair of values for $\lambda^{no}$ and $\phi^{no}$ that, when used jointly with the adjustment for defective completeness, yields a best fitting sequence $cmR_{x,[1981,1994]}$, e.g. one that minimizes absolute

[39]Estimates of relative completeness of death registration in Guatemala after 1950 range between 0.75 and 0.91.

[40]Large values of the sequence (and even sharp discontinuities leading to negative values) at very old ages are not uncommon and may not always be a sign of unusually large age overstatement. It could also be an artifact of random fluctuations of small counts at these ages and/or a result of inappropriate adjustments for defective census or vital registration coverage.

deviations from a vector of 1's. These best estimates of $\lambda^{no}$, $\phi^{no}$ are in the range (0,0.5) and (2.0-2.5) for females and (0-0.5) and (1.5-2.0) for males.[41]

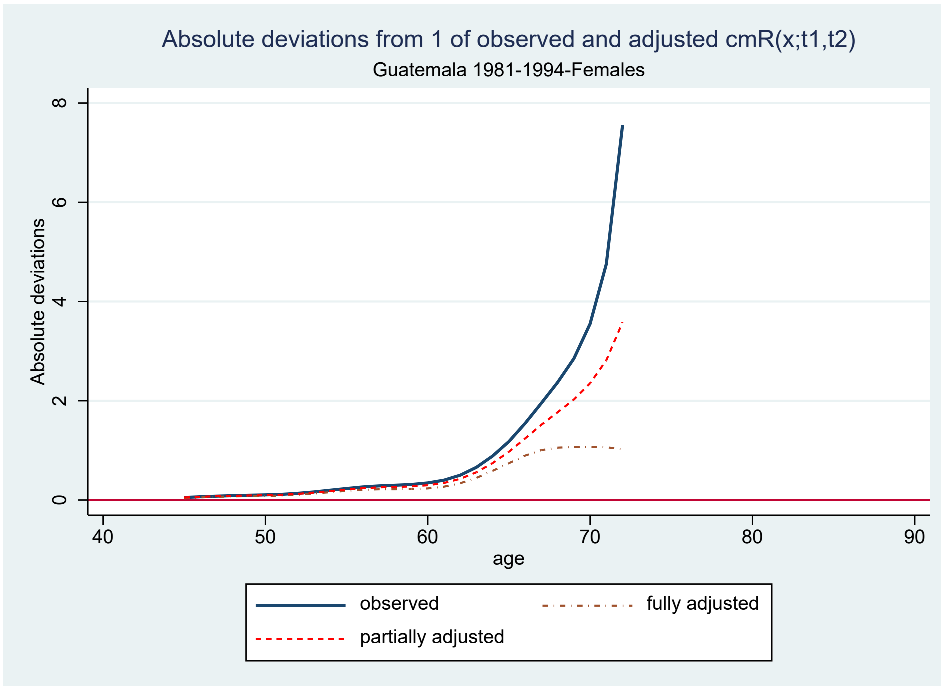
Figure 3.7 plots the median values of deviations of the fully adjusted sequences. The figure plots the median deviations of the fully adjusted sequences *for each of the 36 pairs of parameters values*. Before plotting, we rank the absolute deviations in ascending order so that the lowest value to the left of the graph is associated with the pair of optimal parameter estimates¿ In contrast, the highest value is associated with the worst performer pair. The scale of the $x$-axis is arbitrarily set to the natural numbers reflecting the rank order of the absolute deviations. In the case of females, for example, the minimum median value of absolute deviations from 1 was generated by estimates of $\lambda^{no}$ in the range (0-0.5) and estimates of $\phi^{no}$ in the range (2.0-2.5).

Ideally, the fully adjusted values of the sequence $cmR_{x,[1981,1994]}$ should be equal to or very close to 1. The smallest medians of absolute deviations of fully adjusted values from a vector of 1's plotted in the figure are .014 and .21 for males and females. They represent 7 and 70 percent of the male and female observed values, respectively, and 26 and 78 percent of the partially adjusted values. Although improvements are substantial, we live in an imperfect world and the fully adjusted values for females are less satisfactory than for males. In both cases these sequences are devoid of discontinuities, considerably flatter and closer to 1 than the observed ones but, as revealed by the values attained by absolute deviations, the adjustments are less satisfactory at the oldest ages. This could be an indication of mismatches between the assumed and underlying patterns of age reporting or imperfect adjustment for completeness of census and death registration.

In a final step, we use the inverse of the (male and female) estimated matrices $\widehat{\Theta}^S$, estimates of the two level parameters and compute adjusted (for age misreporting) vectors of age-specific population and intercensal death counts. We then calculate adjusted (for defective coverage and age misreporting) age specific intercensal mortality rates, and an adjusted life tables centered in middle of the intercensal interval. Table 3.12 displays observed, partially and fully adjusted values of life expectancy at ages 5 and 60. Partially adjusted values only reflect adjustment for relative completeness and ignore age misreporting. The relative differences between observed and partially adjusted life expectancy at age 5, on one hand, and observed and fully adjusted values, on the other, are as follows: for life expectancy at age 5 they are about 3.7 percent and 4.1 percent for males and 2.8 percent and 3.9 percent for females. For life expectancy at age 60 the contrasts are sharper: differences for females are 6.3 percent and 12 percent very similar to those for males, 6.1 percent and 12 percent.

[41] The parameters $\lambda^{no}$ and $\phi^{no}$ are real numbers and can attain an uncountable number of values in the permissible range. To short-circuit the search of the optimal pair we looped through all 36 possible combinations of discrete values 0, 0.5, 1, 1.5, 2.0 and 2.5. Thus, strictly speaking the solution we present here only identifies a *range of values* within which the "true" values are contained.

Figure 3.6: Absolute deviations of observed, partially and fully adjusted values of $cmR_{x,[1981,1994]}$ from vector of 1's: Guatemala 1981-1994



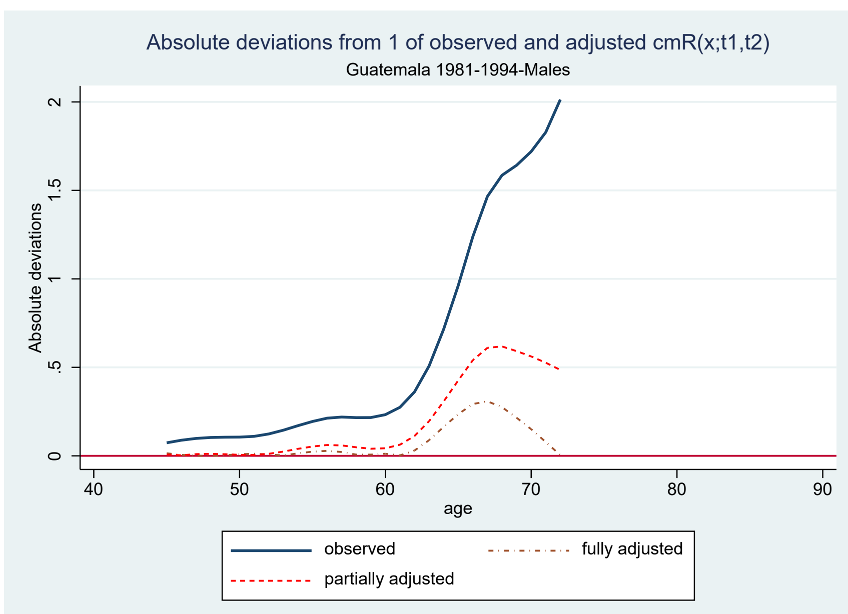
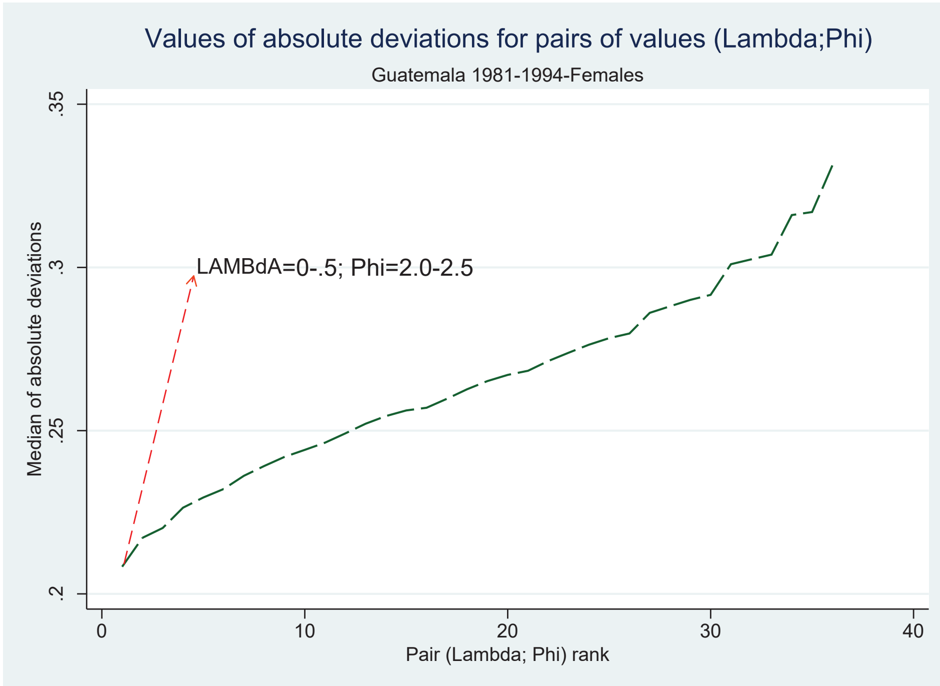



Figure 3.7: Median of deviations of fully adjusted values of $cmR_{x,[1981,1994]}$from vector of 1's for 36 pairs of unknown parameters: Guatemala 1981



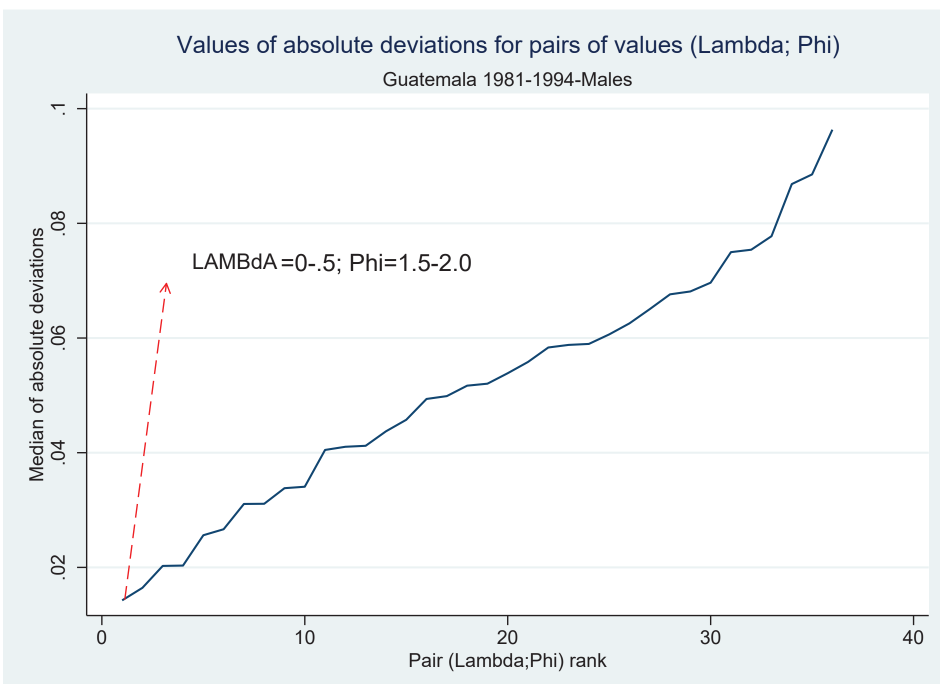



Table 3.12: Observed and adjusted life expectancy at ages 5 and 60: Guatemala, 1981-1994.

| Population | Age | Observed and adjusted life expectancy[1] Observed | Adjusted Partially[2] | Adjusted Fully[3] |
|---|---|---|---|---|
| Males | | | | |
| | 5 | 60.06 | 58.30 | 57.73 |
| | 60 | 17.11 | 16.12 | 15.24 |
| Females | | | | |
| | 5 | 65.13 | 63.37 | 62.70 |
| | 60 | 18.51 | 17.41 | 16.54 |

[1] Relative completeness of first to second census: Males=1.0256; Females=0.984; Relative completeness of death registration: Males=0.899; Females=0.888. Severity of age misreporting: Males: $\lambda^{no}$ value set to middle of range (0-.5); $\phi^{no}$ value set to middle of range (1.5-2); Females: $\lambda^{no}$ value set to middle of range (0-.5) and $\phi^{no}$ value set to middle of range (2-2.5).
[2] Adjusted for completeness only.
[3] Adjusted for completeness and age misreporting.

## 3.5 Estimation of adult mortality for the period 1850-1950

Availability of vital statistics for the period 1850-1950 is scarce (see Table 3.2). It is only in a few country-years that we can estimate life tables using techniques that rely on intercensal deaths. In some cases we use vital statistics for a period of three years centered on a population census and adjust the observed rates using Martin's variant of Brass II method that relaxes the stability assumption (Brass, 1975; Martin, 1980) or, more rarely, Bennett-Horiuchi method.[42] In the remaining cases, where we can only access census counts, we use a generalized version of the ogive approach first formulated by Coale and Demeny (1967). In this section we review the classic version of the ogive approach, a shortcut proposed by Arriaga, and a generalized version of the standard ogive method. In sections 3.5.4 and 3.5.5 we review methods used to estimate adjusted rates of growth for completeness and for migration in four countries that, at the turn of the 20th century were heavily influenced by international migration.

### 3.5.1 Classic ogive method

Assume a stable population with a natural rate of increase $r$ and mortality given by the survival function $S(y)$. The age distribution $N(y), y = 0, \ldots, \infty$, of the population is given

[42] We use Martin's variant to produce life tables for the following country years Argentina 1914, Costa Rica 1927, Mexico 1921, and Uruguay 1908.

by:

$$N(y) = N(0)\exp(-ry)S(y) \tag{3.5.1}$$

At a minimum, estimation of the full function $S(x)$ requires knowledge of $r$ and of $N(x)$ in finite age groups, e.g single or five-year age groups. To reduce noise and minimize effects of age heaping and systematic age misreporting, Coale and Demeny (Coale and Demeny, 1967) compute the cumulative age distribution from (3.5.1), namely

$$cumN(x) = N(0)\int_0^x \exp(-ry)S(y)dy \tag{3.5.2}$$

a function that is relatively insensitive to population transfers across ages older than $x$.[43] When the survival function $S(y)$ is an element of a finite set within which there is variation only due to mortality levels, expression (3.5.2) will have a unique solution for the unknown level of mortality for each age $x$. In most observed cases, data errors, inaccuracies in the observed value of $r$, or mild departures from stability, can yield different solutions for the unknown level of mortality associated with each age $x$. Coale and Demeny compute the median level estimated in a restricted range of ages that excludes very old and very young ages where errors are likely to be more pronounced.

A key difficulty remains unsolved, however. This is that $S(y)$ may belong not to a unique but to one of $M > 1$ distinct families of mortality patterns within each of which there is variability induced by differences due to levels of mortality only. If so, there will be as many as $M$ distinct solutions for each $x$ contained in the range of ages chosen. Even in the absence of multiple solutions due to defective data noted above, multiple plausible models create an identification problem that can only be resolved by *a priori* specifying the family of mortality patterns to which $S(y)$ belongs.

Two remarks about these methods are important. First, the ogive method, or any of its variants, is designed to work only within a range of ages. In particular, it was never meant to include the population below ages 5 or 10. This means that researchers can only discern from the observable population the force of mortality above ages 5 or 10. Because the contrast between mortality patterns is mostly rooted in differences in the relation between child and adult mortality, it is possible that multiple mortality patterns may identify similar levels of mortality at adult ages thus shrinking considerably the identification problem. We use this feature to our advantage in all applications of the generalized ogive method. Second, if we choose a parameterization of mortality that returns an explicit expression for $S(x)$ in terms of a handful of parameters, one for mortality level and two or three ancillary ones to identify the shape of the mortality function at various ages, it becomes feasible to solve simultaneously for the values of these parameters from the observable data. For example, if $S(x)$ is expressed as a Brass-type logit function of a standard survival function and two parameters, one for the level of mortality and the other to reflect the relation between child and adult mortality,

[43] The label 'ogive method' is due to the shape of the cumulative age distribution.

expression (3.5.2) becomes a (highly non linear) function of these parameters. It is then possible to use all observations from a limited range of ages to identify the pair of estimates that reproduces best (in squared error terms) the observed quantity $cumN(x)$. This was not the solution adopted in the original formulation of the ogive procedure because it was implemented jointly with the Coale-Demeny mortality patterns that do not admit simple parameterization. The ogive procedure adopted in LAMBdA restricts searches within two models of the Coale Demeny mortality patterns and the Latin American model in the United Nations mortality patterns.

### 3.5.2 Arriaga's shortcut

We can re-express (3.5.1) as follows:

$$\ln\left(\frac{N(y)}{N(\tilde{x})}\right) = -r(y-\tilde{x}) - I(\tilde{x}, y) \tag{3.5.3}$$

where $I(\tilde{x}, y)$ is the integrated force of mortality between ages $\tilde{x}$ and $y$.[44] When the survival function belongs to a known family it can be used to retrieve the mortality level. Arriaga's suggestion (Arriaga, 1968) is to assume an arbitrary mortality level, compute the difference $\ln(N(y)/N(\tilde{x})) - I(\tilde{x}, y)$ and then regress this variable (via OLS) on the age difference $(y-\tilde{x})$. The procedure is repeated as many times as levels of mortality one may want to consider. The life table consistent with the observed age distribution will yield a slope close to the observed value of $r$. Admitting the possibility of multiple candidates for mortality patterns requires to repeat the search with multiple mortality families and obtain multiple solutions consistent with the observed populations. Evidently, this procedure is a shortcut of the more general ogive method described before. Although is was used quite successfully, it suffers from a number of problems that are less relevant in the Coale-Demeny formulation. The first is that it requires the population to be in one or five year age groups and is highly vulnerable to age misreporting. The second problem is that the dependent variable is noisy and may produce outliers or observations with a disproportionate influence on the estimated slope, the target parameter. This will lead to misidentification of the underlying mortality level.

### 3.5.3 Generalized ogive (GO) method

When the assumption of stability is indefensible, we use expressions from generalized stable population (Preston and Coale, 1982) and write (3.5.1) as:

$$N(y,t) = N(0,t)\exp\left(-\int_0^y r(x,t)dx\right)\exp\left(-\int_0^y \mu(x,t)dx\right) \tag{3.5.4}$$

where the population distribution is observed at time $t$, $r(x,t)$ is the age specific rate of growth at age $x$ and time $t$, and $\mu(x,t)$ is the instantaneous mortality rate at age $x$ and time $t$. Expression (3.5.4) can be treated just as expression (3.5.1) to retrieve the level of mortality

[44] In empirical applications the value of $\tilde{x}$ is usually set to 5.

consistent with the observed populations. There are two important differences between GO and the standard ogive method. First, unlike the standard ogive method, GO does not invoke the assumption of stability but, instead, demands as input a complete age pattern of rates of growth. Second, unlike the standard procedure, significant migration flows do not undermine GO. Indeed, if the researcher has access to age-specific rates of net migration they can be used to recalculate the rates of growth net of migration. We implemented this variant in Argentina, Brazil, Cuba and Uruguay, for years 1850-1920, a period when these countries experienced substantial international migratory flows in both directions (see below).

The core of the procedure based on expression (3.5.4) is analogous to the application of the standard ogive method. In a first stage we compute age specific rates of growth to generate the exponent of the first exponential function in (3.5.4). In a second stage we choose a pivotal age $\tilde{x}$ and compute the quantity $\Lambda_{\tilde{x}}(y) = \ln(N(y,t)/N(\tilde{x},t)) + \int_{\tilde{x}}^{y} r(x,t)dx$.[45] We then search within a single family of life tables the sequence $-\int_{\tilde{x}}^{y} \hat{\mu}(x,t)$ that best approximates the value of $\Lambda_{\tilde{x}}(y)$. Inevitably, different ages $y \in [\tilde{x}, \breve{x}]$ identify different solutions. When that is the case, we use one of two strategies: (a) select the median solution among those obtained within the range of ages from $\tilde{x}+1$ to $\breve{x}$ or (b) select the level of mortality that minimizes

$$\sum_{\tilde{x}}^{\breve{x}} [\Lambda_{\tilde{x}}(y) - \int_{\tilde{x}}^{y} \hat{\mu}(x,t)]^2 \tag{3.5.5}$$

where $[\tilde{x}, \breve{x}]$ is the range of ages selected for the computations. Throughout, we chose the first of these methods to estimate the optimal mortality level.

As in the case of the standard ogive procedure, the application of GO requires choosing *ex ante* a model within a family of mortality patterns. In our application we choose from among three options: two mortality models from the Coale-Demeny families (West, South) and the Latin American family of the United Nations system (United Nations, 1982). To compare results and assess levels of uncertainty we compute life expectancy at ages 45, 65, 75 associated with the unique solution in each of the models. We then calculate mean and variances over three observations (e.g. three different mortality patterns) for all values of life expectancy and their coefficient of variation. In all cases, estimates of the three life expectancy values associated with models South and West are very similar to each other but of lower magnitude than those from the other model. However, in the worst of cases differences never exceed 8 percent of the median value. The UN model reflects mortality patterns that emerge when the secular mortality is already underway and the original relation between early and adult mortality has already shifted as a consequence of the mortality decline itself. Instead models West and South in the Coale-Demeny families reflect likely to have prevailed at adult ages in LAC before the onset on mortality decline (LAMBdA team, 2020). In all cases we use the average mortality rates associated with the unique solutions from models South and West and Latin American model of the UN system.

[45] After some trial and error we settled for $\tilde{x} = 15$ and $\breve{x} = 70$ as lower and upper upper bounds of the age range within which we search for solutions.

### 3.5.4 Estimation of rates of growth for the application of GO

Rates of growth were computed from two successive censuses. In cases when there is historical evidence of heavy international migration, these rates were adjusted using estimates of net international migration (see below). In all other cases, the observed rates of growth (total and age specific) were further adjusted for relative completeness of censuses. Two different methods were employed to calculate adjusted rates of natural increase:

i. *Splines:* we fit cubic splines to observed rates of growth for the entire period 1900-1970. Rates of population growth from 1950 on are all directly adjusted for differential census completeness whereas, with some exceptions, those for the period before 1950 are not so and require more care.[46]

ii. *Third party estimates*: when available, we used estimates adjusted by third parties provided the adjustments are based on exogenous information about accuracy of census counts rather than estimated from models requiring unverifiable assumptions. Most of these rates were estimated and adjusted by Arriaga'sadjusted estimates (Arriaga, 1968). Rates for Argentina, Brazil, Cuba and Uruguay for the earlier periods were estimated after adjusting for net international migration (see below).

### 3.5.5 Adjusting rates of natural increase for net international migration

Argentina, Brazil, Cuba and Uruguay experienced large flows of international migrants soon after 1850-1860. These lasted until about 1945, albeit with a decline after World War I. Table 3.13 displays sources of total net international migration available for all four countries. If the goal is to implement the GO method, total net migration rates are of only limited utility as they convey no age-specific information. To estimate age-specific net migration flows we use age-specific rates of net migration derived from a standard model pattern from among those proposed by Raymer and Rogers (Raymer and Rogers, 2006). Figure 3.8 displays a standard model pattern of net migration rates and, as an illustration, Figure 3.9 displays estimated age specific male net migration rates consistent with total flows (from sources in Table 3.13) for Argentina during the period 1870-1950. These estimates are then used to adjust rates of natural increase before applying GO.[47]

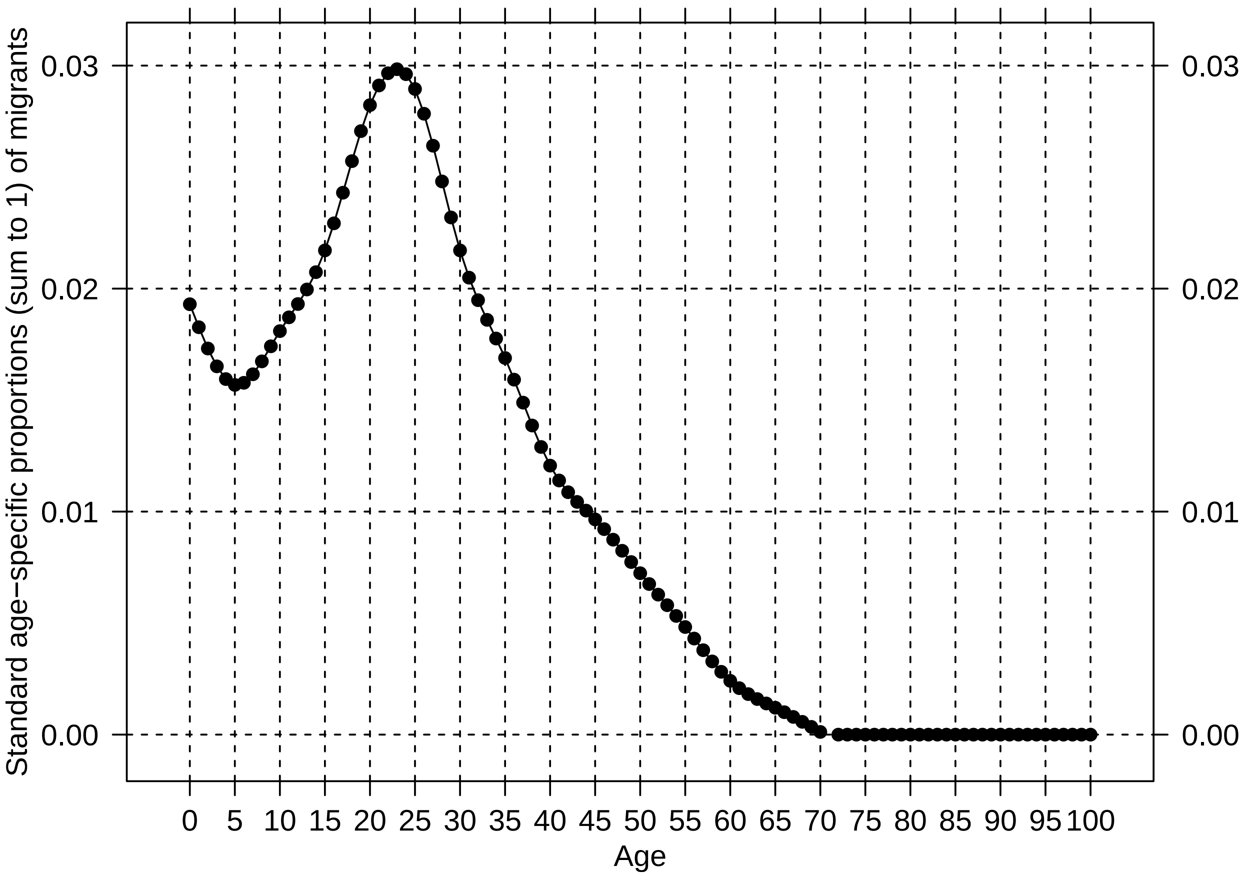
### 3.5.6 Summary of life tables for the period 1850-1950

Table 3.14 displays country-years in the interval 1850-1950 and the methods used to estimate life tables in each case. The four methods are: generalized ogive (GO), GO with migration-

[46] The exceptions are country-years where we could apply Martin's variant of Brass's method (see above) or in countries where we had to adjust for international migration (Argentina, Brazil, Cuba, Uruguay).

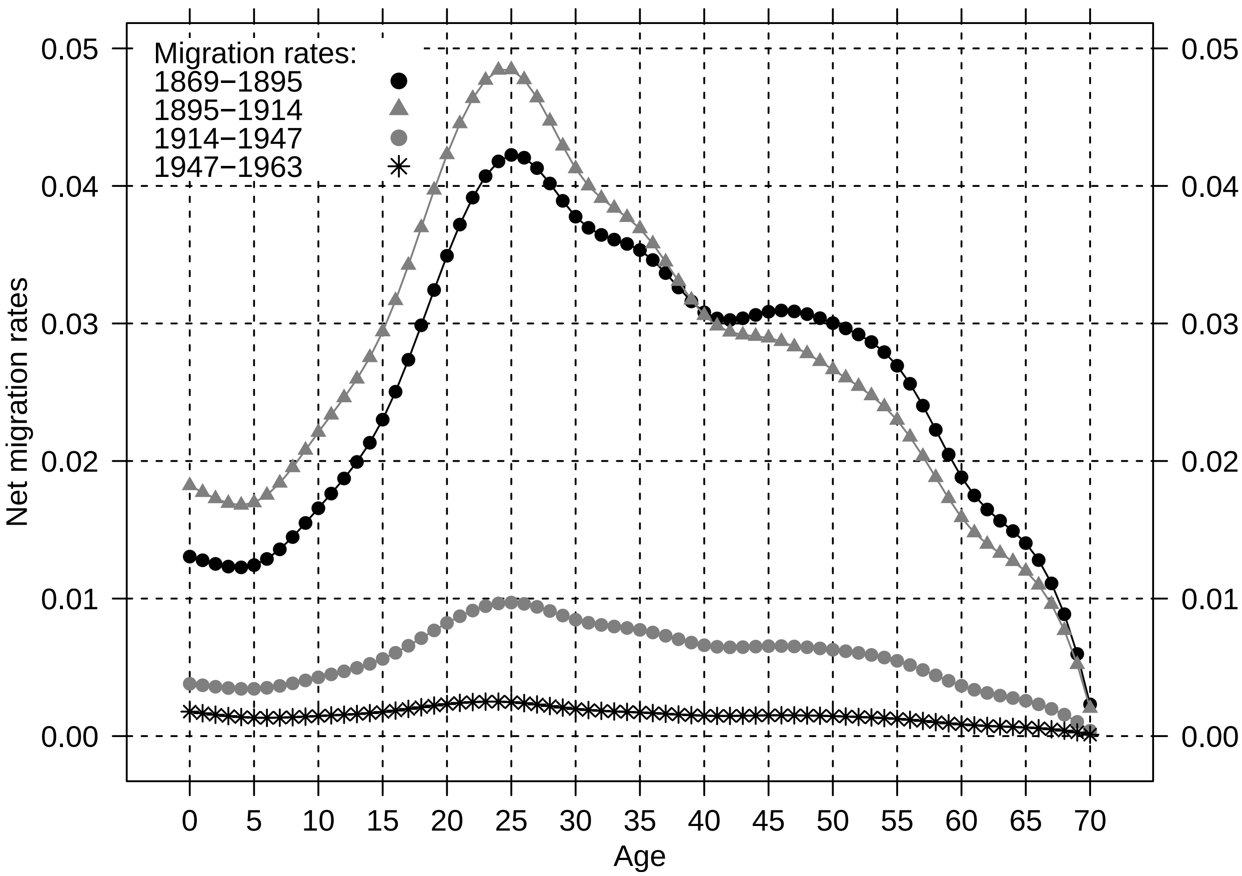
[47] There are, of course, multiple standard age patterns of net migration that could have been used. However, experimentation with alternatives ones that preserve a plausible shape (high rates at youngest ages, increasing at ages of heavy labor force participation and decreasing at older ages) yield very similar estimates of mortality levels when the GO method is applied. The results are largely insensitive to alternative age patterns of migration because the generalized ogive rests on computations of *cumulative* quantities.

Figure 3.8: Standard density function for net international migration.




Source: Fitted model to Raymer-Rogers data (2006).

Figure 3.9: Estimated net international migration rates. Argentina-Males




Source: see text

Table 3.13: Sources of estimates of international migration: 1860-1960.

| Country | Source | 1870-1900 | 1900-1914 | 1914-1950 | 1950+ |
|---|---|---|---|---|---|
| Argentina | | | | | |
| | Somoza | √ | √ | NA | NA |
| | Collver | √ | √ | √ | √ |
| | Sanchez-Albornoz | √ | √ | √ | NA |
| Brazil | | | | | |
| | Collver | √ | √ | √ | √ |
| | Sanchez-Albornoz | √ | √ | √ | NA |
| | Mortara | √ | √ | √ | NA |
| | Ferreira-Levy | √ | √ | √ | NA |
| Cuba | | | | | |
| | Collver | √ | √ | √ | NA |
| | Secretaria Hacienda | √ | √ | √ | NA |
| Uruguay | | | | | |
| | Anuarios estadisiticos | √ | √ | √ | NA |

Sources by periods: (a) Somoza (1971); (b) Collver (1965); (c) Sanchez-Albornoz (1976); (d) Mortara (1954); (e) Ferreira-Levy (1974); (f) Secretaria de Hacienda, Republica de Cuba (1976); (g) Anuarios estadísticos, Republica del Uruguay (1976).

adjusted rates of growth, Martin's variant of Brass's method (Martin) and, finally, two-stage Bennett-Horiuchi (Bennet-Horiuchi Ia-IVa). After application of Martin and the four variantes of Bennett-Horicuchi we further adjust adult mortality for net age over-reporting using the earliest estimates of net overstatement for the country in the post 1950 period.[48]

## 3.6 Estimation of child mortality

The life tables estimated so far are of limited use as they only inform over 5 mortality experiences. Documentation of a complete history of mortality decline in the region should include estimates of mortality below age 5, a task that demands different data and techniques. We now describe the production of estimates of child mortality below age 5 ($_5Q_0$) for the period after 1950.[49] Most of these estimates are based on mixed methods, including indirect techniques with data from World Fertility Surveys (WFS), Demographic and Health Surveys (DHS) and Multiple Indicator Cluster Surveys (MICS), birth histories from surveys, and direct assessments with adjusted vital statistics. Different methods are required for the estimation of under-five mortality for the period 1850-1950, when vital statistics are either unavailable, unreliable, or the data to compute indirect estimates do no exist. A handful of the estimates we compute for this period are derived from adjusted vital statistics. The majority, however, are calculated under weak assumptions about model mortality patterns that establish consistency with adjusted adult mortality.

[48] Estimates of adult mortality retrieved from the application of GO do not require adjustments for net age overstatement as the estimates of mortality rates come directly from Model mortality patterns.

[49] A more thorough rendition of the methods used to estimate infant and under 5 mortality is in Chapter 4.

Table 3.14: Countries by Method of Life Table Estimation for 1850-1950.

| Country | Generalized Ogive (GO) | GO with migration (GOm) | BMartin | 2SBH_4 |
|---|---|---|---|---|
| Argentina | | 1882, 1904 | 1914 | |
| Bolivia | 1925, 1950 | | | |
| Brazil | | 1881, 1895, 1900, 1910 | | |
| Chile | 1859, 1870, 1880, 1890, 1901, 1913 | | | 1925, 1935, 1946 |
| Colombia | 1908, 1915, 1923, 1933 | | | 1944 |
| Costa Rica | 1873, 1887, 1909 | | 1927 | 1938 |
| Cuba | | 1851, 1869, 1882, 1893, 1903, 1913 | | 1925, 1937, 1948 |
| Dominican Republic | 1927 | | | 1942 |
| Ecuador | | | | |
| El Salvador | | | | 1940 |
| Guatemala | 1886, 1907, 1930 | | | 1945 |
| Honduras | 1932, 1937 | | | 1942 |
| Mexico | 1897, 1905, 1915 | | 1921 | 1925, 1935, 1945 |
| Nicaragua | | | | 1945 |
| Panama | 1915, 1925, 1935 | | | 1945 |
| Paraguay | | | | |
| Peru | 1908 | | | |
| Uruguay | | 1904 | 1908 | |
| Venezuela | 1931 | | | 1938, 1945 |

Martin: Martin's variant of Brass's method for quasi-stable populations; two-stage Bennett-Horiuchi (Bennet-HoriuchiI Ia-IVa.

### 3.6.1 Child mortality after 1950

The goal is to estimate infant mortality or the probability of dying during the first year of life, ${}_1Q_0 = 1 - \exp(-\int_0^1 \mu(y)dy)$, early child mortality or the conditional probability of dying between the first and fifth birthday, ${}_4Q_1 = 1 - \exp(-\int_1^5 \mu(y)dy)$, and childhood mortality or the probability of dying before age 5, ${}_5Q_0 = 1 - \exp(-\int_0^5 \mu(y)dy)$. We use three separate sources of information: (i) vital registration including births and deaths as well as census counts below age 5 by single years of age , ii) survey data with birth histories and reports of children ever born and surviving to mothers aged 15-49, and iii) microcensus samples with requisite information on children ever born and surviving by maternal age. The first source of data is the basis for *direct estimates* computed with adjustments whereas the second and third sources are the basis for *indirect estimates*.

These data will be combined to handle three classes of country-years observations. The first are countries with no or highly erratic vital records but suitable data on birth histories and/or children ever born and surviving from either surveys or microcensuses. The second class contains country-years with vital records and survey or census information on children ever born and surviving. The third class includes country-years with (potentially adjustable) vital records but no survey or census information on children born and surviving. The rules followed to compute alternative estimates of parameters of interest are *per force* different in each of these three classes of countries and are dictated by the nature of indirect and direct estimates. We review these below.

Table 3.15: Availability of information to compute indirect estimates of child mortality.

| Country | Census | WFS | DHS | MICS | Vital |
|---|---|---|---|---|---|
| Argentina | 1970, 1980, 1991, 2001 | | | | 1966-2010 |
| Bolivia | 1976, 1992, 2001 | | 1989, 1994, 1998, 2003, 2008 | 2000 | |
| Brazil | 1970, 1980, 1991, 2000 | | 1986, 1996 | | |
| Chile | 1970, 1982, 1992, 2002 | | | | 1955-2009 |
| Colombia | 1973, 1985 | 1976 | 1986, 1990, 1995, 2000, 2005, 2009 | | |
| Costa Rica | 1973, 1984, 2000 | 1976 | | | 1956-2010 |
| Cuba | 1981 | | | | 1964-2010 |
| Dominican Rep | 1970, 1981, 2002 | 1975, 1980 | 1986, 1991, 1996, 1999, 2002, 2007 | 2006 | |
| Ecuador | 1974, 1982, 1990, 2001, 2010 | 1979 | 1987 | | |
| El Salvador | 1971, 1992, 2007 | | 1985 | 1992, 1993 | |
| Guatemala | 1973, 1981, 2002 | | 1987, 1995, 1999 | | |
| Honduras | 1974, 1988, 2001 | | 2005 | | |
| Mexico | 1980, 1990, 2000, 2005, 2010 | 1976 | 1987 | | |
| Nicaragua | 1971, 1995, 2005 | | 1998, 2001, 2006 | | |
| Panama | 1980, 1990, 2000 | 1975, 1976 | | | 1955-2009 |
| Paraguay | 1972, 1982, 1992, 2002 | 1977, 1979 | 1990 | | |
| Peru | 1972, 1981, 1993, 2007 | 1978 | 1986, 1992, 1996, 2000, 2003, 2004, 2005 | | |
| Uruguay | 1975, 1985, 1996 | | | | 1955-2009 |
| Venezuela | 1981, 1990, 2001, 2011 | 1977 | | | 1955-2007 |

Note: World Fertility Surveys (WFS), Demographic Health Surveys (DHS), Multiple Indicator Cluster Surveys (MICS) and microcensus samples.

### Indirect estimates

Indirect methods were implemented in country-years with surveys (DHS, WFS, MICS) and census (microdata from census samples)[50] on children ever born and surviving by mother's age. The list of country-years with suitable information is in Table 3.15. In addition, WFS and DHS data includes complete (WFS) or partial (DHS) maternal birth histories from which it is possible to compute estimates of single year of age probabilities of dying between birth and age 5. To facilitate identification we consider these as indirect estimates also, even though their genesis is quite different from standard indirect estimates. There are then three types of estimates:

i. *Standard indirect estimates*: These were computed following two different methodologies. The first is the classic Brass technique augmented with estimates of time reference associated with the conditional probabilities of dying for women in the age groups 20-24, 25-29 and 30-34 (Brass and Coale, 1968; Trussell, 1975). Throughout we assume models West and South from the Coale-Demeny families of mortality patterns. The second methodology is based on the United Nations life tables also augmented with estimates of time references (Palloni and Heligman, 1985). Throughout we assume the Latin American model from the United Nations families of mortality patterns (United Nations, 1982).

ii . *Non-standard indirect estimates from birth histories*: These were computed directly

[50] We used census microdata directly released to us by CELADE. In some cases we complemented this with standardized files from the IPUMS project.

from mothers' reports on children's dates of birth and death within the first five years before the survey in WFS and DHS. We use conventional events-exposure ratios in single years of age as estimates of single year death risks and the exponentiated (negative) rates as estimates of conditional probabilities of surviving inone year segments.

**Direct estimates from adjusted vital statistics**

In most cases, available information on births and child deaths during this period is defective due to incomplete coverage of both counts and cannot be used without adjustments. We compute adjusted estimates of ${}_1Q_0$, ${}_4Q_1$ and ${}_5Q_0$ from two sources. The first source originates in third party figures that include estimates from different agencies using fully documented adjustments for completeness of death, birth and death registration and whose suitableness can be evaluated. In a number of cases these estimates were computed by a country's statistical offices and, in others, they are produced by international organizations. An important fraction of these adjustments originate in contrasts between raw figures from vital statistics and indirect estimates from one or several methodologies identified before.

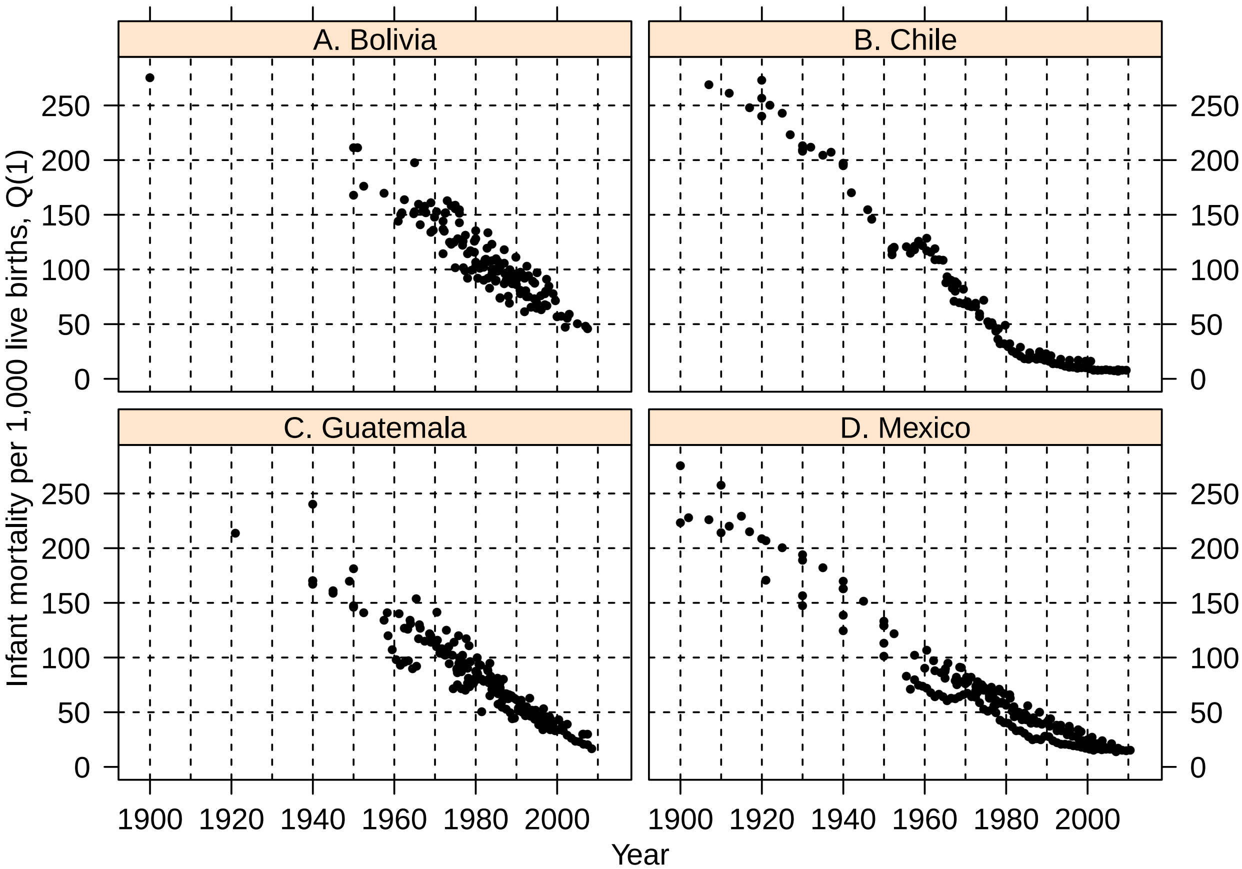
The second source of adjusted estimates is the product of reconciliation between estimates computed directly from vital statistics and estimates from indirect methods. Following conventional rules we compute single year probabilities of dying from vital statistics and then estimate adjustment factors to make them consistent with estimates from standard and non-standard indirect estimates. To compute infant and child mortality from vital statistics we track registered births cohorts by year of birth and then match them to registered deaths in single years of age. In all cases we used conventional Lexis diagrams and separation factors to allocate deaths in a calendar year to a birth cohort and obtain cohort based estimates. In addition, we compute period-based estimates of the same probabilities. We average period and cohort estimates and compute adjustment factors that make them consistent with standard and non standard indirect estimates. These factors are then used to compute adjusted quantities during periods not longer than 10 years and located in the neighborhood of the years from which we are able to compute adjustment factors. Furthermore, we estimate time trends of adjustment factors, a useful tool for the computation of parameters during 1900-1950 when indirect estimates are unavailable.[51]

**Estimation from pooled direct and indirect estimates**

Pooling together the set of indirect and direct (adjusted) estimates for each country results in a time series, with possibly multiple observations for each time point, straddling the period 1950-2010. The country-specific pooled data contain non-independent observations with some degree of redundancy, a desirable property when computations by different agents follow different conventions. Each estimate, however, also provides different, possibly erroneous, information. The errors are caused by inappropriate choice of mortality models, reporting errors in the surveys or censuses (of children born or dead, of maternal age), or inaccurate dating of events(birth histories and vital statistics). Thus, despite redundancies,

[51]In all cases the adjustment factors we compute using vital statistics (births and deaths) and census counts (populations at ages 0 and 1-4) are relative in the sense that they reflect errors both in numerators and denominators of rates.

Figure 3.10: Pooled set of infant mortality rates, ${}_1Q_0$, in four countries.




the pooled set for a single time point is richer and more informative about the true value of parameters that any single estimate for the same time point. If anything, the pooled set reflects measurable uncertainty that can be informative in statistical analyses. A single point estimate chosen arbitrarily from available estimates conveys no such information and conceals the uncertainty surrounding its genesis. Figure 3.10 displays the pooled set of ${}_1Q_0$ in four countries.

To demonstrate the detection of consistent time trends we also include estimates from adjusted vital statistics and from third party sources for years before 1950. The clouds of estimates are dense for years after 1950 and thin out between 1900 and 1950. The estimates are tightly clustered around narrow bands following smooth trajectories with some critical points of acceleration and deceleration, suggesting unequal rates of change over time. For the purposes of producing LAMBdA's pivotal life tables we identify only one point estimate per year. We do this for each year after 1950 by fitting country-specific splines of the following form

$$\texttt{logit}({}_kQ_x) = \beta_0 + \sum_{n=1}^{n=j} \beta_n T_n \tag{3.6.1}$$

where ${}_kQ_x$ stands for one of the three measures of child mortality defined before, $T_n$ are nodes at predefined years from 1950 to 2010 in jumps of five and $\beta_n$ are associated shifts of the spline at those nodes. Year-specific predicted values from the fitted splines are then selected as estimates for that year.[52]

### 3.6.2 Child mortality for the period 1850-1950

With only a few exceptions vital statistics for this period are either unavailable or incomplete (with regional rather than national coverage). To estimate child mortality we employ a strategy that rests on three principles. First, when available we use adjusted adult mortality rates to determine childhood mortality consistent with a handful of selected mortality patterns, e.g believed to fit well the mortality experience during this period. Second, we include third party estimates if and when the methods employed are thoroughly justified and one is able to assess their validity. Third, for countries with vital statistics with national, but perhaps deficient, coverage we generate adjusted estimates using time trends of adjustment factors for the period 1950-2010 (see section above).

Implementation of these three principles yields at least one and in most cases between 2 to 5 different estimates of the target parameters.[53] In all cases the series of estimates stretches back to 1905 form a cloud of points that either moves slowly upwards as it gets closer to the year 1900 or attains a ceiling earlier than 1900. Only in the cases of Argentina, Brazil, Cuba and Uruguay are we able to compute estimates of childhood mortality from adult mortality before 1900. In most other cases we can only verify time trends up to 1900 (a handful of countries) or up to 1920 (most countries).

With the exception of Argentina, Cuba and Uruguay,the mortality decline could not have begun in earnest before 1900. This suggests a procedure to estimate childhood mortality for the missing years: we join the pool of estimates for the period before and after 1950 and fit a country-specific, three-parameter Gompertz function to the entire cloud of point estimates. The function is

$$Q_i = \beta_1 \exp(-\exp(-\beta_2(t_i - \beta_3))) + \varepsilon_i \tag{3.6.2}$$

where $Q_i$ is either ${}_1Q_0$ or ${}_5Q_0$, $t_i$ is the year of estimation (with origin in 1900) and $\beta$'s are parameters. The function, a variant of a standard logistic function, provides two pieces of information: first, a point estimate for each year within the range of years where countries contribute with at least one estimate that is consistent with the post-1950 trend and, second, an estimate of the time trend threshold value that serves as an estimator for levels in the distant past.

[52] In most cases we are able to compute all three indicators of child mortality. However, during this period of time we can occasionally compute ${}_1Q_0$ and ${}_5Q_0$. In these cases ${}_4Q_1$ was defined as ${}_4Q_1 = ({}_5Q_0 -_1 Q_0)/(1 -_1 Q_0)$ .

[53] Because in some countries we chose three model mortality patterns to estimate adult with the GO method, those countries will be endowed with three alternative estimates of the parameters of interest.

Table 3.16: Classification of country years by sources used to estimate ${}_5Q_0$.

| Country | Model mortality | Vital adjusted | Third party |
|---|---|---|---|
| Argentina | 1869, 1882, 1895, 1904, 1914, 1930, 1947 | 1914, 1947 | 1882, 1904, 1914, 1947 |
| Bolivia | 1900, 1925 | 1900, 1950 | 1950 |
| Brazil | 1872, 1881, 1890, 1895, 1900, 1910, 1920, 1930, 1940, 1945 | 1940 | 1940 |
| Chile | 1854, 1859, 1865, 1870, 1875, 1880, 1885, 1890, 1895, 1901, 1907, 1913, 1920, 1925, 1930, 1935,1940, 1946 | 1920, 1930, 1940 | 1920, 1930, 1935, 1940, 1946 |
| Colombia | 1905, 1908, 1912, 1915, 1918, 1923, 1928, 1933, 1938, 1944 | 1938 | 1938, 1944 |
| Costa Rica | 1864, 1873, 1883, 1887, 1892, 1909, 1927, 1938 | 1927, 1938 | 1927, 1938 |
| Cuba | 1841, 1851,1861,1869, 1877, 1882, 1887, 1893, 1899, 1903, 1907, 1913, 1919, 1925, 1931, 1937, 1943, 1948 | 1927, 1938. 1943 | 1925, 1937, 1948 |
| Dominican Rep | 1920, 1927, 1935, 1942 | 1935, 1942 | 1935, 1942 |
| Ecuador | | | |
| El Salvador | 1930, 1940 | 1930, 1940 | 1930, 1940 |
| Guatemala | 1880, 1886, 1893, 1907, 1921, 1930, 1940, 1945 | 1940, 1945 | 1940, 1945, 1949 |
| Honduras | 1930, 1932, 1935, 1937, 1940, 1942, 1945, 1947 | 1940, 1945 | 1940, 1945 |
| Mexico | 1895, 1897, 1900, 1905, 1910, 1915, 1921, 1925, 1930, 1935, 1940, 1945 | 1921, 1925, 1930, 1935, 1940, 1945 | 1940, 1945 |
| Nicaragua | 1940, 1945 | 1940, 1945 | 1940, 1945 |
| Panama | 1911, 1915, 1920, 1925, 1930, 1935, 1940, 1945 | 1940, 1945 | 1940, 1945 |
| Paraguay | | | |
| Peru | 1876, 1908, 1940 | | |
| Uruguay | 1900, 1904, 1908, 1935 | 1908, 1935 | 1908, 1935 |
| Venezuela | 1926, 1931, 1936, 1938, 1941, 1945 | 1938, 1941, 1945 | 1938, 1941, 1945 |

# Chapter 4

# Estimation of Infant and Child Mortality in LAMBdA

## 4.1 Introduction

LAMBdA's life tables were estimated using separate methods to adjust infant (under age 1) and child mortality (between ages 1 and 5), on one hand, and adult mortality (ages 5 and over). This strategy is consistent with the one pursued in the work underlying the United Nations Life Tables (United Nations, 1982). It responds to the need to generate age mortality patterns that reflect relations between child and adult mortality avoiding imposition of *ex ante*, assumed relations, through the use of model patterns that may represent well age-gender patterns of mortality in some populations but only poorly or not at all in others. LAMBdA is as free of model pattern assumptions as possible given the information available to us (see Chapter 3).

The estimation of child mortality (under age 5) follows three principles. The first is to work with a natural constraint that draws a hard separation between the period before and after 1950.[1] As we discuss below, estimation procedures for the period 1850-1950 are qualitatively different from those used in the period after 1950. The difference is largely due to lack of minimally reliable vital statistics and the absence of secondary sources, such as household surveys, to obtain indirect estimates of child mortality for the period before 1950. The second principle is to utilize as many credible sources of information as possible for each unit of observation (country-year), including vital statistics and population censuses after suitable adjustments. The third is to exploit the existence of multiple estimates for each observation to generate robust point estimates. As discussed below, it is only in a very few cases that we are able to explicitly choose a unique value from among those available. In most cases, the selected point estimate for a country-year corresponds to a function of the set of alternative estimates and these may sometime include values for neighboring years.

[1] As mentioned in Chapter 2 we use the expression "before 1950" somewhat loosely to encompass years in the interval 1950-1964. This facilitate reference to periods that, for some countries, may not literally refer to years before 1950. For most countries though, "before 1950" has a literal meaning.

## 4.2 Definitions

Mortality among young children (younger than 5) can be broken down by exact age at death to produce data compatible with abridged life table and international standard measures of young mortality (infant, child, and under-five). Infant mortality, ${}_1Q_0$, includes deaths between birth and exact age 1; child mortality, ${}_4Q_1$, includes deaths from age 1 to 4 years; and under-five mortality, ${}_5Q_0$, includes deaths between birth and exact age five. ${}_1Q_0$ is the probability that a child born in a particular year will die before reaching age 1, if subject to current age mortality rates. It is usually expressed in 1,000 live births. Similarly, ${}_5Q_0$ is the probability that a child born in a year will die before reaching 5 years if subject to current age mortality rates. It is also expressed in 1,000 live births. The probability of dying between exact ages 1 and 4 years is computed algebraically from ${}_1Q_0$ and ${}_5Q_0$ as follows:

$$ {}_4Q_1 = \frac{{}_5Q_0 -_1 Q_0}{1 -_1 Q_0} \tag{4.2.1}$$

## 4.3 Sources

Although the quality of vital statistics in LAC has improved steadily since 1950, there is sizeable time and country heterogeneity in the quality (accuracy, reliability, and timeliness) of these data (see Chapters 2 and 9). Our objective is to assemble a large set of estimates as accurate as possible using multiple and fully or partially independent sources of information. These are primarily recorded vital statistics, population censuses, household surveys of various types and alternative estimates elaborated by various institutions (UNPD/CELADE, Statistics National Offices, and individual researchers).

The period before 1950 presents us, however, with a qualitatively different challenge as few countries possess any information from official vital statistics, regular censuses and, least of all, household surveys. To compute estimates for the period 1850-1950 we rely on data from three different sources: adjusted vital statistics (births, deaths) and population censuses, third party estimates based on replicable algorithms, and our own estimates based on generalized stable populations that use mortality patterns from the Coale-Demeny and United Nations mortality models (see Chapter 3). It is only in this latter situation and for a handful of country-years within the period 1850-1920 that we rely on external models of mortality. But even in these cases we do not select a final estimates of child mortality from the models themselves but rather consider all of them as plausible estimates of the same population parameter.

## 4.4 Estimates for the period 1950-2010

For the period after 1950 infant and child mortality rates were computed combining information from vital statistics and population censuses in countries with good quality data (Argentina, Chile, Costa Rica, Cuba, Panama, Uruguay, and Venezuela). In addition, when available for any country, we computed Brass type indirect estimates and direct estimates from birth histories retrieved from either established surveys (WFS and DHS) from specialized, country-specific,and Multiple Indicator Cluster Surveys (MICS) technical reports. We also generated indirect estimates using available censuses microdata released by CELADE to us and complemented with available standardized files from the IPUMS Project. Finally, we resorted to published studies and third-party life tables some of which used the same sources we employ to compute our own estimates. The result is a large set of alternative estimates each composed of multiple (usually more than three) estimates per country-year. What follows is a comprehensive list of sources used to calculate estimates for the period 1950-2010.

### 4.4.1 Microsamples of national population censuses

Every country in the region has conducted population censuses on regular basis at least since 1960, mostly at ten-year intervals. Their large scale, national coverage and inclusion of suitable items in the household questionnaire make them appropriate for estimating infant and child mortality in the absence of adequate vital statistics.

Beginning in 1970 population censuses main schedules include a set of questions related to womens' reproductive outcomes, such as number of children ever born and surviving. This information is used to estimate infant and child mortality based on the so-called indirect methods, also known as Brass Methods. Although census data are not subject to sampling errors[2], they do not always have full population coverage and the missing individuals could be clustered in areas where the resident population is exposed to selectively high (low) mortality. In addition, some census information is always subject to errors, including maternal age. Finally, at least part of the data on children ever born and surviving is oftentimes subject to recall problems that cause systematic biases in the estimates of child mortality. For example, omission of children born who died soon after birth is characteristic among older women and causes estimates of probabilities of dying to be biased downward. Despite these weaknesses, Brass type of estimates have proven to be highly robust under a very wide set of conditions and in populations with sharply different levels of economic well-being, social and political systems, and cultural backgrounds. Throughout, a paramount concern was to attain consistency of estimates that we sought to secure using redundant information. Whenever two or more successive censuses (or surveys) were available, estimates associated with each of them would frequently overlap for some years. Significant differences between these estimates signaled errors in one or more of the sources and required us to screee each of them and discard those that were inconsistent with time trends or with levels of adult mortality. Consistency of estimates, on the other hand, was desirable although not an impregnable criteria for perfectly consistent estimates could also result from perfectly consistent errors in the sources.

Table 4.1 displays censuses used to retrieve estimates of infant and child mortality via Brass indirect methods. The basic source for the estimates were the microcensus samples for each country and the pertinent information on children ever born and surviving for each household.

We applied standard variants of the Brass technique to estimate child mortality over a

[2]Of course, this only applies to official country aggregate figures, not to quantities computed from micro census samples.

Table 4.1: Censuses used to compute indirect estimates

| Censuses used to compute indirect estimates | |
|---|---|
| COUNTRY | CENSUS YEARS |
| Argentina | 1970, 1980, 1991, 2001 |
| Bolivia | 1976, 1992, 2001, 2012 |
| Brazil | 1970, 1980, 1991, 2000, 2010 |
| Chile | 1970, 1982, 1992, 2002 |
| Colombia | 1973, 1985 |
| Costa Rica | 1973, 1984, 2000, 2011 |
| Cuba | 1981 |
| Dominican Republic | 1970, 1981, 2002, 2010 |
| Ecuador | 1974, 1982, 1990, 2001, 2010 |
| El Salvador | 1971, 1992, 2007 |
| Guatemala | 1973, 1981, 2002 |
| Honduras | 1974, 1988, 2001 |
| Mexico | 1980, 1990, 2000, 2005, 2010 |
| Nicaragua | 1971, 1995, 2005 |
| Panama | 1980, 1990, 2000, 2010 |
| Paraguay | 1972, 1982, 1992, 2002 |
| Peru | 1972, 1981, 1993, 2007 |
| Uruguay | 1975, 1985, 1996, 2010 |
| Venezuela | 1974, 1982, 1990, 2001, 2010 |

period of 5 to 10 years before the census. In all cases we excluded information about mothers aged 15-19 and 45-49 and supported estimation using standard multipliers. The censuses also generated information on population counts in the age groups 0 and 1-5 which, when combined with yearly births and observed death counts, enable us to compute unadjusted estimates of ${}_1Q_0$ and ${}_4Q_1$ (see below).

Brass type mortality estimates cannot be obtained without assuming an underlying mortality pattern. To reduce errors due to misindentification of the mortality pattern we use the Latin American mortality model which is based on adjusted data for the period 1950-1980 and the corresponding Brass multipliers (Palloni and Heligman, 1985). As additional insurance against biases, and to explicitly incorporate a measure of uncertainty, we also considered estimates computed using multipliers based on the South and West models in the Coale-Demeny life table system (Trussell, 1975). Finally, and whenever the quality of the vital statistics and census counts permitted, we computed probabilities of child mortality directly from raw information and added these to the set of permissible estimates for a particular country-year (see below).

### 4.4.2 Indirect estimates from household surveys

Starting in early 1970s most countries of the LAC region fielded a large number of household surveys, all with different objectives, sponsoring institutions, and executing agencies. While not all were directed at assessing mortality, most of them became vehicles for the introduction of simple modules that elicited easy-to-get information which could then be transformed into estimates of child mortality. These modules included items such as mothers' total number of children born and surviving as well as their partial or complete birth (or pregnancy) history. As a result, the surveys constitute the most important and sometimes the only– source of information on infant and child mortality and fill in the niche left by imperfect, albeit improving, national systems of vital statistics.

The quality of data on mortality levels and trends retrieved from these surveys depends on their design, nature of field work, and precision of central office operations (robustness of consistency checks, verification strategies, and methods for data cleaning, etc...). It is also a function of the quality of the two types of data elicited, namely, birth (pregnancy) histories that yield direct estimates and, secondly, retrospective recall of children ever born and surviving children.

All household surveys we use to generate mortality estimates share problems with microcensus data, including potential lack of population coverage, sampling errors, erroneous declaration of ages, occurrence and timing of events, and under (over) reporting of events. We use two tools to handle systematical and random errors. First, when a country fields multiple surveys that are close together it is theoretically possible to use redundant information to establish consistency, albeit not validity, of estimates for a period during which there is overlap. Second, to attenuate the influence of noise we compute estimates for periods of 5 to 10 years preceding the surveys.

We include four types of household surveys: World Fertility Surveys (WFS), Demographic and Health Surveys (DHS), Multiple Indicator Cluster Surveys (MICS), and other available national surveys focuses on Contraception, Health, Household Budgets, etc.

### World Fertility Surveys

Table 4.2 displays the country-years with available data from the WFS that supported calculation of direct and indirect estimates of infant and child mortality. Two types of information included in WFS were used. First, tabulations of mothers by age, children ever born and children surviving as input for the computation of Brass's indirect estimates. Second, we generate direct estimates using birth and pregnancy histories elicited from all mothers aged 20-44. These were joined with indirect estimates and, if referring to the same year or window of time, both sets were retained as alternative estimates for a single population parameter.

### Demographic and Health Surveys

Table 4.3 displays the country and years for which we retrieved information from DHS. As in the case of WFS estimates of infant and child mortality, these were computed using two sources, namely, tabulations of mothers by age and children ever born and surviving and birth histories. The latter are more limited in DHS than in WFS but enable us to support computation of infant and child mortality rates covering a period of five years before the surveys. As in the case of WFS, we assembled indirect and direct estimates, assessed their consistency and in most cases include both sets in the country-years database. Whenever inconsistencies were identified we discarded the estimates that were not consistent either with a detectable time trend of child mortality or with our own estimates of mortality above age 5 for the same period. In the few cases when we had to apply the rule we always retained

Table 4.2: World Fertility Surveys used to compute indirect estimates

| World Fertility Surveys | |
|---|---|
| COUNTRY | WFS YEARS |
| Argentina | NA |
| Bolivia | NA |
| Brazil | NA |
| Chile | NA |
| Colombia | 1976 |
| Costa Rica | 1976 |
| Cuba | NA |
| Dominican Republic | 1980 |
| Ecuador | 1979 |
| El Salvador | NA |
| Guatemala | NA |
| Honduras | NA |
| Mexico | 1976 |
| Nicaragua | NA |
| Panama | 1976 |
| Paraguay | 1979 |
| Peru | 1978 |
| Uruguay | NA |
| Venezuela | 1977 |

the indirect estimate. For the most part we almost retained the indirect estimates. In the bulk of cases where we discarde or chose the indirect estimates. It is important to note that DHS surveys stretch over a period that starts at the earliest five to ten years after WFS. As a consequence there is no overlap of estimates that can support consistency tests. However, since infant and child mortality were declining rapidly during the period 1975-2000 in all countries of the region, one can use the time trends inferred from WFS and DHS estimates to check consistency. Both sets of estimates were included in the data base and no decision was made about using one or the other in case of inconsistencies. Instead, we employed a modeling strategy to conciliate both sets of indirect estimates as well as point estimates for neighboring years derived from vital statistics or other sources).

### Multiple Indicator Cluster Surveys

The Multiple Indicator Cluster Surveys (MICS) program consists of various surveys rounds carried out in developing countries under the technical supervision and funding of the United Nations Children's Fund (UNICEF). The goal is to provide high quality data on the situation of women and children in low- and middle-income countries. All six rounds were implemented to monitor a large number of indicators measuring the progress of various international commitments (World Summit for Children, Millennium Development Goals, and Sustainable Development Goals) for improving childrens health and social situation. The MICS data are comparable to the DHS. Data sets include womens reproductive history, childrens health and development, and households characteristics. As is the case of DHS, this information was used to estimate early child mortality using indirect and direct methods of estimation. Table 4.4 displays countries for the MICS we include in our database.

Table 4.3: Demographic and Health Surveys used to compute indirect estimates

| Demographic and Health Surveys(DHS) | |
|---|---|
| COUNTRY | DHS |
| Argentina | NA |
| Bolivia | 1989, 1994, 1998, 2003, 2008 |
| Brazil | 1986, 1994 |
| Chile | NA |
| Colombia | 1986, 1990, 1995, 2000, 2005, 2009, 2010 |
| Costa Rica | NA |
| Cuba | NA |
| Dominican Republic | 1986,1991,1996,1999,2002,2007,2013 |
| Ecuador | 1987 |
| El Salvador | 1985 |
| Guatemala | 1987,1995,1999 |
| Honduras | 2005,2011 |
| Mexico | NA |
| Nicaragua | 1998,2001,2006 |
| Panama | NA |
| Paraguay | 1990 |
| Peru | 1986,1992,1996,2000,2004,2005,2009,2010-2014 |
| Uruguay | NA |
| Venezuela | NA |

## Other surveys

Table 4.5 lists alternative surveys that contain either information on children ever born or birth histories. While these surveys are not part of a dedicated cross-country data collection effort as are WFS, DHS, and MICS, they were implemented in selected countries under the sponsorship of national governments and/or international organizations. Most of these surveys are USAID-funded, were carried out with CDC's technical assistance and their purpose was to monitor child health and mortality, reproductive health and contraceptive prevalence. First, there are Contraceptive Prevalence Surveys (CPS) that gathered basic information on children ever born and surviving children that enables us to compute indirect estimates of child mortality. These include surveys in Colombia (1978), Costa Rica (1978, 1981), Dominican Republic (1983), Mexico (1979), and Peru (1981). Later, these surveys were broadened to incorporate child and maternal health issues and were renamed the Maternal and Child Health/ Family Planning Surveys (MCH/FP). They were carried out in Guatemala (2008) and Honduras (1984). During the late 1980s, reproductive health were included in the Reproductive and Health Surveys (RHS), which were adapted to the data needs in every country. The topics covered are family planning, maternal and child health, infant and child mortality, anthropometric measures, immunization, sexual health, HIV/AIDS, and health care practices, some countries included birth histories. These surveys were carried out in countries like Costa Rica (1986), El Salvador (1998, 2002, 2008), Guatemala (2002), Nicaragua (1992), and Paraguay (1995, 2004, 2008).

Table 4.4: Multiple Indicators Cluster Surveys used to compute indirect estimates

| Multiple Indicator Cluster Surveys | |
|---|---|
| COUNTRY | MICS |
| | |
| Argentina | NA |
| Bolivia | 2000 |
| Brazil | NA |
| Chile | NA |
| Colombia | NA |
| Costa Rica | NA |
| Cuba | NA |
| Dominican Republic | 2006 |
| Ecuador | NA |
| El Salvador | 1993,1992 |
| Guatemala | NA |
| Honduras | NA |
| Mexico | NA |
| Nicaragua | NA |
| Panama | NA |
| Paraguay | NA |
| Peru | NA |
| Uruguay | NA |
| Venezuela | NA |

A different set of surveys were financed exclusively by national governments, or jointly by national governments and international cooperative organizations. Such are the cases of Brazil (1972, 1973, 1976, 1977, 1978, 1984, 2006), Bolivia (1975, 1981, 1988), and Honduras (1970, 1972, 1983), which included information on children ever born and surviving children. Finally, we also consider household surveys aimed at gathering information of economic outcomes, including household budgets and income, labor force participation, and occupations. Some of these also collected information on children ever born and surviving children. Indirect mortality estimates from these surveys were computed for Brazil (1986, 2005, 2007, 2008, 2009), Colombia (1978, 1980, 1986, 2005, 2007, 2008, 2009), Cuba (1974), and Dominican Republic (1978, 1980).

### 4.4.3 Vital Statistics

Even though the systems of vital statistics in LAC countries improved considerably after 1950, there are only a handful of countries in which one could confidently use unadjusted estimates of child mortality even after 1975/80 (Argentina, Chile, Costa Rica, Cuba, Mexico, Panama, Uruguay, and Venezuela). For a handful of countries, however, one can document coverage in the neighborhood of 95 percent for variable periods in the time interval 1950-2010 and, in these cases and these cases only, we include estimates computed from vital statistics (and population counts in census years) as candidate estimates and treat them

Table 4.5: Other surveys used to compute indirect estimates

| Other surveys(*) | |
|---|---|
| COUNTRY | SURVEY-YEAR |
| | |
| Argentina | NA |
| Bolivia | 1975,1988-89 |
| Brazil | 1972-73,1976-78,1984,1986, 2005, 2006-2009 |
| Chile | NA |
| Colombia | 1978,1980 |
| Costa Rica | 1978,1981,1986 |
| Cuba | 1974,1979,1987 |
| Dominican Republic | 1983 |
| Ecuador | 1982,1989,1994,1999,2004 |
| El Salvador | 1973,1992-93,1998,2008 |
| Guatemala | 1978,1987,1989, 2002,2008 |
| Honduras | 1970,1972,1983-84,1987,1991,1996, 2001 |
| Mexico | 1979,1992,2006, 2009 |
| Nicaragua | 1978,1985,1992 |
| Panama | NA |
| Paraguay | 1995, 2004, 2008 |
| Peru | 1974,1976,1981 |
| Uruguay | NA |
| Venezuela | NA |
| (*) The names of surveys listed in this table are in a table in the Appendix | |

as plausible as those derived from birth histories or from indirect procedures. Because for some country-years we are able to compare estimates from vital statistics and from indirect methods and birth histories, it is possible to compute 'adjustment' factors that apply to 5 to 10 year windows of time (see above discussion of time reference of indirect estimates). We then apply the same adjustment factors to correct estimates from vital statistics to years preceding the 5 to 10 year windows. This type of adjustment was used for the period 1975-1990 in Mexico, Panama and Venezuela. In all cases the adjustment factors hovered in the range .94-1.01. Table 4.6 contains a list of country-years for which we were able to compute estimates of child mortality using vital statistics.

### The raw database

All the information collected and estimates pertaining to the period 1950-2012 were assembled and housed in a database containing a cloud of mortality point estimates that includes infant mortality, ${}_1Q_0$, early child mortality, ${}_4Q_1$, and child mortality, ${}_5Q_0$. The value of early child mortality was calculated using the estimates for ${}_5Q_0$ and ${}_1Q_0$ (see equation (4.2.1)). The data base is a simple rectangular array containing the country name, mid-point of the period to which each estimate applies, the estimate, data source, type of estimate, year the

Table 4.6: Country-years and vital statistics used to estimate child mortality since 1950

| Country-years and vital statistics used to estimate child mortality since 1950 | |
|---|---|
| COUNTRY | VITAL STATISTICS YEARS |
| | |
| Argentina | 1966-2012 |
| Bolivia | NA |
| Brazil | Not used |
| Chile | 1952-2012 |
| Colombia | Not used |
| Costa Rica | 1950-2012 |
| Cuba | 1964-2012 |
| Dominican Republic | Not used |
| Ecuador | Not used |
| El Salvador | Not used |
| Guatemala | Not used |
| Honduras | Not used |
| Mexico | 1955-2012 |
| Nicaragua | Not used |
| Panama | 1955-2012 |
| Paraguay | Not used |
| Peru | Not used |
| Uruguay | 1955-2013 |
| Venezuela | 1955-2009 |

data was collected, and source's name. The protocol to obtain estimates described before yields single or multiple estimates for each country-year, and all of them are subject to uncertainty that depends on the quality and amount of data available associated with each set. The next step is to reconcile alternative estimates for each country-year observation derived from multiple data sources and compute"optimal" single year point estimates. In particular, we are interested in values for years contained in the period 1950-2012 that coincide with pivotal years.[3] To generate these unique estimates we use non-parametric and local estimation methods on each country separately.

### Estimation of country-year unique estimates

Our aim is to use the full set of estimates of child mortality for each country. To do so we design a procedure that reconciles different estimates associated with multiple sources, each with its own idiosyncracies, for a long stretch of time. The procedure consists of combining

[3]A pivotal year is the mid-point of an intercensal year for which we are able to compute adjusted adult life tables. See definition in Chapter 3.

two non-parametric techniques. First, and to preserve the underlying shape of each country observed trends, we applied Lowess to the cloud of points formed by pairs of an estimate of child mortality and a calendar year. Lowess is a flexible non-parametric local regression technique that generates estimates for observed points and are depedent on the window or bandwidth of estimation. We opted to generate estimates associated with a bandwith that led to reproduce best the observed points (in a squared error sense) for each country. However, since Lowess generates estimates for calendar years that contribute to estimation but not *for calendar years for which there are no observed estimates*, we resorted to a second technique that yields the "missing" values in the time series. This second technique consists of estimating splines with knots every five years and retrieving the parameters of the splines using estimates generated by Lowess as point of supports. We then use estimates of the spline parameters to predict the values of the child mortality functions for single calendar years. Figure 4.1 displays observed (blue) and predicted (red) values for ${}_1Q_0$ and ${}_4Q_1$ for Guatemala that obtain after application of the two stage procedure.

**Sex-specific estimates: general considerations**

In all cases the two-stage procedure described above could adopt one of two variants. The first is to apply it to estimates of mortality functions corresponding to females and males combined and subsequently derive sex-specific values of the mortality functions. The second variant consists of applying the two stage procedure *separately* to sex-specific mortality estimates and then join the results to generate estimates of mortality for both sexes combined. This variant is riskier since it (a) rests on two separate and independent applications of model fitting and (b) ignores relations between sex-specific mortality functions for a given country-year. By contrast, the first variant relies on only one application of model fitting and offers ample room to establish consistency with observed relations between sex-specific mortality estimates. For this reason alone, all life tables in LAMBdA were generated with sex-specific estimates of infant and child mortality derived from the application of the first variant of the two stage strategy. Below we describe the stages of its implementation.

First, the two-stage procedure described above was applied to obtain estimates of ${}_1Q_0$ and ${}_5Q_0$ for males and females *combined*.[4] To generate sex-specific estimates we proceed in two steps. First, in each country we assess the (linear) relation between the logit of (observed) estimates of each of the two mortality functions for both sexes combined and the logit of observed estimates for females. We then use estimates of the country-specific parameters of this linear relations, combine them with final estimates of total mortality in each country-year obtained from the two stage procedure described above, and compute predicted values for the two child mortality functions for females. Second, we use the entire database consisting of all direct and indirect observed sex-specific estimates to estimate an orthogonal regression relating the logit of male and female values of the child mortality functions. We then use the parameters of the orthogonal regression relation, combine them with the predicted values for the female functions obtained in the first step , and compute predicted values for the male mortality functions. The estimated orthogonal regression equations are:

[4]Throughout we work separately with these two indicators of infant mortality and child mortality.

$$\texttt{logit}({}_1Q_0^{males}) = .071 + .956 * \texttt{logit}({}_1Q_0^{females}) \tag{4.4.1}$$

and

$$\texttt{logit}({}_4Q_1^{males}) = .104 + .956 * \texttt{logit}({}_4Q_1^{females}) \tag{4.4.2}$$

It is also possible use a simpler strategy: this is to circumvent the second step altogether by simply assuming sex ratios at birth and then using these in combination with the predicted value of the female functions obtained in the first step to derive final estimates of male mortality functions. The extra simplicity of this strategy is deceiving, however, since there are multiple sets of sex ratios among which one could choose in each case and no precise guidelines about how to choose among them . We prefer instead to opt for the more complicated strategy that builds into final estimates observed patterns of sex differentials of child mortality.

## 4.5 Estimation for the period 1850-1950

Very few countries have suitable, continuous vital statistics for the period before 1950. As a consequence we employ very different procedures to arrive at final estimates of child mortality for that. They varied depending on the country-year and the data available to us for each. A maximum of three techniques were used in each case and, when more than one of them be applied, we include the corresponding estimates in the pool of plausible estimates for the country-year.

### 4.5.1 Adjusted vital statistics

In countries with partial vital statistics before 1950 we proceeded to compute estimates of infant and child mortality using the raw death, birth and census population counts and we correct them using adjustment factors derived from the sequence of values already estimated for the years after 1950.

Let $t$ be the mid-point of an intercensal period closest to the first year for which we have a robust estimate of infant and child mortality (see above). We then compute infant mortality centered for year $t$ as:

$${}_1Q_0 = \frac{{}_1d_0(t)}{b(t)} \tag{4.5.1}$$

where ${}_1d_0(t) ={}_1 d_0(t) ={}_1 D_0(t-1) * a +_1 D_0(t) +_1 D_0(t+1) +_1 D_0(t+2) * (1-a)$ and $b(t) = B(t-1) + B(t) + B(t+1)$. The quantity $a$ is a separation factor drawn from the Coale-Demeny model life tables whereas ${}_1D_0(t-1)$, ${}_1D_0(t)$, ${}_1D_0(t+1)$, ${}_1D_0(t+2)$ are deaths in years $t-1$, $t$, $t+1$ and $t+2$ and, finally, $B(t-1)$, $B(t)$, and $B(t+1)$ are births in years $t-1$, $t$ and $t+1$.

The next step is to choose from all years after 1950 the one closest to $t$ for which we have an estimate of infant mortality after fitting the lowess-spline functions, say, ${}_1Q_0^{splines}(t*)$. We then compute an adjustment factor or the ratio using the observed and estimated (from spline fitting) values or ${}_1R_0 ={}_1 Q_0^{splines}(t*)/{}_1Q_0^{vital}(t)$ and apply it to all years before 1950 for which we have vital statistics available to get adjusted values, namely, ${}_1Q_0^{adj}(t^*) ={}_1 Q_0^{vital}(t^*) *_1 R_0$

for $t^* < 1950$. A similar procedure is used to adjust child mortality (1-4) centered in a census year $t$. In this case we compute:

$$ {}_4Q_1^{vital}(t) = \frac{{}_4d_1(t)}{{}_4P_1(t)} \tag{4.5.2} $$

where ${}_4P_1(t)$ is the census population aged 1 to 4, $t$, ${}_4d_1(t) =_4 D_1(t-1) +_4 D_1(t) +_4 D_1(t+1)$, and the adjustment ratio is ${}_4R_1 =_4 Q_1^{vital}(t)/{}_4Q_1^{splines}(t)$. The adjusted mortality rates are computed as ${}_4Q_1^{adj}(t^*) =_4 Q_1^{vital}(t^*)/{}_4R_1$ for all years $t^* < 1950$ for which there are available censuses and vital statistics.[5]

Table 4.7 include countries and years in which the foregoing procedures were used.

Table 4.7: Country-years with estimates of child mortality computed from adjusted vital statistics before 1950

| **Country-years with estimates of child mortality computed from adjusted vital statistics** | |
|---|---|
| | |
| country | vital statistics years |
| | |
| Argentina | 1904, 1914, 1947 |
| Bolivia | NA |
| Brazil | Not used |
| Chile | 1920, 1925,1930,1935,1940,194 |
| Colombia | 1938, 1944 |
| Costa Rica | 1900,1910,1920,1927,1930,1938 |
| Cuba | 1919,1931,1943 |
| Dominican Republic | 1935,1942 |
| Ecuador | Not used |
| El Salvador | 1930-1940 |
| Guatemala | 1940 |
| Honduras | 1940, 1945 |
| Mexico | 1900, 1910,1921,1925,1935,1940 |
| Nicaragua | 1940 |
| Panama | 1940 |
| Paraguay | Not used |
| Peru | 1940 |
| Uruguay | 1908,1935 |
| Venezuela | 1938,1945 |

[5]In many cases we were able to estimate a *time trend* of adjustments factors beginning in a year within the period 1950-1954 and extending up to 2010. When this was possible and we could also compute observed values of child mortality functions before the year 1950, we extrapolated the time trend of adjusted values backwards and assigned adjustment factors to years contained in the interval $1930 - 1950$ which are then used to adjust the observed values of the mortality functions.

### 4.5.2 Estimates based on the generalized ogive

A number of countries lacked vital statistics before 1950 either partially or totally. Most of them, however, fielded one of more population censuses. First, when two censuses were available we used a generalized version of the classic ogive method (Coale et al., 1983) which retrieves estimates of adult mortality (ages above 5) supported by a choice of model mortality patters. The method is described in full elsewhere (see chapter 3). The sensitivity of estimates of life expectancy above age 5 to variation of mortality patterns is quite low. However, inferences about child mortality given an estimate of life expectancy supported by a model of mortality is a more delicate matter. To circumvent this problem we generate three estimates using models West and South from the Coale-Demeny pattern and the new Latin American mortality model. Because estimates of the latter were always between those associated with the West and South we took the average of all three as our final estimate and linked it to the mid point of the intercensal year.

Second, when only one census was available we used the classic ogive that, unlike the generalized ogive, assumes stability and proceed to generate three estimates and their average as described before.

### 4.5.3 Estimates based on a 3-parameter Gompertz function

The third method we use consists of joining together any estimates of infant and child mortality available from the first and second procedure for the period before 1950 with those for the period 1950-1960 and fit a 3-parameter Gompertz function. The predicted values from this function for all years before 1950 are defined as yearly estimates for that period.

The functional form for the 3-parameter Gompertz for ${}_1Q_0$ and ${}_4Q_1$ is as follows:

$$Z(t) = \beta_1 \exp(-\exp(-\beta_2(t - \beta_3))) + \varepsilon_t \tag{4.5.3}$$

where $Z(t)$ stands for either ${}_1Q_0(t)$ or ${}_4Q_1(t)$ and $t$ corresponds to the year for which the mortality measures were estimated. The parameters $\beta_1$ and $\beta_2$ have similar roles to those for the parameters in the classic two parameter Gompertz function. The third parameter, $\beta_3$, represents the time (year) before which the time trend becomes invariant or the function reaches its maximum. We will refer to it as the "ceiling" parameter.

Two caveats are important. First, in a few cases the time trend of mortality functions descends to rapidly from high values located associated with the first years we were able to gather estimates for. In these cases the ceiling parameter is not estimable. Second, in all cases where a ceiling parameter was estimable we ensured that estimates of *adult mortality* for years that preceded the year of attainment of the ceiling value was invariant. If that was not the case we assumed that the descending adult mortality trend was correct, ignored the ceiling value of child mortality and computed values of child mortality for those years using the two remaining Gompertz parameters in combination with minimally modified higher value of the ceiling parameter.

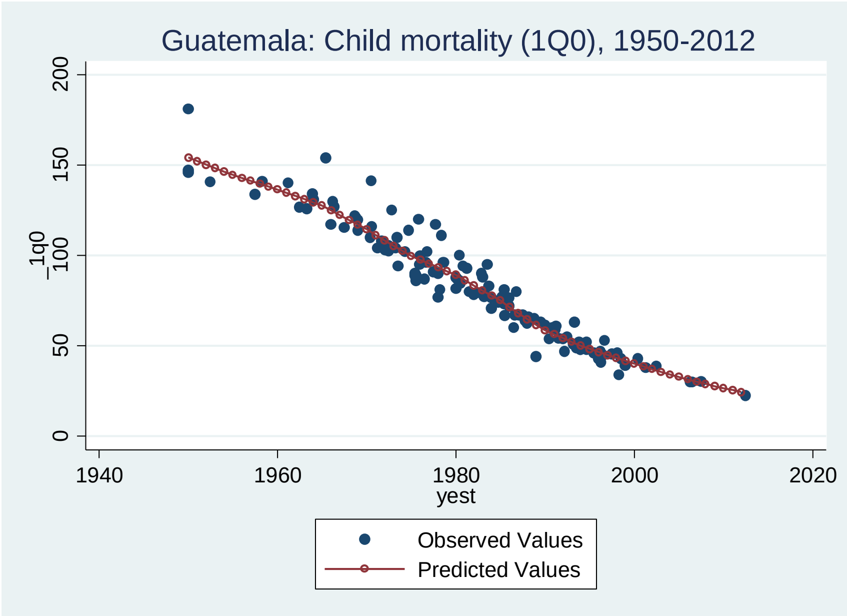
## 4.6 Final database

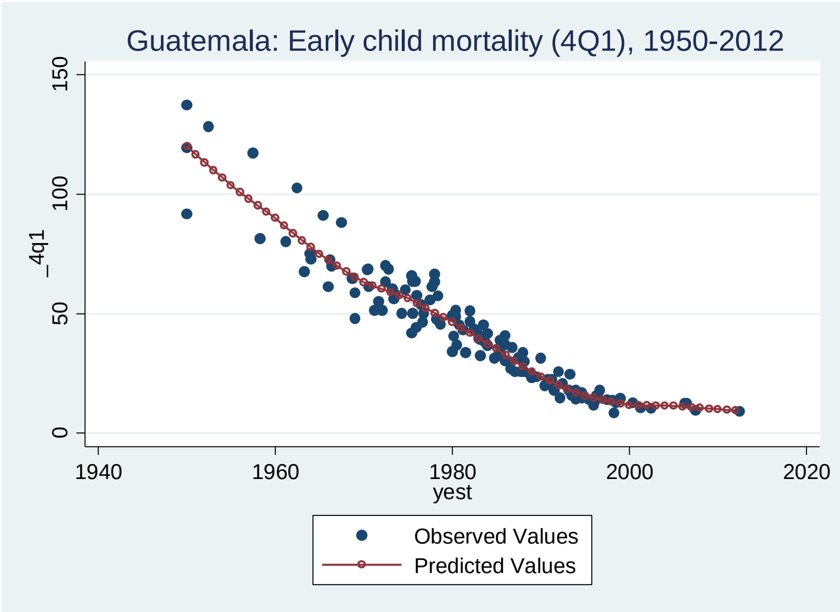
The final database contains yearly point estimates on infant (${}_1Q_0$) and early child (${}_4Q_1$) mortality by sex, spanning the period from 1850 through 2012 for all 19 LAC countries. Table 4.7 contains all the information available in LAMBdA.

## 4.7 Comparisons of LAMBdA and alternative estimates for a limited time interval

How do LAMBdA estimates compare with alternative ones? Figures 4.2 and 4.3 plot LAMBdA values of child mortality, ${}_5Q_0$, for males and females, respectively, with those obtained by the consortium UN-IGME group (UN IGME, 2017). These are available for a shorter period of time following 1960-70. Although the agreement is not perfect there is much consistency between the two sets of estimates despite the different procedures used to generate them.

Figure 4.1: Observed and estimated values of infant and child mortality: Guatemala 1940-2015



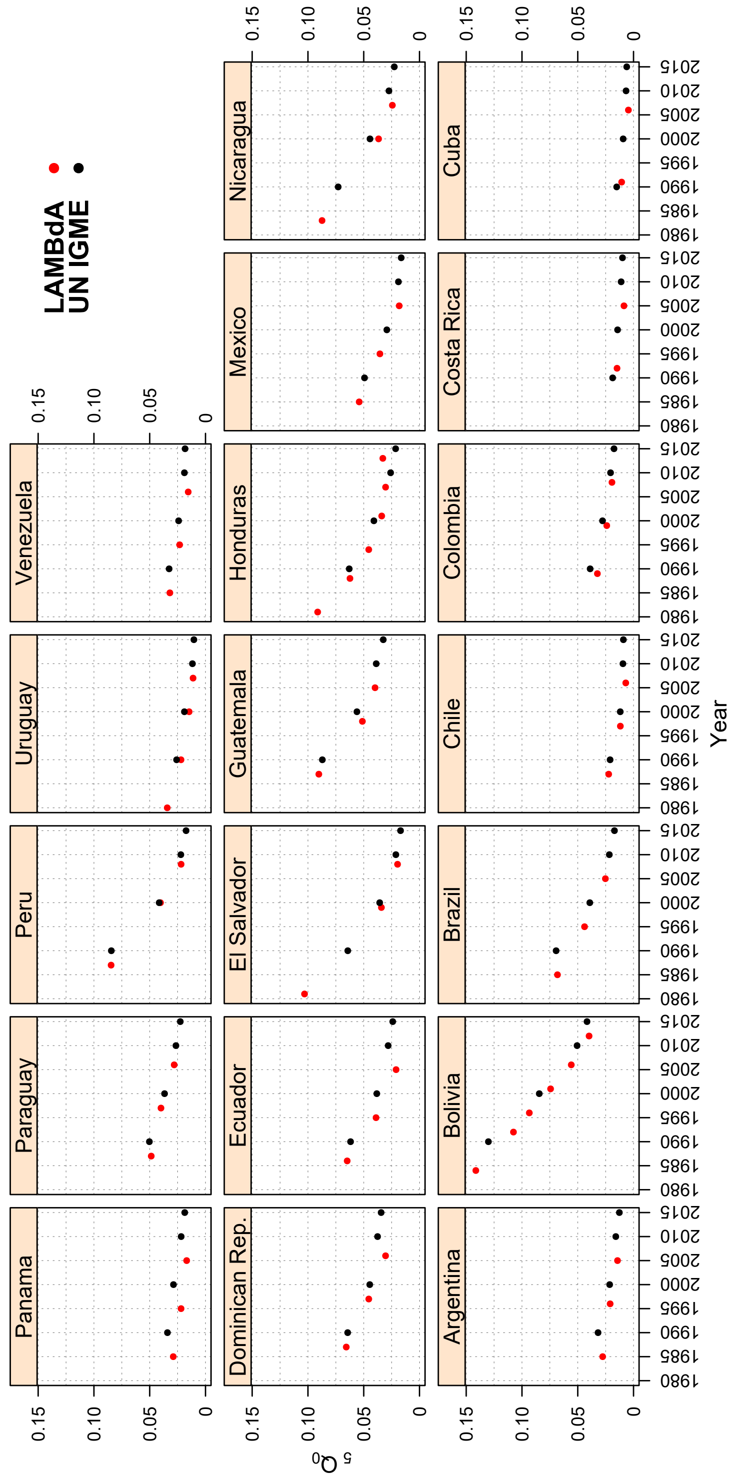
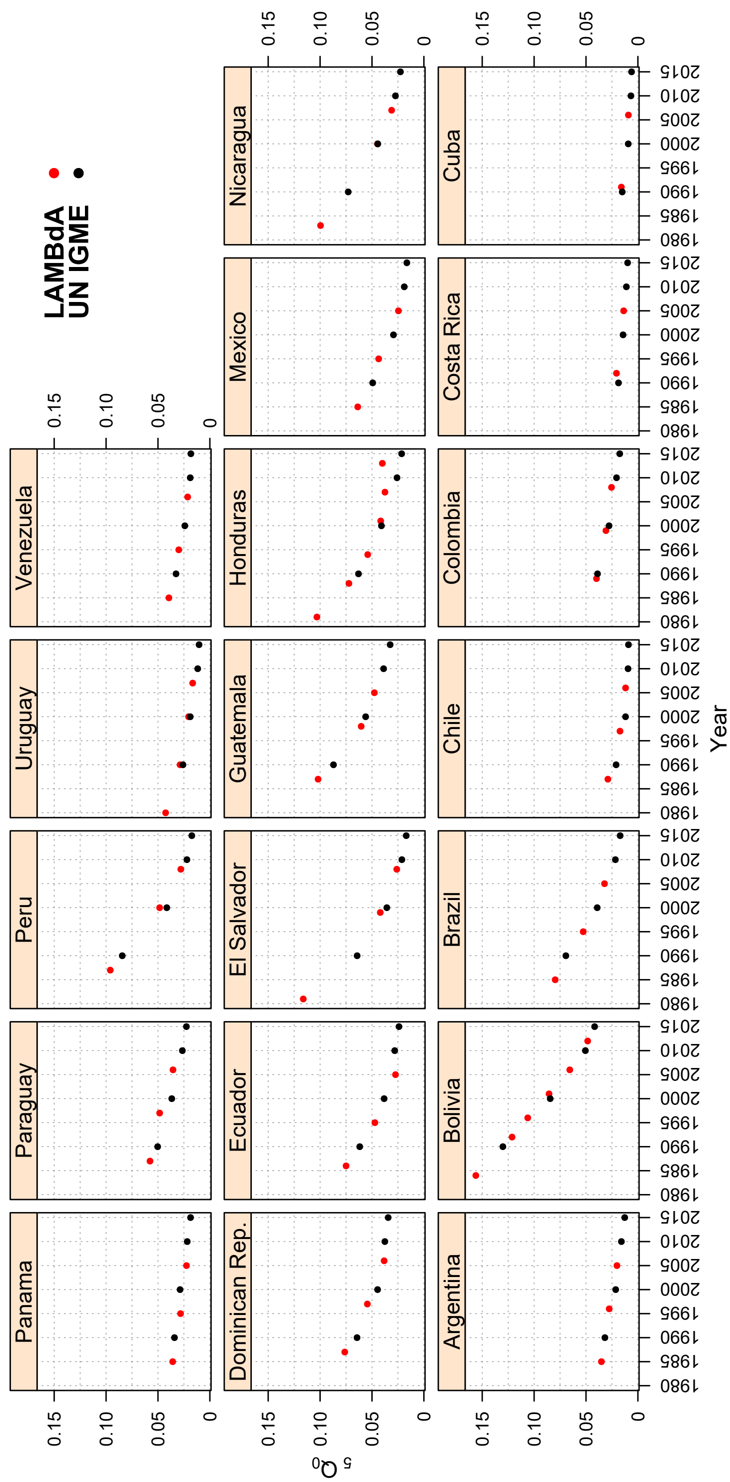


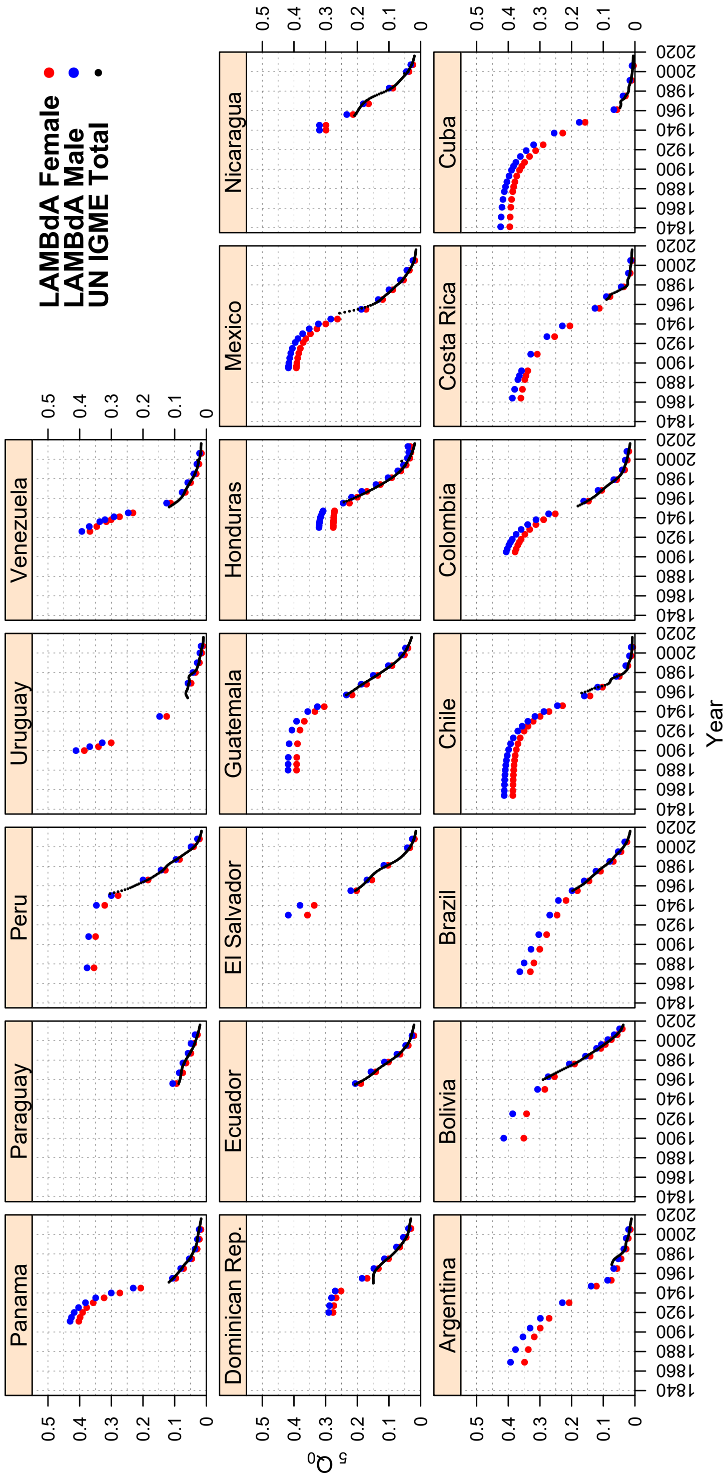

Figure 4.2: Estimates of child mortality ($_5Q_0$) for females: LAMBdA and IGME

Figure 4.3: Estimates of child mortality ($_5Q_0$) for males: LAMBdA and IGME



More consequential are the country-specific shapes of the long-run time trends estimated in LAMBdA (but not in UN-IGME)(Figure 4.4). In all but two cases the curvature of the estimates, the implied, trends and ceiling values (those prevailing before 1850-1900) are well-behaved. The exceptions are Panama, Uruguay and Venezuela, countries where the mortality decline in the neighborhood 1875 and 1915 is quite precipitous and no ceiling parameter could be estimated.

Figure 4.4: Time trends in estimates of child mortality ($_5Q_0$) by sex: LAMBdA and IGME



# Chapter 5

# Computing Single Year Population and Death Counts from Data Clustered in Five Year Age Groups

## 5.1 Introduction

The bulk of data to construct life tables contained in LAMBdA originate in two independent sources, namely population censuses and vital statistics. Most Latin American and Caribbean countries conducted periodic population censuses and maintained and made public registration of yearly vital events since 1950. These raw data on population and counts of vital events are available in the UN and WHO (PAHO) databases as well as in each country statistical offices. With some exceptions, the data are available to us are by calendar year and five year age groups (except for the first two age groups, 0 and 1-4, and the last, open, age group, 85+).

In this document we report results from a series of consistency checks for splitting $n$x1 data into 1x1 format using two approaches: Sprague multipliers (Sprague, 1880; Shryock et al., 1976) and cubic splines with several nodes (McNeil et al., 1977). We illustrate the use of these methods by applying them to aggregate counts of populations and deaths in two countries, Mexico and Guatemala. Our goal is to assess differences of results between the two methods. The aim of these tests is to unveil singularities that could signal anomalies in the disaggregation of counts in the data.

The main take away message is that the key statistics whose behavior we chose to study that result from the two methods exhibit no major discrepancies. We find that when departures between methods are detected, they are mainly in younger ages (age group 5-10 and 10-15) and are probably due to boundary conditions on which estimation of McNeil's splines depend, on one hand, and on left censoring imposed at age 5 when using Sprague multipliers. But our checks reveal no deviant behaviors of the estimated single year mortality rates that could induce systematic biases or distortions if one uses Sprague vs. McNeil's spline in the disaggregation of death and/or population counts.

To implement consistency checks we apply both methods to identical raw data and then compute selected statistics to assess the behaviors of the methods. In the next sections we first describe each method and then review two types of consistency checks, one for population and death counts and the other for mortality rates.

## 5.2 From age groups to single-years of age

In this section we briefly describe application of Sprague multipliers and McNeil splines.

### 5.2.1 Sprague multipliers

One of the earliest methods to disaggregate data into single ages involves simple linear interpolation between adjacent points. This method, however, often led to discontinuities at the beginning and end points of age groups, an inherent limitation of fitting independent line segments. One of the early solutions to this problem was to smoothly join the end points of the interpolation through a process known as "osculatory interpolation" (Shryock et al., 1976). The basic idea is to simultaneously estimate overlapping interpolation equations to achieve a smooth transition between age groups.

Sprague (1880) developed one of this methods in the late 1800's by looking at leading differences between 5 successive age groups using a fourth degree polynomial, a method called 'the fifth-difference equation' approach. The use of a fourth degree polynomial allows one to impose 3 conditions that assure a smooth transition between adjacent polynomials. Briefly, the equation is based on two fourth degree polynomials with the following conditions: (1) cross at the same ordinate (y-axis), (2) same slope, and (3) similar radius of curvature at points $y_{n+2}$ and at $y_{n+3}$, where $n$ is the lenght of the age-interval. Thus, the predicting equation is given by:

$$\begin{aligned} y_{n+2+x} = y_n + \frac{x+2}{1!}\Delta y_n + \frac{(x+2)(x+1)}{2!}\Delta^2 y_n + \frac{(x+2)(x+1)x}{3!}\Delta^3 y_n + \\ \frac{(x+2)(x+1)x(x-1)}{4!}\Delta^4 y_n + \frac{x^3(x-1)(5x-7)}{5!}\Delta^5 y_n \end{aligned} \tag{5.2.1}$$

where $\Delta^i$ represents the $i$th difference.

For practical purposes, however, we applied coefficients derived from equation (5.2.1) by Shryock et al. (1976)(see part B of Figure 5.1). These coefficients are typically applied to 5 consecutive age groups and the interpolation is limited to the middle age groups (i.e., excluding the two end points in a 5 consecutive sequence of age groups). An important advantage of using the Sprague multipliers is that the sum of the interpolated single-age values is consistent with the total number of counts within the age group.

**Two alternative applications of Sprague multipliers**

We used two alternative forms of the data when applying Sprague multipliers. The first (labeled "older") consists of 17 age groups corresponding to ages 5-9, 10-14, ..., 85+. The second one (labeled "younger") is similar to the first but excludes the open-ended age group (i.e., 85+). Strictly speaking, one should not use the open-ended age group as it is an improper point of support for splitting the two preceding five year age groups. However, there is a powerful argument to justify the use of data at ages 85+. This is that death and

Figure 5.1: Interpolation coefficients based on the Sprague formula.

**A. FOR INTERPOLATION BETWEEN GIVEN POINTS AT INTERVALS OF 0.2**

| Interpolated point | Coefficients to be applied to-- | | | | | |
|---|---|---|---|---|---|---|
| | $N_{1.0}$ | $N_{2.0}$ | $N_{3.0}$ | $N_{4.0}$ | $N_{5.0}$ | $N_{6.0}$ |
| | First interval | | | | | |
| $N_{1.0}$ | +1.0000 | .0000 | .0000 | .0000 | .0000 | |
| $N_{1.2}$ | +.6384 | +.6384 | -.4256 | +.1824 | -.0336 | |
| $N_{1.4}$ | +.3744 | +.9984 | -.5616 | +.2304 | -.0416 | |
| $N_{1.6}$ | +.1904 | +1.1424 | -.4896 | +.1904 | -.0336 | |
| $N_{1.8}$ | +.0704 | +1.1264 | -.2816 | +.1024 | -.0176 | |
| | Next-to-first interval | | | | | |
| $N_{2.0}$ | .0000 | +1.0000 | .0000 | .0000 | .0000 | |
| $N_{2.2}$ | -.0336 | +.8064 | +.3024 | -.0896 | +.0144 | |
| $N_{2.4}$ | -.0416 | +.5824 | +.5824 | -.1456 | +.0224 | |
| $N_{2.6}$ | -.0336 | +.3584 | +.8064 | -.1536 | +.0224 | |
| $N_{2.8}$ | -.0176 | +.1584 | +.9504 | -.1056 | +.0144 | |
| | Middle interval | | | | | |
| $N_{3.0}$ | .0000 | .0000 | +1.0000 | .0000 | .0000 | .0000 |
| $N_{3.2}$ | +.0128 | -.0976 | +.9344 | +.1744 | -.0256 | +.0016 |
| $N_{3.4}$ | +.0144 | -.1136 | +.7264 | +.4384 | -.0736 | +.0080 |
| $N_{3.6}$ | +.0080 | -.0736 | +.4384 | +.7264 | -.1136 | +.0144 |
| $N_{3.8}$ | +.0016 | -.0256 | +.1744 | +.9344 | -.0976 | +.0128 |
| | Next-to-last interval | | | | | |
| $N_{4.0}$ | | .0000 | .0000 | +1.0000 | .0000 | .0000 |
| $N_{4.2}$ | | +.0144 | -.1056 | +.9504 | +.1584 | -.0176 |
| $N_{4.4}$ | | +.0224 | -.1536 | +.8064 | +.3584 | -.0336 |
| $N_{4.6}$ | | +.0224 | -.1456 | +.5824 | +.5824 | -.0416 |
| $N_{4.8}$ | | +.0144 | -.0896 | +.3024 | +.8064 | -.0336 |
| | Last interval | | | | | |
| $N_{5.0}$ | | .0000 | .0000 | .0000 | +1.0000 | .0000 |
| $N_{5.2}$ | | -.0176 | +.1024 | -.2816 | +1.1264 | +.0704 |
| $N_{5.4}$ | | -.0336 | +.1904 | -.4896 | +1.1424 | +.1904 |
| $N_{5.6}$ | | -.0416 | +.2304 | -.5616 | +.9984 | +.3744 |
| $N_{5.8}$ | | -.0336 | +.1824 | -.4256 | +.6384 | +.6384 |
| $N_{6.0}$ | | .0000 | .0000 | .0000 | .0000 | +1.0000 |

**B. FOR SUBDIVISION OF GROUPS INTO FIFTHS**

| Interpolated subgroup | Coefficients to be applied to-- | | | | |
|---|---|---|---|---|---|
| | $G_1$ | $G_2$ | $G_3$ | $G_4$ | $G_5$ |
| | First panel | | | | |
| First fifth of $G_1$ | +.3616 | -.2768 | +.1488 | -.0336 | |
| Second fifth of $G_1$ | +.2640 | -.0960 | +.0400 | -.0080 | |
| Third fifth of $G_1$ | +.1840 | +.0400 | -.0320 | +.0080 | |
| Fourth fifth of $G_1$ | +.1200 | +.1360 | -.0720 | +.0160 | |
| Last fifth of $G_1$ | +.0704 | +.1968 | -.0848 | +.0176 | |
| | Next-to-first panel | | | | |
| First fifth of $G_2$ | +.0336 | +.2272 | -.0752 | +.0144 | |
| Second fifth of $G_2$ | +.0080 | +.2320 | -.0480 | +.0080 | |
| Third fifth of $G_2$ | -.0080 | +.2160 | -.0080 | .0000 | |
| Fourth fifth of $G_2$ | -.0160 | +.1840 | +.0400 | -.0080 | |
| Last fifth of $G_2$ | -.0176 | +.1408 | +.0912 | -.0144 | |

**B. FOR SUBDIVISION OF GROUPS INTO FIFTHS--Continued**

| Interpolated subgroup | Coefficients to be applied to-- | | | | |
|---|---|---|---|---|---|
| | $G_1$ | $G_2$ | $G_3$ | $G_4$ | $G_5$ |
| | Middle panel | | | | |
| First fifth of $G_3$ | -.0128 | +.0848 | +.1504 | -.0240 | +.0016 |
| Second fifth of $G_3$ | -.0016 | +.0144 | +.2224 | -.0416 | +.0064 |
| Third fifth of $G_3$ | +.0064 | -.0336 | +.2544 | -.0336 | +.0064 |
| Fourth fifth of $G_3$ | +.0064 | -.0416 | +.2224 | +.0144 | -.0016 |
| Last fifth of $G_3$ | +.0016 | -.0240 | +.1504 | +.0848 | -.0128 |
| | Next-to-last panel | | | | |
| First fifth of $G_4$ | | -.0144 | +.0912 | +.1408 | -.0176 |
| Second fifth of $G_4$ | | -.0080 | +.0400 | +.1840 | -.0160 |
| Third fifth of $G_4$ | | .0000 | -.0080 | +.2160 | -.0080 |
| Fourth fifth of $G_4$ | | +.0080 | -.0480 | +.2320 | +.0080 |
| Last fifth of $G_4$ | | +.0144 | -.0752 | +.2272 | +.0336 |
| | Last panel | | | | |
| First fifth of $G_5$ | | +.0176 | -.0848 | +.1968 | +.0704 |
| Second fifth of $G_5$ | | +.0160 | -.0720 | +.1360 | +.1200 |
| Third fifth of $G_5$ | | +.0080 | -.0320 | +.0400 | +.1840 |
| Fourth fifth of $G_5$ | | -.0080 | +.0400 | -.0960 | +.2640 |
| Last fifth of $G_5$ | | -.0336 | +.1488 | -.2768 | +.3616 |

**C. FOR SUBDIVISION OF GROUPS INTO TENTHS OR HALVES**

| Interpolated subgroup | Coefficients to be applied to-- | | | | |
|---|---|---|---|---|---|
| | $G_1$ | $G_2$ | $G_3$ | $G_4$ | $G_5$ |
| First tenth of $G_3$ | -.0076 | +.0510 | +.0660 | -.0096 | +.0002 |
| Second tenth of $G_3$ | -.0052 | +.0338 | +.0844 | -.0144 | +.0014 |
| Third tenth of $G_3$ | -.0022 | +.0154 | +.1036 | -.0195 | +.0027 |
| Fourth tenth of $G_3$ | +.0006 | -.0010 | +.1188 | -.0221 | +.0037 |
| Fifth tenth of $G_3$ | +.0027 | -.0133 | +.1272 | -.0203 | +.0037 |
| Sum of coefficients for first five-tenths = coefficients for first half of $G_3$ | -.0117 | +.0859 | +.5000 | -.0859 | +.0117 |
| Sixth tenth of $G_3$ | +.0037 | -.0203 | +.1272 | -.0133 | +.0027 |
| Seventh tenth of $G_3$ | +.0037 | -.0221 | +.1188 | -.0010 | +.0006 |
| Eighth tenth of $G_3$ | +.0027 | -.0195 | +.1036 | +.0154 | -.0022 |
| Ninth tenth of $G_3$ | +.0014 | -.0144 | +.0844 | +.0338 | -.0052 |
| Last tenth of $G_3$ | +.0002 | -.0096 | +.0660 | +.0510 | -.0076 |
| Sum of coefficients for last five-tenths = coefficients for second half of $G_3$ | +.0117 | -.0859 | +.5000 | +.0859 | -.0117 |

Source: Shryock et al. (1976, Table C-5, pp.555) population counts at very old ages tend to be inflated, and sometimes considerably so, by age overstatement, a flaw not easily corrected (see Chapter 3). In fact, in many cases the computed death rates that one obtains when using the results from the "younger" variant suddenly decline at older ages (between 75 and 84). This results from age over-statement that transfers counts from age groups 75-79 and 80-84 into the open age group 85+. Thus, using the "younger" data form generate counts of both population and deaths for ages between 75 and 84 that are too low. Because the bias for population counts has less impact than the bias of death counts, the rates computed from them will be biased downward and more so at very old ages. In contrast, when the age group 85+ is included as a legitimate point of support, most of the biases are removed. Irrespective of which form of the data we use, *the counts in the open age groups are observed counts*, that is, the original, raw, values. In section 5.4.1 we show that, except for irregularities in the oldest age groups associated with the "younger" data form, the two alternative applications yield very similar results.

**Sprague multipliers in LAMBdA**

Available data on population and death counts for LAC countries are usually reported in five-year age groups (i.e., 0-4, 5-9,..., 85+). We applied Sprague multiplies to these data to generate single-age counts of both population and deaths. As noted above, the are two possible applications of the Sprague techique. In what follows we briefly describe this method when using ages 5 to 84 (excluding the open age group). In this case we applied the five-panel interpolation multipliers (part B in Figure 5.1) to each age-group that falls in the middle of a five consecutive set of age groups. For the first two and the last two age groups in this five consecutive sequence, we use the first two (labeled "First panel" and "Next-to-first panel") and last two ("Next-to-last panel" and "last panel") multiplier panels, respectively.

For example, an interpolation of an aggregate count for age group $N_{25-29}$ into five single-age counts should satisfy:

$$N_{25-29} = N_{25} + N_{26} + N_{27} + N_{28} + N_{29}$$

To estimate an interpolated value at single-age 27 we use the multipliers from the "Middle panel" in part B in Figure 5.1 and five consecutive age groups centered around age group 25-29: $N_{15-19}$, $N_{20-24}$, $N_{25-29}$, $N_{30-34}$, and $N_{35-39}$. Thus, $N_{27}$ is estimated as (using the coefficients in the third row of the "Middle panel"):

$$N_{27} = 0.0064 * N_{15-19} - 0.0336 * N_{20-24} + 0.2544 * N_{25-29} - 0.0336 * N_{30-34} + 0.0064 * N_{35-39}$$

These computations produce single-age counts for all age groups except the firdt two (age 5-9 and 10-14) and last two (age 75-79 and 80-84). For these age groups we used the first two panels (labeled "First panel" and "Next-to-first panel") and last two ("Next-to-last panel" and "last panel") shown in part B of Figure 5.1. For example, an estimate for age 7, $N_7$ in age group 5-9, requires the use of the multipliers in the "First panel" and the following four consecutive age groups: $N_{0-4}$, $N_{5-9}$, $N_{10-14}$, and $N_{15-19}$. This leads to the following estimation equation:

$$N_7 = 0.3616 * N_{0-4} - 0.2768 * N_{5-9} + 0.1488 * N_{10-14} - 0.0336 * N_{15-19}$$

Similar computations with other panels enable us to disaggregate counts at ages 10-14, 75-79 and 80-84.[1] Finally, an analogous series of calculations that includes the open age group leads to estimates associated with the"older" version of the multipliers.

### 5.2.2 Cubic splines: McNeil spline

In the late 1970's, McNeil et al. (1977) proposed an alternative method for splitting counts that usually come in age groups into single years of age (1x1). McNeil's method consists on fitting polynomial splines, of which the cubic is the most typical one, to the cumulative distribution of counts (i.e., population or deaths) within each calendar year. We briefly describe the method below.

Let $y(x) = \sum_{i=1}^{x-1} P_i$ be a cumulative count up to age $x$, and assume that $y(x)$ is known for a finite set of ages, i.e., $y(x) \in I$, where the cardinality of $I = n$. An important difference of McNeil's splines relative to splines fitted in most statistical software is the use of $k$ knots, where $k = n - 2$. Thus, the number of knots in McNeil's splines does not correspond to the optimum (i.e., the smallest set of possible knots).

A spline function of degree $m$ is defined as a function of the form (see equation 4 in McNeil et al., 1977)

$$s(x) = \sum_{j=0}^{m} \alpha_j x^j + \sum_{i=1}^{n} \beta_i (x - x_i)_+^m \tag{5.2.2}$$

where $(x - x_i)_+ = (x - x_i)$ if $(x - x_i) \geq 0$ and 0 otherwise, the $x_i$'s satisfy $x_1 < x_2 < x_3 < \cdots < x_n$ and $\alpha_j$, $\forall j \in [0, m]$, and $\beta_i$, $\forall i \in [1, n]$, are parameters to be estimated. Equation (5.2.2) leads to $m + n + 1$ parameters to be estimated. The $x_i$ are called the *knots* of the spline; these are the points at which we know the values of the function.

The smooth spline function defined on the interval $(a, b)$ must pass through the values at each of the points $x_i$. In addition to these interpolation conditions, the smooth function is required to satisfy the boundary conditions on the $k_{th}$ through $m_{th}$ derivatives of $s(x)$ at $a$, $s^{(i)}(a)$, and $b$, $s^{(j)}(b)$, in order that there be a total of $k$ constraints each at $a$ and $b$:

$$\begin{aligned} s^{(i)}(a) &= \eta_i \\ s^{(j)}(b) &= \iota_i \end{aligned}$$

#### McNeil spline in LAMBdA

We implement equation (5.2.2) in LAMBdA with its corresponding boundary conditions as follows. First, we accumulate population and death counts starting at age 5 and ending at age 84 (i.e., last age group is 80-84). If we add an arbitrary upper bound, say at 120, we will have a total of 17 equations (for ages $5, 10, \ldots, 80$ and 120) but with the understanding that we will not predict single years beyond age 85. Second, we will have a total of 19 parameters (four for the cubic and one each for ages $15, \ldots, 80$ and $\omega$). Third, we need two impose two boundary conditions. In the case of mortality and populations the upper derivative

[1] Application of Sprague multipliers was carried out using two STATA do files, one for population and the other for death counts. These are spragueDeaths2017.do and spragueCensus2017.do both located in LAMBdA's web site.

should be 0, but it is unclear what should be the constraint on the derivatives at the lower bound. In LAMBdA we assume that the function is linear between ages 5 and 10, and set the derivative at age 7.5 equal to one fifth of the value of the function at age 10.[2]

## 5.3 Illustrations: Guatemala and Mexico in the period 1900-2010

We disaggregate population and dead counts using both Sprague ("older" and "younger" versions) and McNeil's spline applied on the aggregate counts in Guatemala and Mexico. We selected these countries because they represent cases with relatively poor (Guatemala) and good (Mexico) quality of vital statistics data and choose two time periods when the quality of the data was inferior (1950's) and much improved (2000's).

We assess differences between procedures using two approaches. First, we evaluate discrepancies on the predicted values on both death and population counts. We do so by computing absolute proportionate differences by age and then calculate a weighted average where the weights, $w(x)$, correspond to the proportion of 'events' (population or deaths) in each age group (see equations (5.3.1) and (5.3.2)). For example, the weight associated to single-age 13, `Abs.diff`(13), is the proportion of total 'events' in age group 10-14.

$$\texttt{Abs.diff(x)} = \left|\frac{\texttt{Sprague}(x) - \texttt{McNeil}(x)}{\texttt{McNeil}(x)}\right| \tag{5.3.1}$$

$$\overline{\texttt{Abs.diff}} = \sum_{x=5}^{84} \texttt{Abs.diff(x)} * w(x) \tag{5.3.2}$$

Second, we compute mortality rates in single years of age and compare the values obtained from each of the two procedures. Arguably this is the target of interest since these rates are the main input in the construction of life tables. We follow a similar procedure and first compute absolute proportionate differences of age-specific mortality rates and then calculate a summary, weighted average where the weights, $w(x)$, correspond to the proportion of deaths in each age group:

$$\texttt{Abs.diff.m(x)} = \left|\frac{\texttt{Sprague.}m(x) - \texttt{McNeil.}m(x)}{\texttt{McNeil.}m(x)}\right| \tag{5.3.3}$$

$$\overline{\texttt{Abs.diff.m}} = \sum_{x=5}^{84} \texttt{Abs.diff.m(x)} * w(x) \tag{5.3.4}$$

Note that in the computation of mortality rates we use predicted deaths and population counts from each disaggregation method, respectively. Thus, the rates we obtain are those that would be observed if we applied the same disaggregation procedure on both the numerator and denominator of the mortality rates.

[2] Estimation of McNeil's function was carried out using an R-code located in LAMBdA's web site.

## 5.4 Results

In the two sections that follow we compare results from Sprague procedure using the "older" and "younger" data forms. This is followed by a comparison of results from Sprague with "old" data form and McNeil's spline procedure.

### 5.4.1 Comparison of Sprague using "younger" vs. "older" data

Our final objective is to compute adjusted (for completeness and age misreporting) single year mortality rates, $M_x$. These are functions of both, single year population and death counts as well as adjustments factors depending on estimated completeness of death and population counts and on age misreporting. Thus, a proper, simple, and expeditious comparison of the two alternative applications of Sprague multipliers is best done using adjusted mortality rates, rather than (estimated) single year age group raw counts. The graphs in Figure 5.2 below consist of two panels each and compare four sets of $M_x$'s. In the first panel of each graph we plot the adjusted death rates computed with the "younger" data form of the counts from Sprague and the smoothed (via lowess) values associated with them. The second panel in the graphs does the same but using the "older" data form of the counts. The comparisons is done for a handful of countries (excluding Guatemala and Mexico) , including representation of those with high and poor(er) quality vital statistics and two time periods, one close to 1950 and another with information for the most recent calendar year.[3]

We highlight three results. First, and as expected, the "older" and "younger" adjusted but unsmoothed $M_x$'s are identical up until age 75 but differ for the last two closed age groups (75-79, 80-84); that is, those that use the open age group (85+) as a point of support.

Second, in some cases the unsmoothed "younger" $M_x$'s computed with counts obtained from Sprague multipliers that exclude the open age group, experience implausible drops in the last two age groups. These irregularities are due to age overstatement in death and population counts that transfers counts upwards across age 85. As should be expected, they are much more salient for the earliest period and quite imperceptible in the most recent one. In both cases the irregularities vanish after smoothing. No such irregularities are present in the set of "older" $M_x$'s, that is, those computed with Sprague multipliers that include the open age group as a crutch that helps to offset the impact of age overstatement .

The third result is that both the smoothed and unsmoothed rates computed with the "older" rates are very similar and, in addition, the smoothed "older" and "younger" rates are very close to each other.[4]

In view of these results, all the pivotal life tables in LAMBdA are computed using the set of "older" single year counts of deaths and populations, considered to be uniformly more robust and less affected by age overstatement than the alternative estimates based on the "younger" counts. In the following section we compare these estimates with those that obtain using McNeil splines.

[3]The countries selected are Argentina, Colombia, Guatemala, Paraguay, and Venezuela for periods 1950-1955 and 2000-2010.

[4]The minuscule dip around age 45 is a result of a discrete jump induced by the adjustment for age misreporting at ages 45+. It vanishes after using a local smoother.

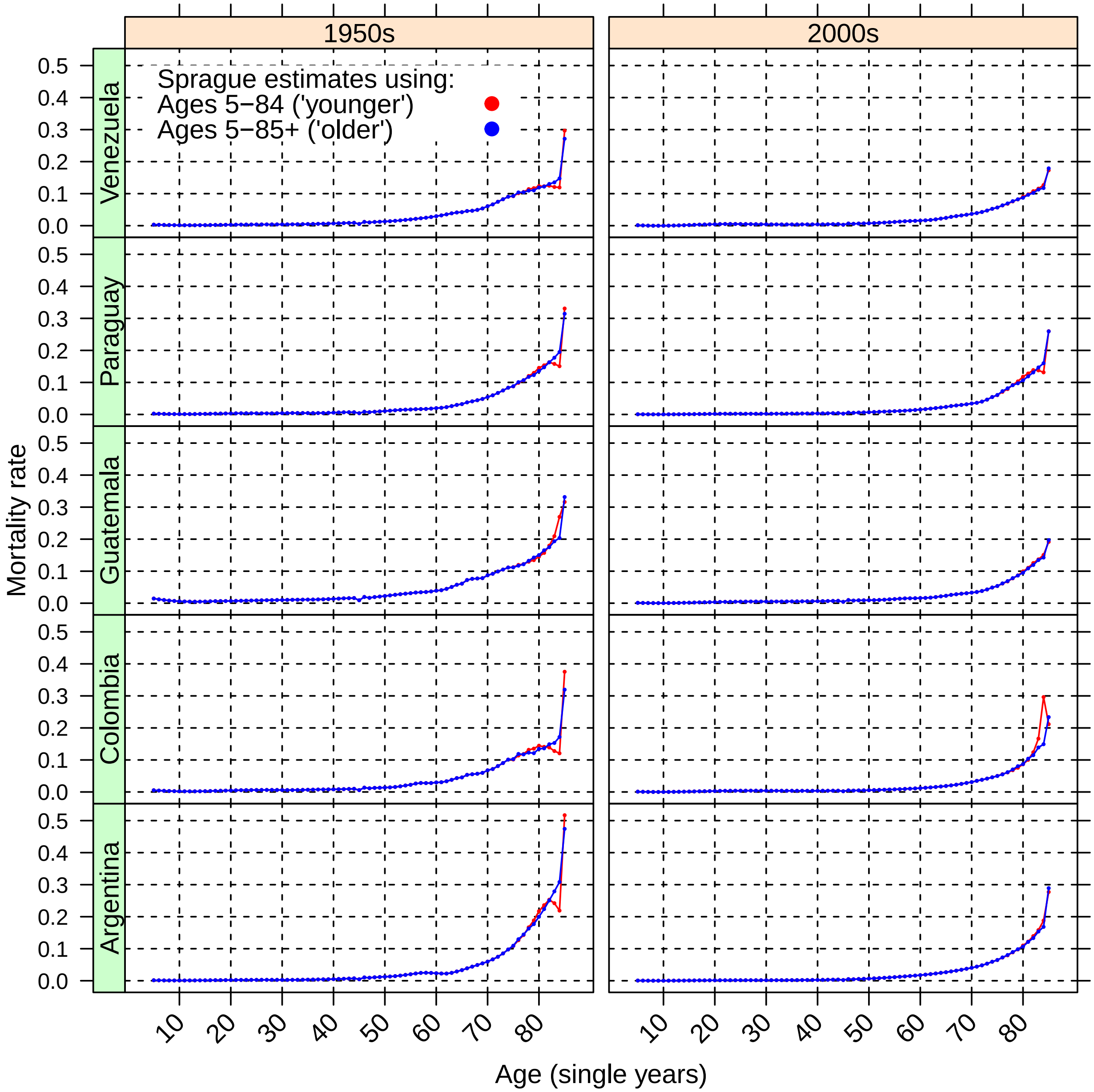
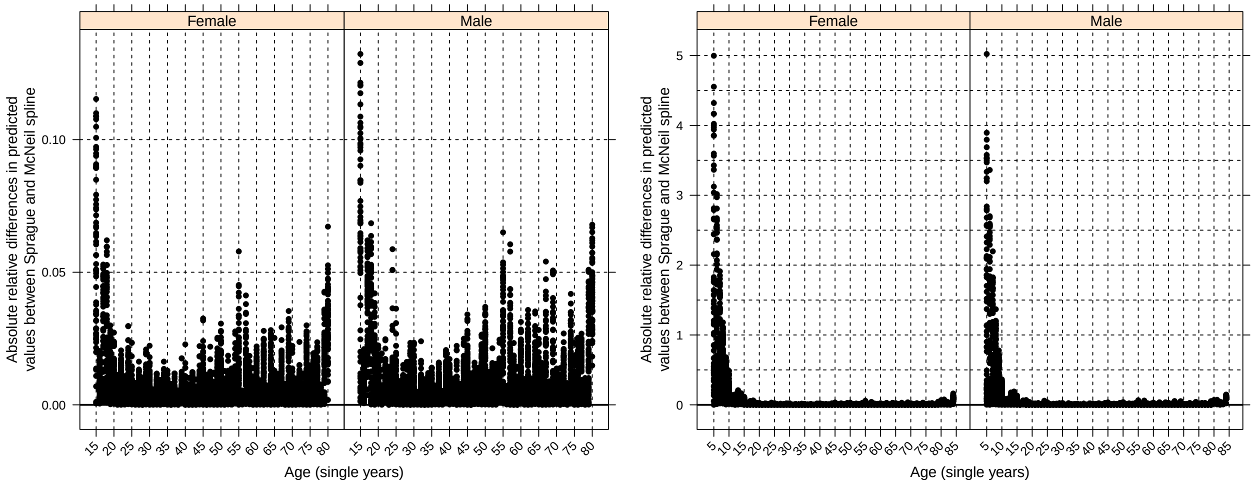



Figure 5.2: Comparisons among males of single-age mortality rates using two different applications of Sprague multipliers: Argentina 1953 and 2007, Colombia 1957 and 2008, Guatemala 1957 and 2005, Paraguay 1956 and 2006, and Venezuela 1955 and 2006.




(a) Ages 15 to 79 (b) Ages 5 to 84

Figure 5.3: Age patterns of absolute proportionate difference in predicted death counts between Sprague and McNeil splines: Guatemala, 1921-2009

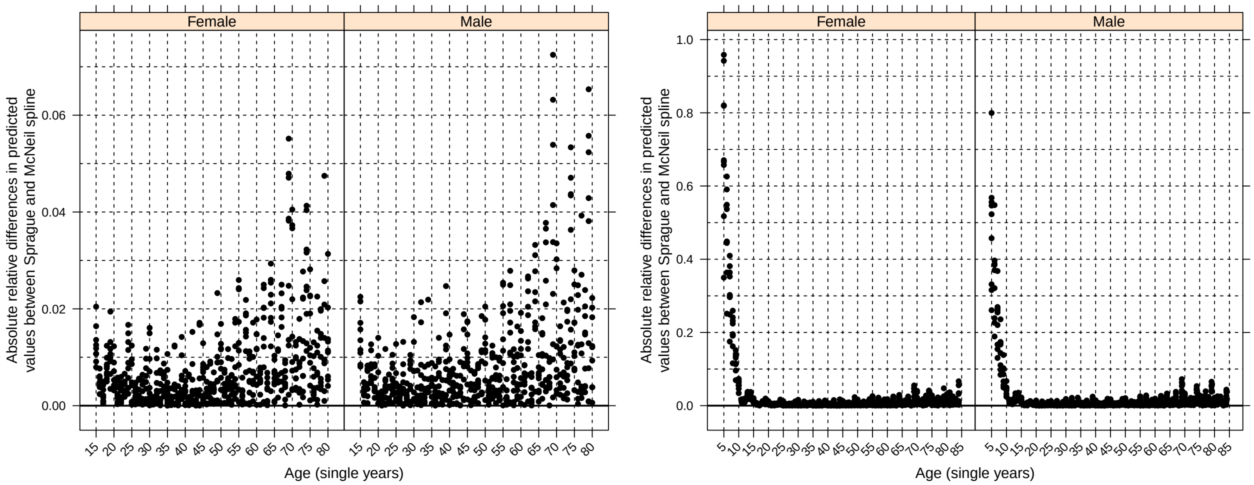
### 5.4.2 Comparison of Sprague using "older" data vs. McNeil spline

We summarize key results using a series of figures. First, we present results on death and population counts separately for Guatemala (Figures 5.3 and 5.4) and Mexico (Figures 5.5 and 5.6). We then plot summary results in the form of weighted averages by year and sex in Figures 5.7 and 5.8 for death counts and in Figures 5.9 and 5.10 for population counts. Second, we repeat the above for mortality rates and show age patterns of absolute proportionate differences by sex, Figures 5.11 and 5.12, and plot weighted averages by year and sex in Figures 5.13 and 5.14.

#### Death and population counts

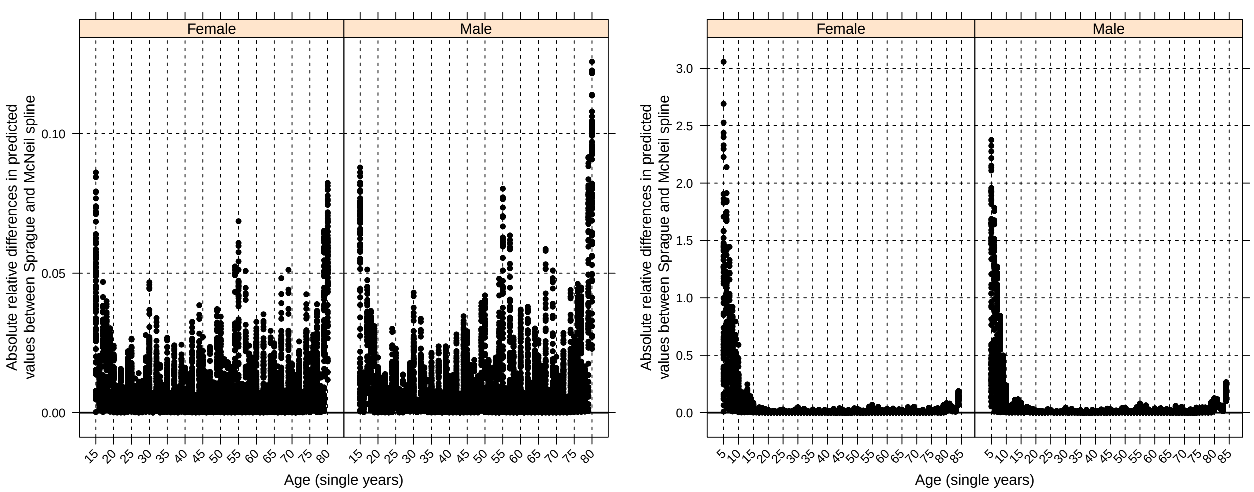
Absolute proportionate differences between Sprague and McNeil's spline show similar age patterns in both countries: the differences between the two sets of estimates are less than percent except at ages <15 and >80 (Figures 5.3–5.6). The results also point to slightly larger absolute differences in deaths than in population counts in Guatemala while the opposite is true in Mexico. Finally, the age pattern of absolute proportionate differences is very similar in males and females in both countries.

The differences at ages younger than 15 ages are likely the result of rather shaky computations with both McNeil's splines and Sprague multipliers. Thus, McNeil 's computations for these ages groups rely heavily on boundary conditions that are explicitly introduced by the researcher. Different boundary conditions leads to different numbers. As noted before, it is unclear what the constraint on the derivatives should be at the lower bound and it is possible that the condition we imposed here does not represent well the underlying changes in the cumulative counts. Modifying the constraint or boundary condition with the aim of reducing discrepancies between the two methods would lead to an infinite regress and is



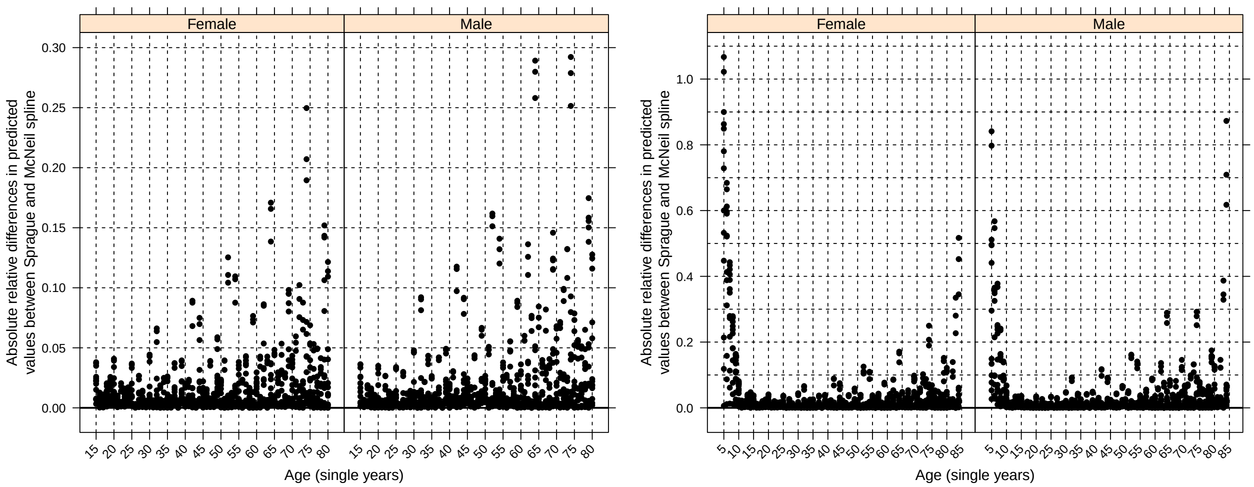

(a) Ages 15 to 79 (b) Ages 5 to 84

Figure 5.4: Age patterns of absolute proportionate difference in predicted population count between Sprague and McNeil splines: Guatemala, 1921-2009.




(a) Ages 15 to 79 (b) Ages 5 to 84

Figure 5.5: Age patterns of absolute proportionate difference in predicted death counts between Sprague and McNeil splines: Mexico, 1900-2010.




(a) Ages 15 to 79 (b) Ages 5 to 84

Figure 5.6: Age patterns of absolute proportionate difference in predicted population count between Sprague and McNeil splines: Mexico, 1900-2010.

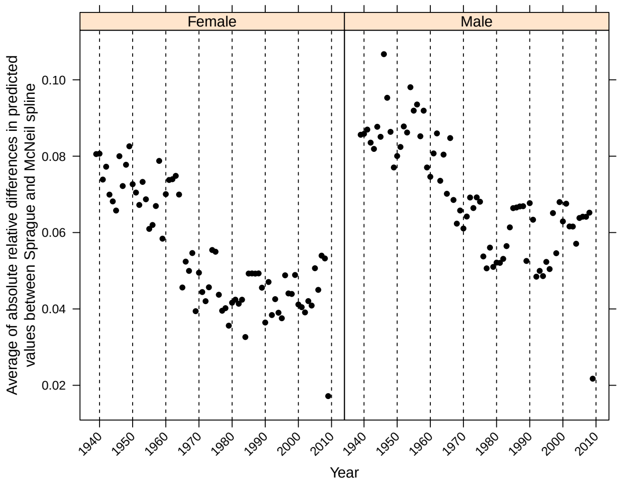
totally impractical. Second calculations based on Sprague multipliers use the two extreme panels (age groups 5-9 and 10-14) that "close" the cubic function are most sensitive to small variations in the quantities for 5 year age groups.

Averaging the absolute proportionate differences over ages highlights similarities between both methods (Figures 5.7–5.10). As one might expect, the large discrepancies between the methods at ages <15 and > 80 described before distort both average levels and time trends. First, in the case of levels there is a 10-fold increase in the average proportionate differences when including ages <15 and > 80 in death counts with a smaller impact on population counts (contrast panels a and b of figures for death and population counts). These patterns are similar in both Guatemala and Mexico.

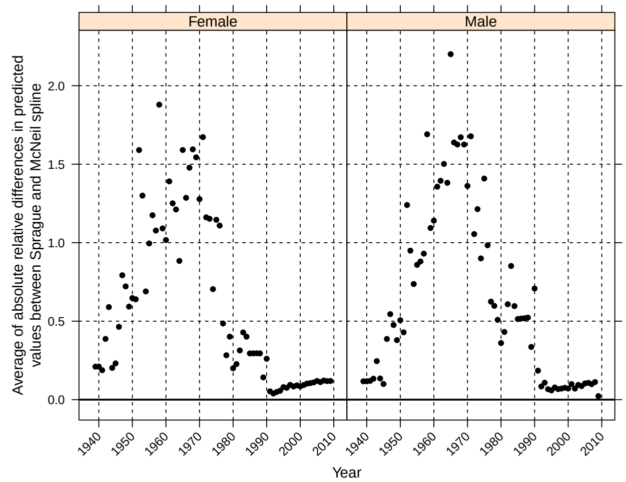
Second, in the case of time trends we observe that average proportionate errors in the 15-80 experience a decline and suggesting more similar results between Sprague and McNeil's spline in recent times. The average differences in death counts attain minima in the range 2-6 percent. In Guatemala, for example, the disaggregation of deaths leads to an average absolute difference of less than 11 percent after 1970 for both males and females (Figures 5.7). In the case of Mexico, the average proportionate differences in deaths counts is similar that in Guatemala for females and slightly higher for males (between 10 and 16 percent after 1970)(Figure 5.8). In the case of population counts, however, the average proportionate differences between the two sets of results is minuscule in both countries and attains values less than reaching 2 percent for males and females (Figures 5.9 and 5.10).

**Mortality rates**

In this section we compare estimates of mortality rates for single years of age . We do so for years when we have both death and population counts, namely "pivotal" years. Age patterns

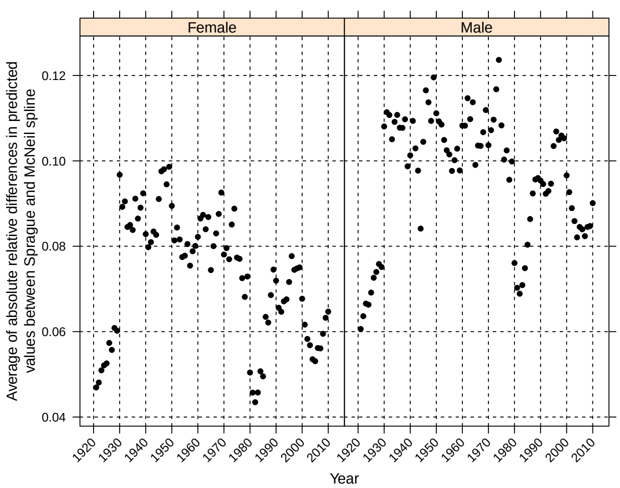



(a) Ages 15 to 79

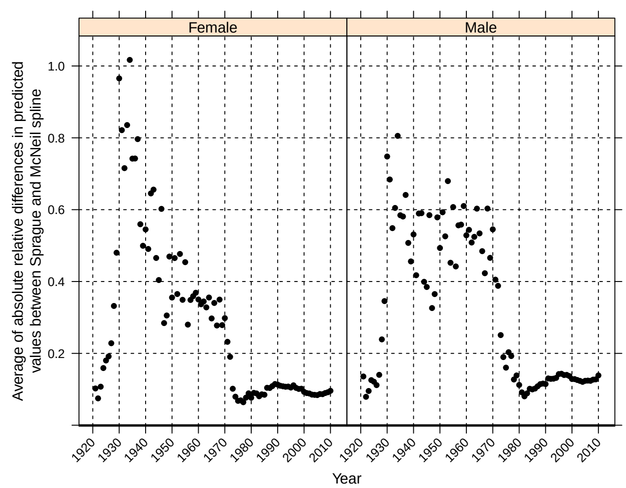



(b) Ages 5 to 84

Figure 5.7: Time trends in average of absolute proportionate difference in predicted death counts across ages between Sprague and McNeil splines: Guatemala, 1921-2009.

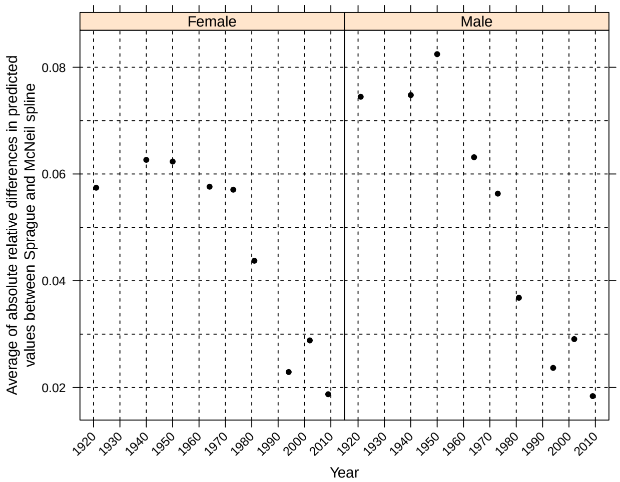



(a) Ages 15 to 79

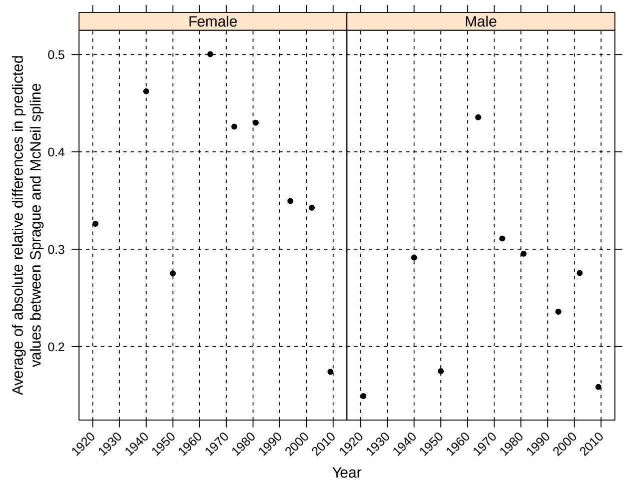



(b) Ages 5 to 84

Figure 5.8: Time trends in average of absolute proportionate difference in predicted death counts across ages between Sprague and McNeil splines by year: Mexico, 1900-2010.

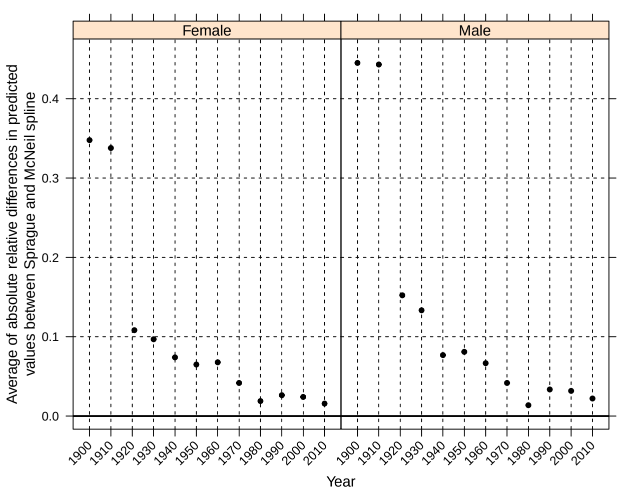



(a) Ages 15 to 79

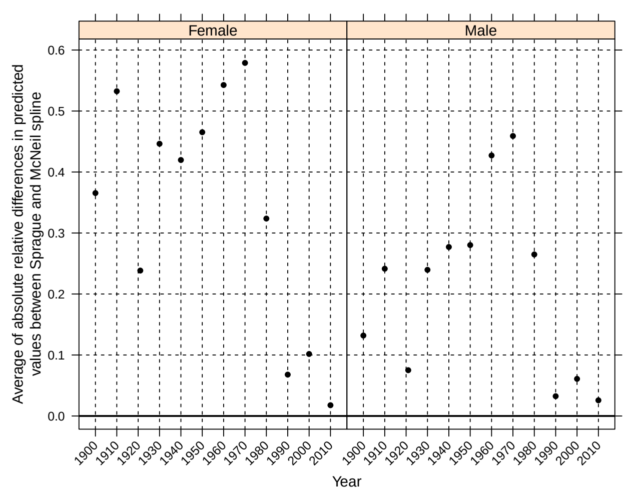



(b) Ages 5 to 84

Figure 5.9: Time trends in average of absolute proportionate difference in predicted population count across ages between Sprague and McNeil splines by year: Guatemala, 1921-2009.

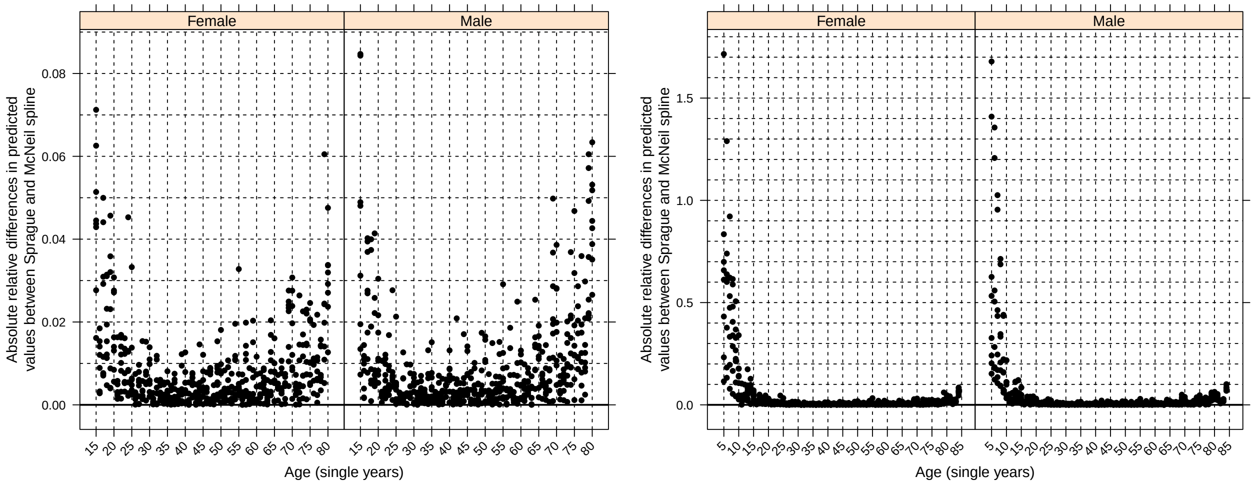



(a) Ages 15 to 79




(b) Ages 5 to 84

Figure 5.10: Time trends in average of absolute proportionate difference in predicted population count across ages between Sprague and McNeil splines by year: Mexico, 1900-2010.




(a) Ages 15 to 79

(b) Ages 5 to 84

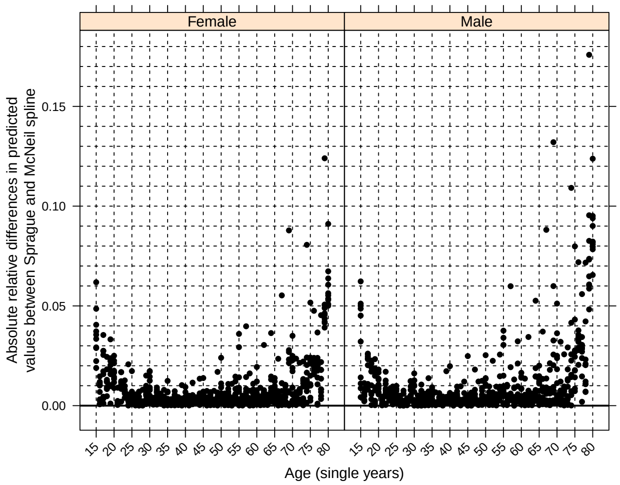
Figure 5.11: Age patterns of absolute proportionate difference in predicted mortality rates between Sprague and McNeil splines: Guatemala, 1921-2009

of absolute proportionate differences are shown in Figures 5.11 and 5.12. Once again, there are large absolute proportionate differences at ages <15 (panel b), when the quantities of interest are very small (all of them close to 0), but consistently lower values when these ages are excluded. For example, absolute proportionate differences are less than 10 percent across ages 15-80, with much lower differences at ages 20-70, within a 5 percent range, for both males and females. This pattern is similar in both countries.

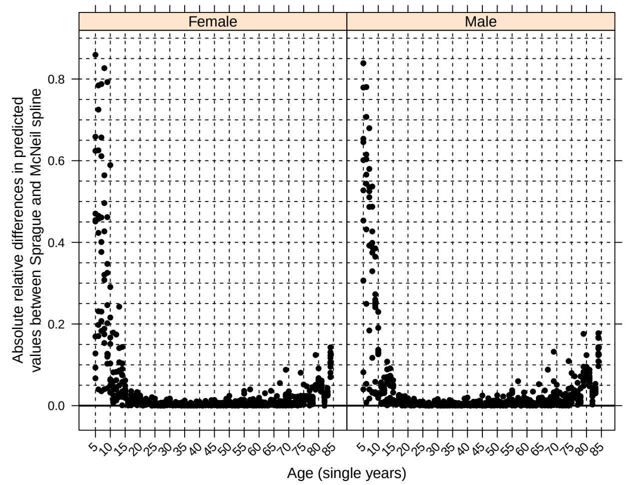
Next we average the absolute proportionate differences using as weights the proportion of deaths in each age group (Figures 5.13 and 5.14). The figures that average absolute proportionate differences are low in both countries for males and females as for most years the differences are within 4 percent except for years around 1970 in Guatemala and in 1950 in Mexico. Furthermore, these average differences decline over time so after 1970 they are consistently below 0.2 percent.

Finally, to further characterize age patterns of absolute proportionate differences, Figures 5.15 and 5.16) display differences in log scale. These figures support two main inferences. First, age patterns of proportionate differences are very similar in the two and in both time periods suggesting that Sprague and McNeil methods should yield comparable results in other LAC countries with profiles contained between those of Guatemala and Mexico and in the span of years covered here (1950-2010). Second, as described above, discrepancies in mortality rates are largest at the lower end of the age range, much less so at the upper end, and quite small everywhere else. Overall these results suggest acceptable levels of similarity between the two disaggregation methods for population and death counts.[5]

[5] The irregularity in seen in Guatemala 2010, where minima of mortality rates in the age group 5-14 are lower than 10 or so, are the only one of this kind observed in the entire set of country-years. Note that this takes place between ages 5 and 14 where the rates are closest to zero. In fact, mortality rates at age

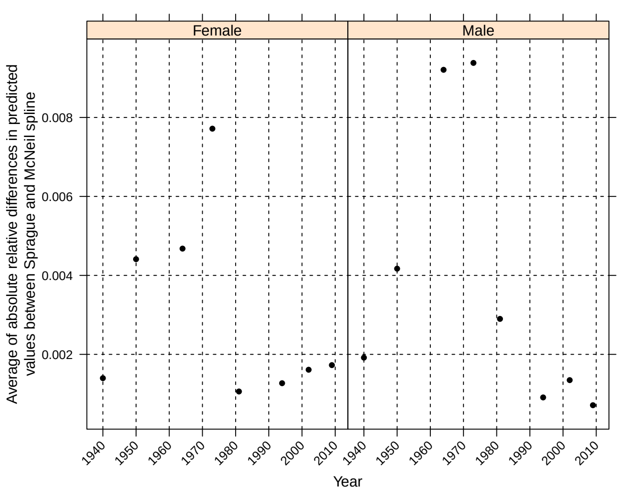



(a) Ages 15 to 79

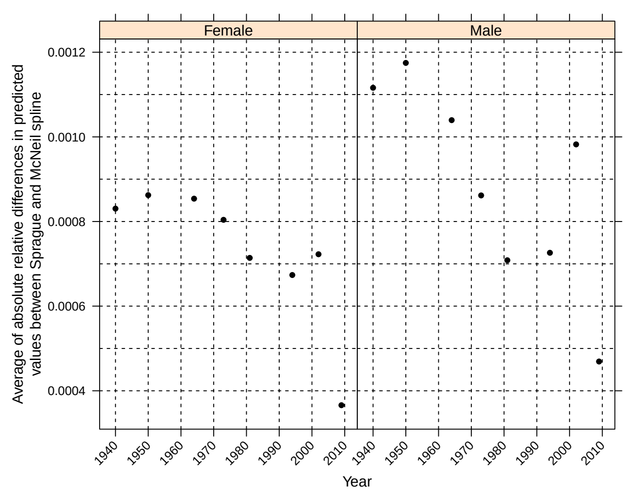



(b) Ages 5 to 84

Figure 5.12: Age patterns of absolute proportionate difference in predicted mortality rates between Sprague and McNeil splines: Mexico, 1900-2010

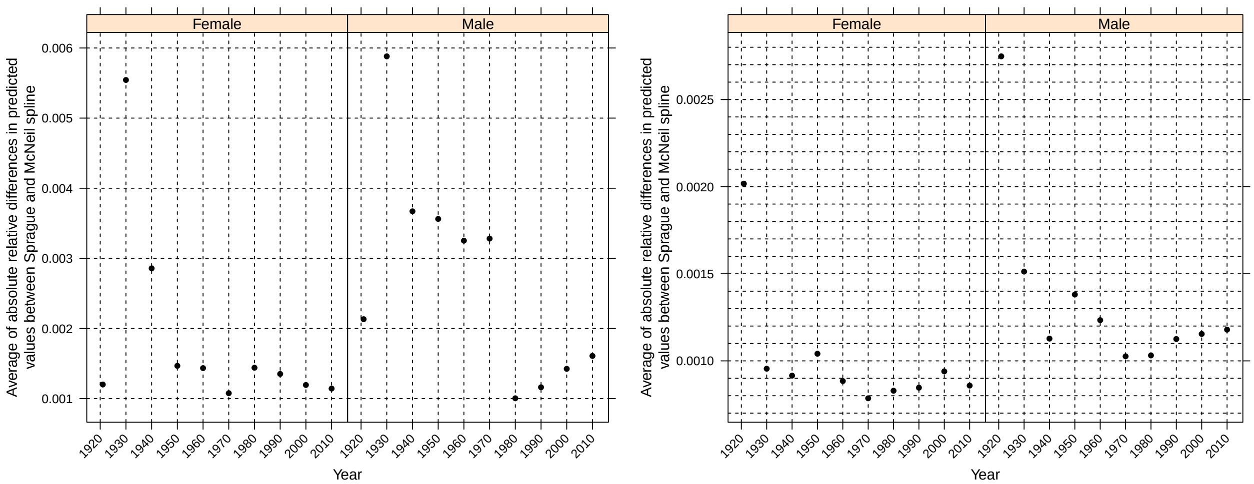



(a) Ages 15 to 79




(b) Ages 5 to 84

Figure 5.13: Time trends in average of absolute proportionate difference in predicted mortality rates between Sprague and McNeil splines: Guatemala, 1921-2009

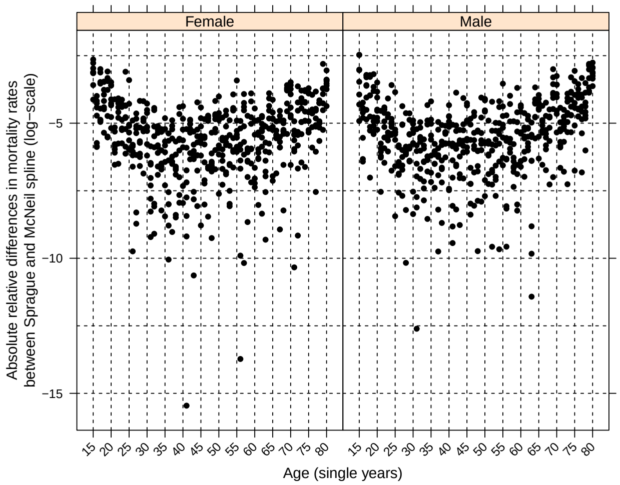



(a) Ages 15 to 79 (b) Ages 5 to 84

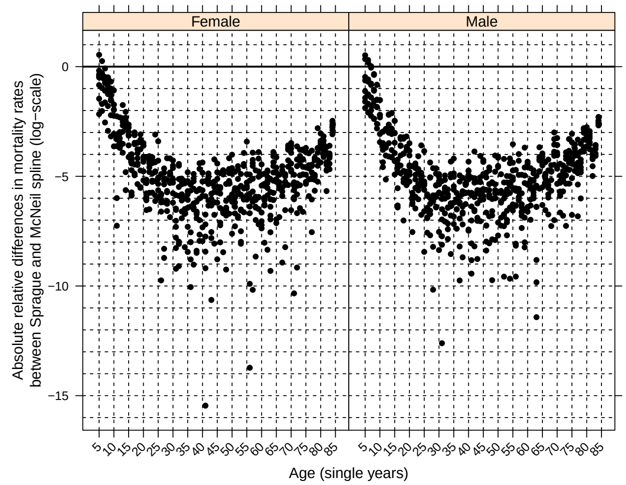
Figure 5.14: Time trends in average of absolute proportionate difference in predicted mortality rates between Sprague and McNeil splines: Mexico, 1900-2010

By and large, the checks implemented above reveal that differences in results associated with each of the two procedures to split multiple year age groups into single ages is highly unlikely to induce systematic biases or distortions of estimates of mortality rates and/or age-patterns of rates.

5 computed with McNeil's spline and Sprague multipliers are .0024 and .0008 respectively, a threefold ratio but of rates that hover around .0015.

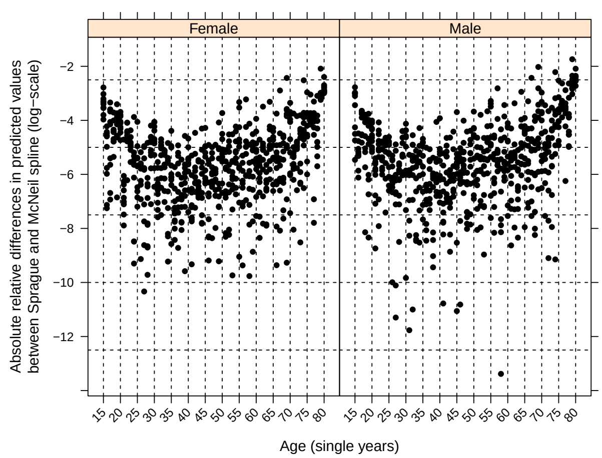



(a) Ages 15 to 79

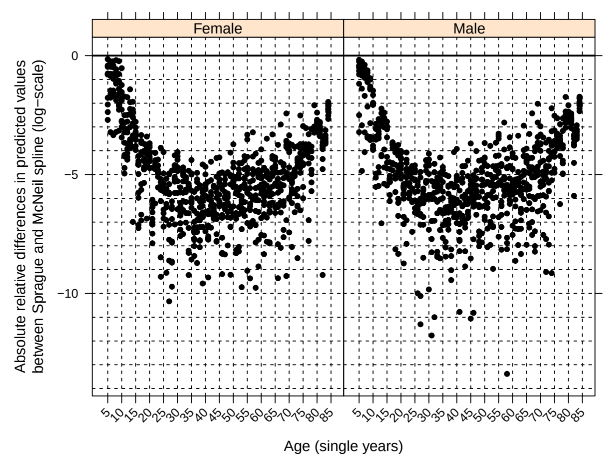



(b) Ages 5 to 84

Figure 5.15: Absolute proportionate difference in predicted mortality rates between Sprague and McNeil splines by age: Guatemala, 1921-2009




(a) Ages 15 to 79




(b) Ages 5 to 84

Figure 5.16: Absolute proportionate difference in predicted mortality rates between Sprague and McNeil splines by age: Mexico, 1900-2010

# Chapter 6

# Assessment of the module of mortality by causes of death and the problem of ill-defined deaths

## 6.1 Introduction

As early as 1950, most LAC countries began establishing consisting reporting of cause of death data. In most cases, death counts are available yearly after 1950 by year of death, sex, and single years of age. These data are of varying quality and are also tightly dependent on changing systems of international classification of diseases (ICD). Table 6.1 contains the definitions of groups of causes of death utilized in LAMBdA by ICD codes and the ICD codes equivalences across multiple ICD's classification systems employed during the period 1950-2010. Figure 6.1 shows available cause of death data in LAC by country, year, ICD, and last age group reported. The figure clearly shows differences in the use of ICD classification across countries and in the level of disaggregation of deaths by age. Before 1970 all countries, except Ecuador, use ICD-7 and included an open age group starting at 85+. For example, between 1970 and 1995, all countries shifted to ICD-8 and ICD-9 and consistently reported deaths up to age 85+. After 1995 all countries transitioned to ICD-10 and the level of disaggregation of deaths by age increased up to age 95+.

Once the data is adjusted for relative completeness[1], the most important deficiency of cause of death data in LAC is the large fraction of deaths classified as ill-defined. For some countries the size of this category is as high as 60 percent of all deaths at ages older than 60. The problem at older ages is exacerbated because assigning an underlying cause of death is a challenging task due to the large number of comorbidities afflicting older adults. The problem is complicated even more by the need to use different cause-of-death coding schemes and changing disaggregation of deaths at older ages.

Failure to correct for ill-defined deaths will lead to incorrect inferences in analyses of

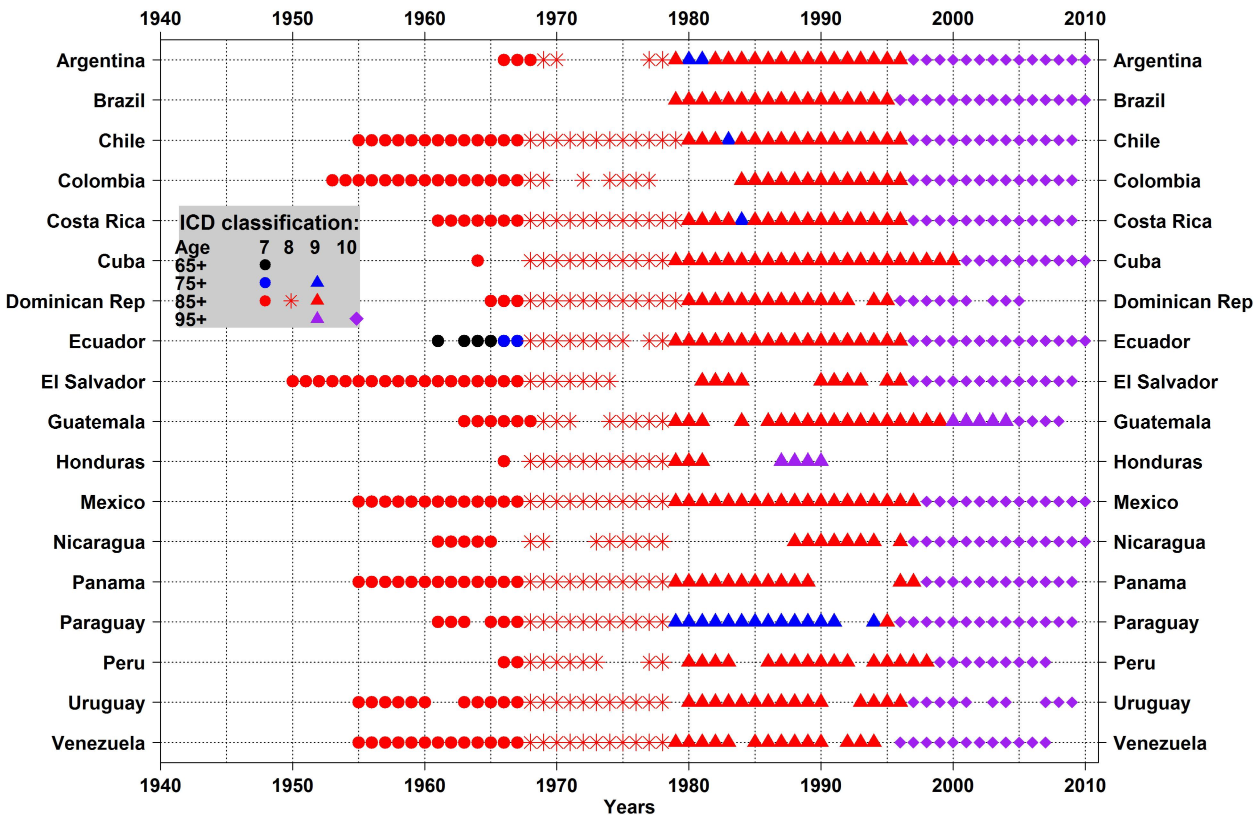
[1]Unless otherwise noted, all adjustments proposed for data by causes of death focus on mortality rates adjusted for relative completeness of death registration. The key assumption that makes the adjustment possible is that relative completeness of death counts is independent of causes of death.

Table 6.1: Groups of causes of deaths used in LAMBdA and associated ICD codes across multiple ICD classification systems

| **Main Groups** <br> **Specific Causes** | **ICD Classification** <br> **6th/7th: A list** | **8th: A list** | **9th: B list** | **10th: Detailed list** |
|---|---|---|---|---|
| 1. Neoplasms (all) | 044-059 | 045-060 | 08-14,16,17 | C00-C97 |
| 2. Respiratory | 044,049,050 | 045,050,051 | 08,10 | C00-C14,C32-C34 |
| 3. Digestive | 045-048 | 046-049 | 09 | C15-C26 |
| 4. Breast | 051 | 054 | 113 | C50 |
| 5. Circulatory Diseases | 070,079-086 | 080-088 | 25-30 | I00-I87,G45 |
| 6. Heart | 079-084 | 080-084 | 25-28 | I00-I02,I05-I09,I11,I20-I22,I24-I28,I30-I38,I40,I42,I44-I51 |
| 7. Hypertension | 083,084 | 082 | 26 | I10-I13 |
| 8. Cerebrovascular | 070 | 085 | 29 | I60-I64,I67,I69,G45 |
| 9. Arteriosclerosis | 085,086 | 086-088 | 030 | I70,I72-I74,I77-I78,I80-I87 |
| 10. Respiratory Diseases | 087-097 | 089-096 | 31,32 | J00-J98 |
| 11. Acute Upper Respiratory Infections | 087 | 089 | 31 | J00-J06 |
| 12. Influenza, Pneumonia Acute Bronchitis | 088-092 | 090-092 | 320-322 | J10-J16,J18,J20-J22 |
| 13. Chronic Bronchitis, Emphysema, Asthma | 093 | 093 | 323 | J40-J46 |
| 14. Digestive Diseases | 098-107 | 004-005,097-104 | 33-34,015 | K00-K92 |
| 15. Cirrhosis | 105 | 102 | 347 | K70,K73,K74 |
| 16. Ulcers | 099-100 | 098 | 341 | K25-K28 |
| 17. Diarrhea | 101,104 | — | — | — |
| 17. Diabetes Mellitus | 063 | 064 | 181 | E10-E14 |
| 18. Infectious Diseases | 001-043 | 001-003,006-044 | 01-07[a] | A00-B99 |
| 19. Respiratory TB | 001 | 006 | 020 | A15-A16 |
| 28. Malaria | 037 | 031 | 052 | B50-B54 |
| 29. Diarrhea | — | 005 | 016 | A09 |
| 20. Accidents, Homicide, Suicides | 138-150 | 138-150 | 47-56 | V01-Y89 |
| 21. Motor Vehicle Accidents | 138 | 138 | 471 | V02-04, V090-V092, V12-V14, V190-V192, V194-V196, V20-V28, V30-V45, V50-V56, V60-V79, V803-V805, V810, V811, V820, V821, V83-V86, V870-V878, V880-V890, V892 |
| 22. All other Accidents | 139-147 | 139-146 | 470-74,479, 48-53 | V01, V05, V06, V091, V093, V099, V10, V11, V15-V18, V193, V198, V199, V800-V02, V806-V09, V812-V819, V822, V824-V829, V879, V889, V891, V893, V899, W00-W009, X00-X09, Y85, Y86, V90-V99, W10-W45, W49-W60, W64-W70, W73-W81, W83-W94, W99, X10-X54, X57-X59, Y85, Y86 |
| 23. Homicides | 149 | 148 | 55 | X85-X99, Y00-Y09, Y35, Y871, Y890 |
| 24. Suicides | 148 | 147 | 54 | Y870, X60-X84 |
| 25. Senility, Ill Defined | 136,137 | 136,137 | 46 | R00-R99 |
| 26. Residual: Total - (1+5+10+14+17+18+20+25) | | | | |
| 27. Total =1+5+10+14+17+18+20+25+26 | | | | |
| 30. Waterborne | 002-005,012-014, 016, 034 | 001-003,007-010, 028 | 010,011,014, 016,022-025, 029,077 | A00-A08, A17-A19 |
| 31. Airborne | 001,017,018, 021,022,028, 031,032 | 006,015-017, 022-025 | 020,021,033-035, 040-042 | A15,A16,A36,A37,B05 |
| 32. Vectorborne | 024,033,036-042 | 011,026,030, 031,039-043 | 030,044,052, 072,074-076 | A20,A95,B50-B54,B65 |

[a] Excludes code 015

Figure 6.1: Cause of death data in LAC by country, year, ICD-classification and last age group reported.



the role of causes of death on the timing and pace of the mortality decline. Our general strategy for adjustments in LAMBdA is to start out with populations where the proportion of ill-defined death counts exceeds 10 percent by age, fit a linear regression model, and obtain estimates of the fraction of deaths due to each cause that are ill-defined (see below). We then re-allocate ill-defined deaths across causes of death by sex, age, country, and year, and obtain adjusted cause-specific mortality rates. Finally, for populations with small fraction of ill-defined deaths(less than 10 percent), we re-distribute these deaths proportionally according to the observed cause of death distribution by sex, age, country, and year.

Figure 6.2 displays LAC countries where the proportion of ill-defined deaths exceeded 10 percent. Shaded areas in the figure represent group of countries roughly defined by their stage of the demographic transition; the top part of the figure shows countries in a late stage while the bottom part corresponds to those that had an early transition. Symbols correspond to different age groups, triangles represent ages 10-14, dots represent ages 15-49 and stars represent those aged 50+. This figure shows that ill-defined deaths represent a higher fraction of all deaths in countries that experienced a late (top) and an intermediate (middle) onset of mortality decline but much less so in countries with early mortality decline (bottom). Moreover, ill-defined deaths are mostly concentrated at younger (triangles) and older adult ages (stars).

## 6.2 Correction for ill-defined deaths

Correction of ill-defined deaths can be accomplished in a number of ways. LAMBdA adjustment followed a procedure that combines both the usual "proportional distribution method", whereby ill-defined causes are allocated to well-defined groups according to the observed distribution in well defined groups, and a regression-based procedure. The former method is used for country-years where there are less than 10 percent of ill-defined deaths while the latter is used where there are more than 10 percent ill-defined deaths.

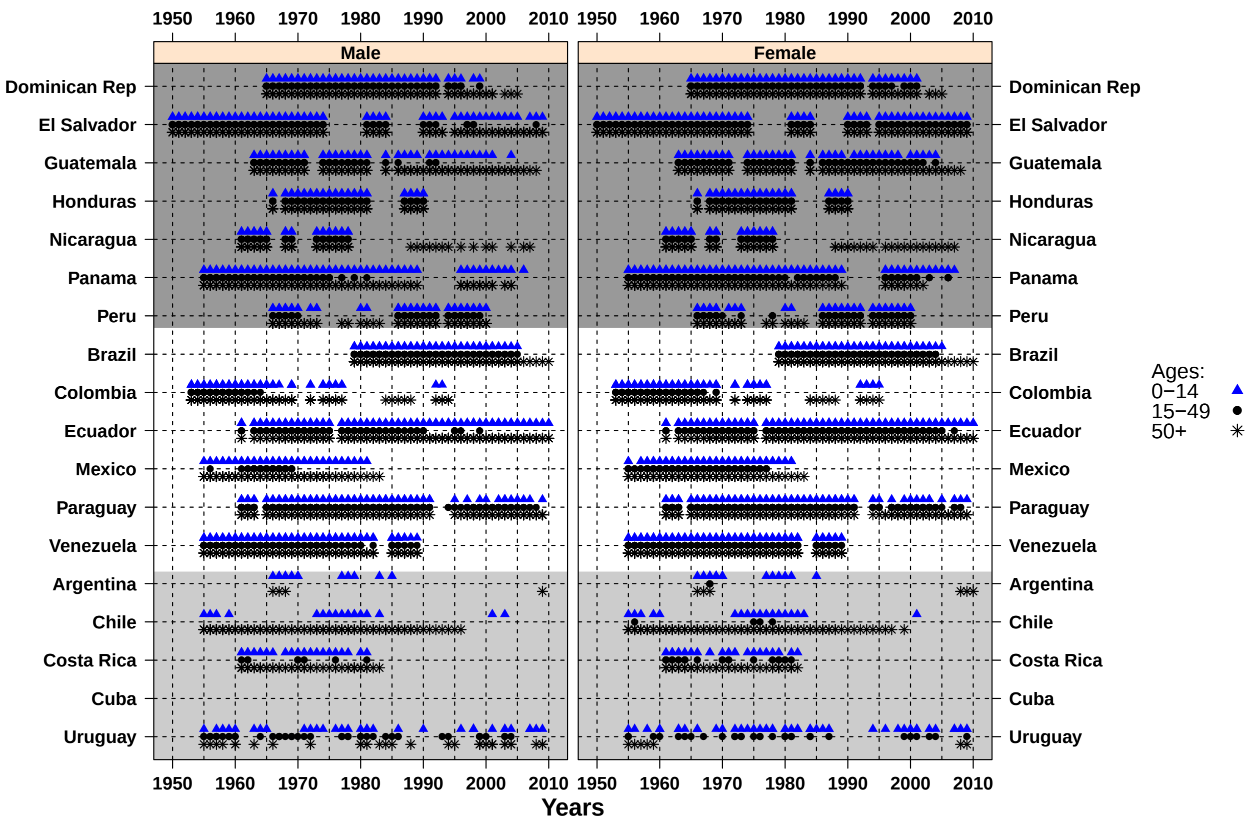
The regression-based approach proceeds as follows. Let $P^O(ill)$ be the probability of observing a death classified as 'ill-defined'; let $P(ill|j)$ be the probability that death due to cause $j$ will be classified as ill-defined category; let $P^T(j)$ and $P^O(j)$ be the true and observed probabilities, respectively, of a death being due to cause $j$. Then

$$\begin{aligned} P^O(ill) &= \sum_{j=1}^{k} P^T(j)P(ill|j) \quad \texttt{and} \\ P^O(j) &= P^T(j)(1-P(ill|j)) \end{aligned} \tag{6.2.1}$$

It then follows that

$$\begin{aligned} P^O(ill) &= \sum_{j=1}^{k} P^O(j)\frac{P(ill|j)}{1-P(ill|j)} \\ P^O(ill) &= \sum_{j=1}^{k} P^O(j)\beta_j \end{aligned} \tag{6.2.2}$$

Figure 6.2: LAC countries where the proportion of ill-defined deaths exceeds 10 pecent for Males and Females.




Note: Shaded areas indicate countries with ‘late’ (top), ‘intermediate’ (middle), and ‘early’ (bottom) demographic transition, respectively.

where $\beta_j = P(ill|j)/(1 - P(ill|j))$ represents the odds of a death due to cause $j$ being classified as ill-defined. Expression 6.2.2 illustrates a regression equation relating observed values of $P(ill)$ and of probabilities of death due to each cause, $P^O(j)$ for $j = 1, \dots, k$. We assume that the $\beta_j$'s are functions of (a) country, (b) time, and then convert the expression into a GLS model that can be estimated in a pooled cross-section time series with random and fixed effects. The estimation requires one constraint, namely, the constant should be equal to 0. In addition, it is important to use a careful definition of groups of causes of deaths, one that secures meaningful groups but also minimizes possibilities of very small counts that can produce noise and lead to high variability of estimates. We also include dummies for groups of countries rather than dealing with individual countries and dummies for periods instead of individual years. Finally, we control for total levels of mortality. The equation then becomes

$$\beta_j = \beta_{0j} + \sum_{\forall r} \beta_{rj} C_r + \sum_{\forall s} \beta_{sj} T_s + \theta_j \ln(D_T)$$

where $C$ and $T$ are dummies for group of countries and years, and $D_T$ is the total mortality rate. The regression model we estimate is as follows:

$$P^O(ill) = \sum_{j=1}^{k} \beta_{0j} P^O(j) + \sum_{j=1}^{k} \sum_{\forall r} \beta_{rj} P^O(j) C_r + \sum_{j=1}^{k} \sum_{\forall s} \beta_{sj} P^O(j) T_s + \sum_{j=1}^{k} \theta_j P^O(j) \ln(D_T)$$

Once we have estimates of $\beta_j$, say $\widehat{\beta}_j$, we calculate the fraction of deaths due to cause $j$ that are ill-defined using a 4 step process: (1) estimate $\widehat{P(ill|j)}$, (2) compute $\widehat{P^T(j)}$, (3) use (1) and (2) to estimate $\widehat{P(ill \cap j)}$, and (4) compute the final proportions by using the ratio $\widehat{P(ill \cap j)}/P^O(ill)$. These steps are summarize below:

1. *Probability that deaths due to cause $j$ will be classified as ill-defined category*, $P(ill|j)$. These values are estimated as $\widehat{P(ill|j)} = exp(\widehat{\beta}_j)/(1 + exp(\widehat{\beta}_j))$

2. *True probability of cause of death $j$*, $P^T(j)$. From 6.2.1, this value is estimated as

$$\widehat{P^T(j)} = \frac{P^O(j)}{1 - \widehat{P(ill|j)}} = \frac{P^O(j)}{1 - \widehat{P(ill|j)}} * \frac{\widehat{P(ill|j)}}{\widehat{P(ill|j)}} = \frac{P^O(j) * \widehat{Odds}}{\widehat{P(ill|j)}} = \frac{P^O(j) * \widehat{\beta}_j}{\widehat{P(ill|j)}}$$

3. *Probability of being ill-defined and cause $j$*, $P(ill \cap j)$. By definition, $P(ill|j) = P(ill \cap j)/P^T(j)$. It thus follows that $P(ill \cap j)$ can be estimated as $\widehat{P(ill \cap j)} = \widehat{P(ill|j)} * \widehat{P^T(j)}$.

4. *Fraction of deaths due to cause $j$ that are ill-defined.* These fractions are estimated as $\widehat{P(j|ill)} = \widehat{P(ill \cap j)}/P^O(ill)$.

Table 6.2: Cause of death classification, country groups, and time periods for estimation of cause-of-death models.

| Group | Causes of death |
|---|---|
| Cancers | Neoplasms, Respiratory Neoplasms, Digestive Neoplasms, Breast Neoplasms |
| Cardiovascular | Circulatory Diseases, Heart, Hypertension, Cerebrovascular, Arteriosclerosis |
| Respiratory | Respiratory Diseases, Acute Upper Respiratory Infections, Influenza |
| | Pneumonia, Acute Bronchitis, Chronic Bronchitis, Emphysema, Asthma |
| Digestive/diabetes | Digestive Diseases, Cirrhosis, Ulcers, Diabetes Mellitus |
| Infections | Infectious Diseases |
| Accidents | Accidents, Homicides, Suicides |
| Residual | All other causes |
| **Time periods** | **Years** |
| 1950-1969 | 1950-1969 |
| 1970-1989 | 1970-1989 |
| 1990-2010 | 1990-2010 |
| **Mortality Regime** | **Countries** |
| Laggard | Dominican Republic, El Salvador, Guatemala, Honduras, Nicaragua |
| | Panama and Peru. |
| Intermediate | Brazil, Colombia, Ecuador, Mexico, Paraguay, Venezuela |
| Forerunner | Argentina, Chile, Costa Rica, Cuba, Uruguay |

"Mortality Regime" refers to a three-category classification of countries' mortality decline: early ("forerunners"), late("laggards") and intermediate.

An important caveat is on order. Because we will estimate the equation via GLS, the predicted values will not be equal to the observed values. In other words, the following inequality will prevail:

$$\widehat{P(ill)} = \sum_{j=0}^{k} P^O(j)\widehat{\beta}_j \neq P^O(ill)$$

To adjust for this we simply *normalize* the estimated proportions so the vector $\widehat{P(j|ill)}$ has length 1. That is, we compute: $\widehat{P(j|ill)}/\sum_{j=1}^{k}\widehat{P(j|ill)}$.

## 6.2.1 Empirical estimates

All models were estimated using group of causes, groups of countries, and time periods to increase the robustness of estimates. The classifications are shown in Table 6.2.

## 6.2.2 Statistical modeling

We fitted 3 models as follows:

1. Basic:

$$M_{illd} = \beta_{0j}M_j + \sum_{\forall k}\beta_k C_r * M_j + \sum_{\forall m}\beta_m T_s * M_j \tag{6.2.3}$$

2. Basic+ln($CMR_T$):

$$M_{illd} = \texttt{Basic} + \theta_j M_j * \ln(CMR_T) \tag{6.2.4}$$

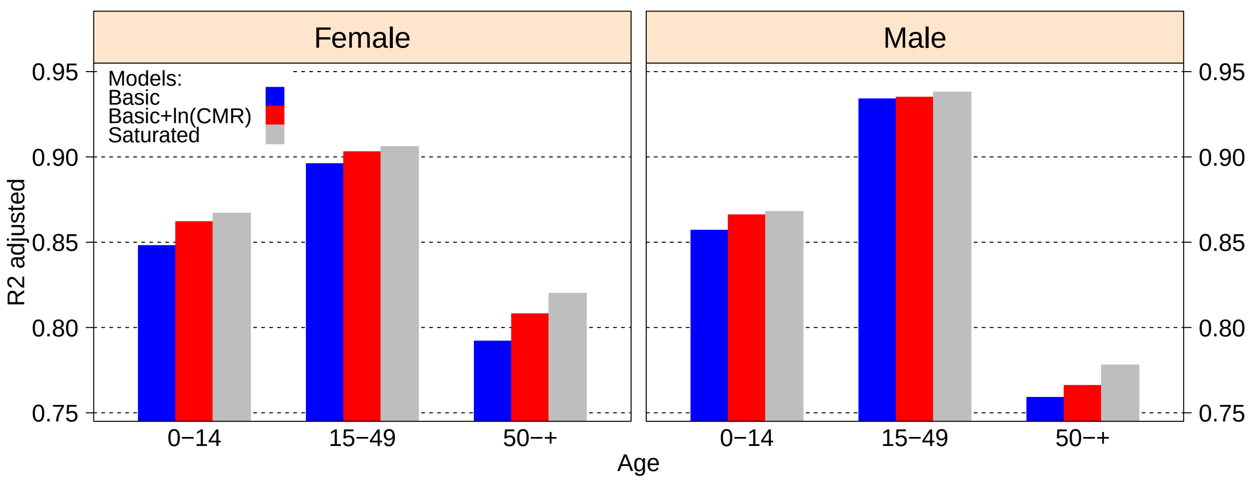
3. Saturated:

$$M_{illd} = \texttt{Basic} + \theta_j M_j * \ln(CMR_T) + \texttt{all interactions} \tag{6.2.5}$$

where $M_{illd}$ and $M_j$ are mortality rates for ill-defined and for cause of death $j$, respectively; $C_r$ and $T_s$ are a set of dummy variables for country-region and time-period, respectively; and $CMR_T$ is the crude mortality rate at time $T$.

### 6.2.3 Results

The figure below shows $R^2$ adjusted in the age-sex specific models shown above.




In order to improve model fitting, we also estimate models of the form Basic+ln($CMR_T$) shown in equation 6.2.4. Table 6.3 shows corresponding coefficient estimates. These models explain at least 85 percent of variance among women and males at ages below 50 but in all cases they fit less well at older ages. We then use the estimated coefficients to compute conditional probabilities and the fraction of deaths due to cause $j$ that are ill-defined by country-year-sex-age.

For comparison purposes, we also computed similar fractions using a standard approach whereby ill-defined deaths are all allocated according to the observed distribution of well-defined causes of death. To simplify the comparisons between the two approaches, we plotted these proportions by groups of cause of death, country, year and age for males. Figures 6.3–6.9 displays results for cancers, cardiovascular, respiratory, digestive/diabetes, infections, accidents and residual causes, respectively. By and large, the two reallocation procedures yield very different results and these differences vary by age groups. Thus, at ages $< 15$ the regression-based method assigns more ill-defined deaths to cancers, CVD, and respiratory diseases whereas at ages 15-49 assigns a greater fraction of deaths to digestive/diabetes conditions. Finally, at ages >50, the regression-based approach assigns more deaths to respiratory diseases and about the same fraction as the traditional approach to CVD, and residual causes of death.

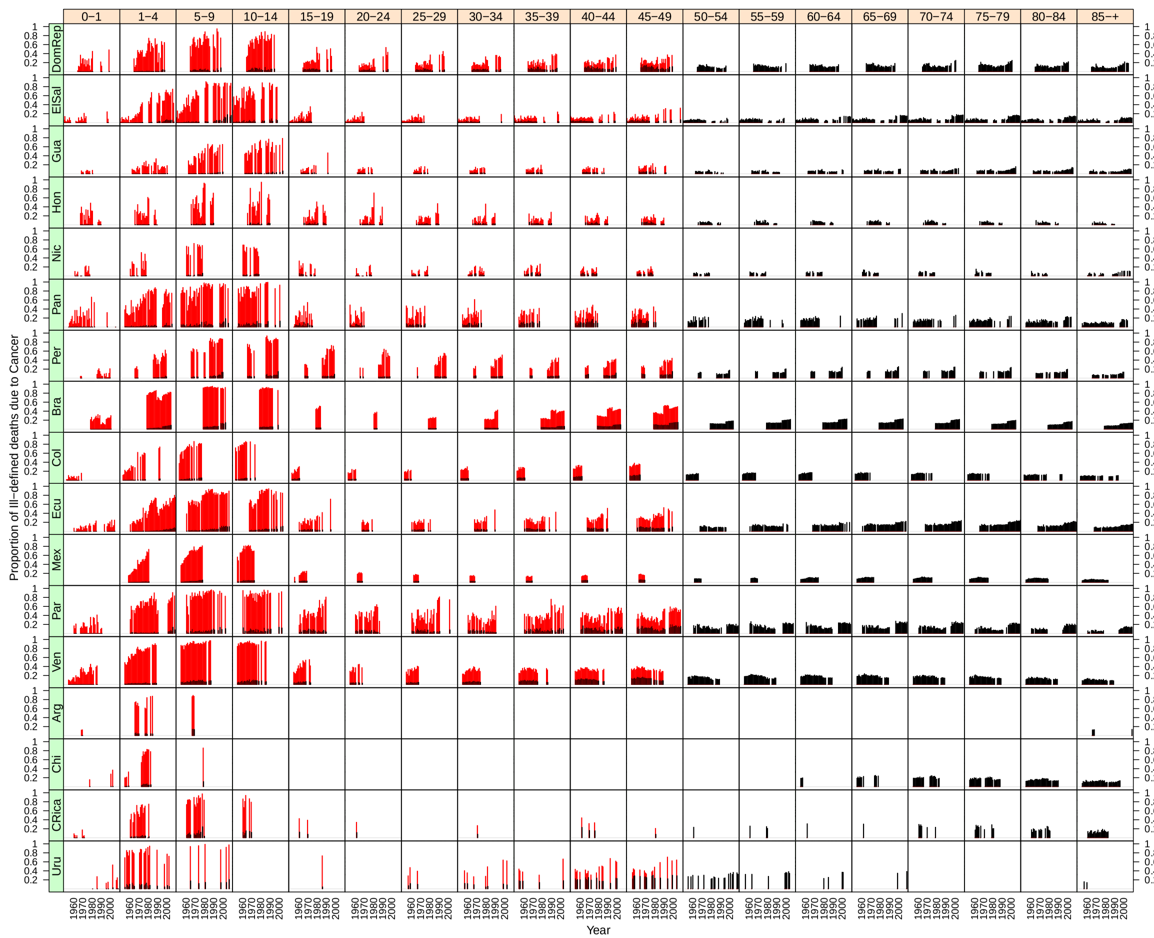
Table 6.3: Cause-of-death coefficient estimates from Model 6.2.4 by age and sex.

| | Men | | | | | | Women | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Age 0-14 | | Age 15-49 | | Age 50+ | | Age 0-14 | | Age 15-49 | | Age 50+ | |
| variable | coef | p-value | coef | p-value | coef | p-value | coef | p-value | coef | p-value | coef | p-value |
| cancer | 40.02 | 0.29 | 2.12 | 0.06 | -3.50 | 0.10 | 55.69 | 0.13 | -1.29 | 0.03 | 0.41 | 0.85 |
| cvd | 12.76 | 0.20 | 0.83 | 0.26 | 2.02 | 0.01 | 23.35 | 0.02 | 7.69 | 0.00 | 1.15 | 0.09 |
| resp | 3.70 | 0.00 | -1.77 | 0.06 | 2.84 | 0.03 | 2.57 | 0.00 | -0.98 | 0.27 | 6.39 | 0.00 |
| dig_diab | 0.38 | 0.96 | 1.57 | 0.01 | -5.86 | 0.00 | 19.37 | 0.00 | -2.09 | 0.01 | -4.35 | 0.02 |
| infe | -3.33 | 0.00 | 0.13 | 0.80 | -10.88 | 0.00 | -3.15 | 0.00 | 0.32 | 0.59 | -20.08 | 0.00 |
| acc | -1.41 | 0.80 | 0.01 | 0.94 | 0.10 | 0.97 | -4.22 | 0.53 | 2.83 | 0.00 | 24.77 | 0.00 |
| res | -0.97 | 0.00 | -0.22 | 0.19 | 1.89 | 0.00 | -1.99 | 0.00 | 0.01 | 0.96 | -1.04 | 0.03 |
| time50Xcancer | 3.68 | 0.57 | -0.54 | 0.01 | -1.95 | 0.00 | 8.68 | 0.20 | 0.65 | 0.00 | -1.35 | 0.01 |
| time50Xcvd | -7.30 | 0.01 | 0.07 | 0.65 | -0.27 | 0.04 | -8.32 | 0.00 | -1.61 | 0.00 | -0.17 | 0.19 |
| time50Xresp | -2.43 | 0.00 | -0.21 | 0.60 | 0.79 | 0.00 | -1.18 | 0.00 | -0.33 | 0.31 | 0.18 | 0.54 |
| time50Xdig_diab | 1.02 | 0.88 | 0.76 | 0.00 | 2.99 | 0.00 | -17.01 | 0.01 | 1.43 | 0.00 | 1.26 | 0.02 |
| time50Xinfe | 1.53 | 0.00 | -0.21 | 0.29 | 0.10 | 0.88 | 1.22 | 0.00 | -0.21 | 0.34 | 0.53 | 0.41 |
| time50Xacc | 1.51 | 0.21 | 0.04 | 0.38 | 0.60 | 0.43 | 0.42 | 0.77 | 0.05 | 0.76 | 1.58 | 0.36 |
| time50Xres | 0.19 | 0.02 | 0.34 | 0.00 | 1.84 | 0.00 | 0.38 | 0.00 | 0.38 | 0.00 | 2.61 | 0.00 |
| time70Xcancer | 10.97 | 0.07 | -0.70 | 0.00 | -1.19 | 0.00 | 4.90 | 0.41 | -0.10 | 0.46 | -0.69 | 0.09 |
| time70Xcvd | 0.79 | 0.74 | 0.19 | 0.12 | -0.05 | 0.64 | -1.60 | 0.51 | -0.11 | 0.43 | 0.16 | 0.13 |
| time70Xresp | -2.26 | 0.00 | -0.44 | 0.28 | 0.24 | 0.22 | -0.80 | 0.00 | -0.56 | 0.07 | -0.26 | 0.23 |
| time70Xdig_diab | 1.89 | 0.78 | 0.60 | 0.00 | 2.38 | 0.00 | -16.13 | 0.01 | 0.50 | 0.15 | 0.79 | 0.13 |
| time70Xinfe | 1.48 | 0.00 | 0.12 | 0.54 | 0.56 | 0.36 | 1.16 | 0.00 | 0.18 | 0.39 | 0.54 | 0.35 |
| time70Xacc | -1.16 | 0.25 | -0.03 | 0.46 | -0.65 | 0.29 | -0.75 | 0.49 | -0.16 | 0.20 | 0.44 | 0.75 |
| time70Xres | 0.02 | 0.76 | 0.07 | 0.09 | 0.07 | 0.28 | 0.03 | 0.68 | 0.12 | 0.03 | 0.04 | 0.51 |
| earlyXcancer | -6.93 | 0.20 | -0.54 | 0.04 | 1.93 | 0.00 | -6.06 | 0.25 | 0.06 | 0.83 | 2.65 | 0.00 |
| earlyXcvd | -0.35 | 0.90 | 0.07 | 0.75 | -0.24 | 0.02 | 0.41 | 0.89 | -0.39 | 0.34 | -0.20 | 0.04 |
| earlyXresp | 0.66 | 0.04 | 0.59 | 0.40 | 0.07 | 0.71 | 0.30 | 0.32 | 1.14 | 0.20 | 0.48 | 0.01 |
| earlyXdig_diab | -0.60 | 0.00 | -0.77 | 0.23 | -2.92 | 0.00 | -0.48 | 0.00 | -0.75 | 0.15 | -2.77 | 0.00 |
| earlyXinfe | -0.74 | 0.00 | -0.20 | 0.51 | -0.90 | 0.33 | -0.78 | 0.00 | -1.17 | 0.02 | -3.45 | 0.00 |
| earlyXacc | 0.44 | 0.71 | -0.06 | 0.56 | 2.48 | 0.02 | 1.08 | 0.43 | 0.21 | 0.47 | -0.39 | 0.77 |
| earlyXres | 0.03 | 0.74 | 0.13 | 0.25 | -0.49 | 0.00 | 0.14 | 0.08 | -0.29 | 0.06 | -0.03 | 0.72 |
| interXcancer | 5.88 | 0.27 | 0.07 | 0.69 | 1.36 | 0.00 | 10.02 | 0.03 | 0.91 | 0.00 | 3.49 | 0.00 |
| interXcvd | -4.92 | 0.02 | 0.07 | 0.43 | 0.14 | 0.08 | -5.37 | 0.00 | -1.04 | 0.00 | -0.19 | 0.01 |
| interXresp | -0.62 | 0.00 | 1.00 | 0.00 | 0.66 | 0.00 | -0.89 | 0.00 | 0.96 | 0.00 | 2.26 | 0.00 |
| interXdig_diab | 0.42 | 0.00 | -1.24 | 0.00 | -1.13 | 0.00 | 0.62 | 0.00 | -1.64 | 0.00 | -2.14 | 0.00 |
| interXinfe | 0.24 | 0.01 | 0.24 | 0.01 | -1.80 | 0.00 | 0.51 | 0.00 | 0.15 | 0.14 | -5.53 | 0.00 |
| interXacc | -0.74 | 0.35 | -0.09 | 0.00 | -0.64 | 0.24 | -1.40 | 0.13 | 0.20 | 0.03 | -4.49 | 0.00 |
| interXres | 0.01 | 0.81 | 0.09 | 0.00 | -0.21 | 0.00 | -0.07 | 0.06 | -0.31 | 0.00 | 0.02 | 0.74 |
| ln_cmrXcancer | 7.80 | 0.29 | 0.27 | 0.21 | -0.54 | 0.18 | 11.07 | 0.10 | -0.22 | 0.03 | 0.57 | 0.14 |
| ln_cmrXcvd | 2.85 | 0.12 | 0.14 | 0.32 | 0.31 | 0.02 | 4.17 | 0.01 | 1.27 | 0.00 | 0.15 | 0.21 |
| ln_cmrXresp | 0.31 | 0.00 | -0.28 | 0.10 | 0.58 | 0.03 | 0.33 | 0.00 | -0.02 | 0.89 | 1.30 | 0.00 |
| ln_cmrXdig_diab | 0.15 | 0.09 | 0.26 | 0.02 | -1.00 | 0.00 | 0.36 | 0.00 | -0.50 | 0.00 | -1.11 | 0.00 |
| ln_cmrXinfe | -0.51 | 0.00 | -0.04 | 0.64 | -2.54 | 0.00 | -0.51 | 0.00 | -0.04 | 0.72 | -4.75 | 0.00 |
| ln_cmrXacc | -0.21 | 0.84 | -0.02 | 0.28 | -0.04 | 0.94 | -0.73 | 0.54 | 0.53 | 0.00 | 4.23 | 0.00 |
| ln_cmrXres | -0.22 | 0.00 | -0.07 | 0.03 | 0.27 | 0.00 | -0.40 | 0.00 | -0.06 | 0.15 | -0.23 | 0.01 |

Abbreviations used: resp= respiratory; dig_diab= digestive diseases and T2D; inf= infectious diseases; acc= accidents; res= residual; time50=1 if year is included in period 1950-1969; time70=1 if year is included in period 1970-1989 (left out category=years after 1989; early=1 if country experiences an early onset of mortality decline; inter=1 if country's timing of onset of mortality decline is intermediate between early and late(left out category=countries with a late onset mortality decline); CMR=Crude Mortality Rate.

Source: Data from LAMBdA

Figure 6.3: Estimated proportion of ill-defined deaths due to cancer for Males by country, age, and year. Red values are estimated using our approach and black values come from a proportional distribution



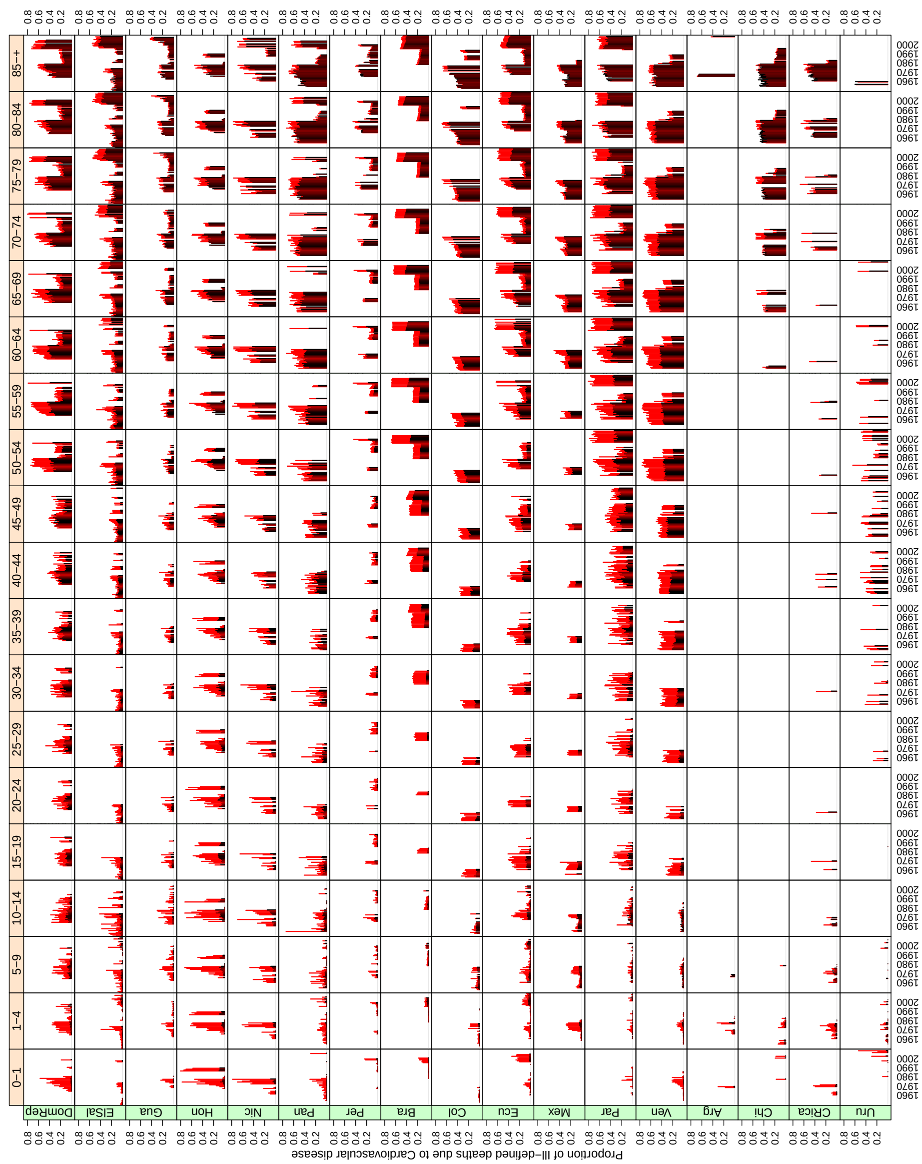



Figure 6.4: Estimated proportion of ill-defined deaths due to cardiovascular diseases for Males by country, age, and year. Red values are estimated using our approach and black values come from a proportional distribution

Figure 6.5: Estimated proportion of ill-defined deaths due to respiratory diseases for Males by country, age, and year. Red values are estimated using our approach and black values come from a proportional distribution

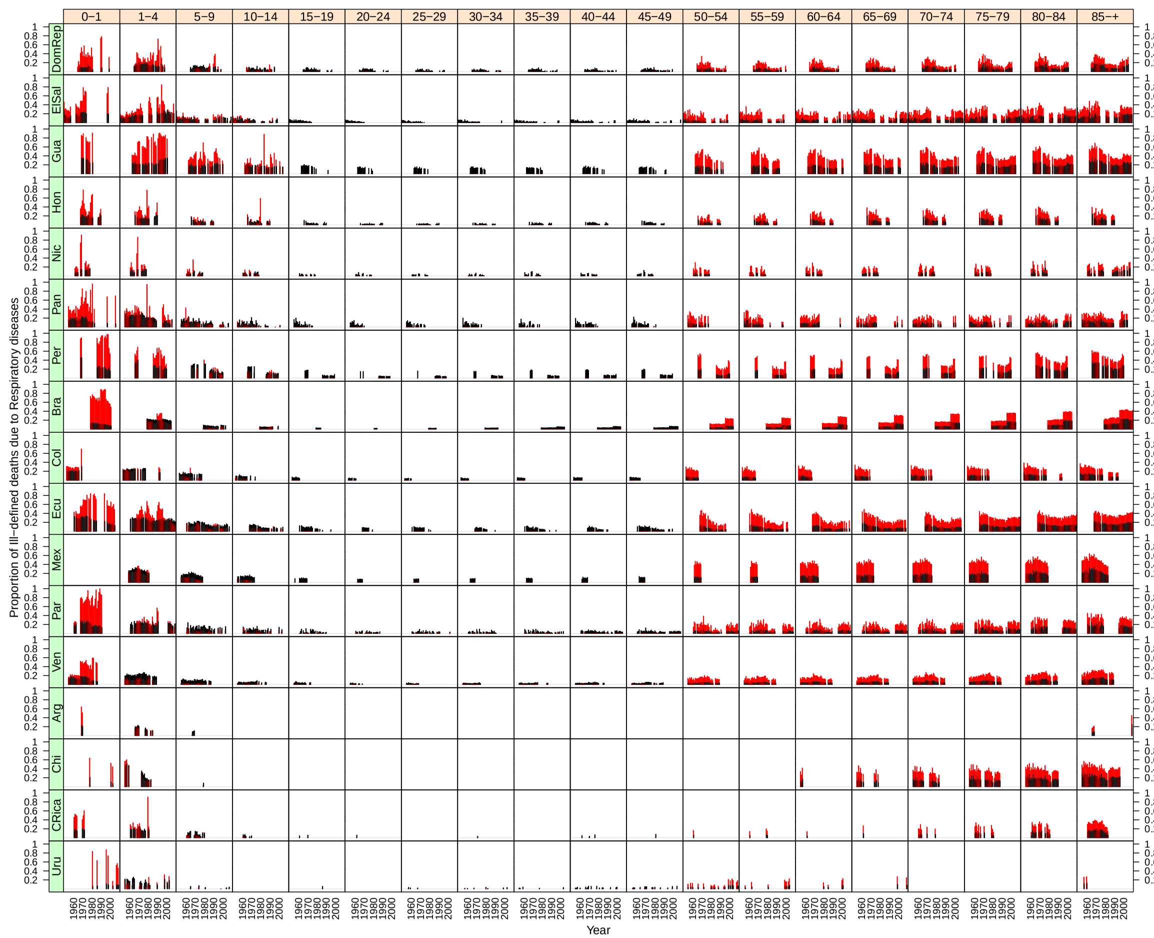
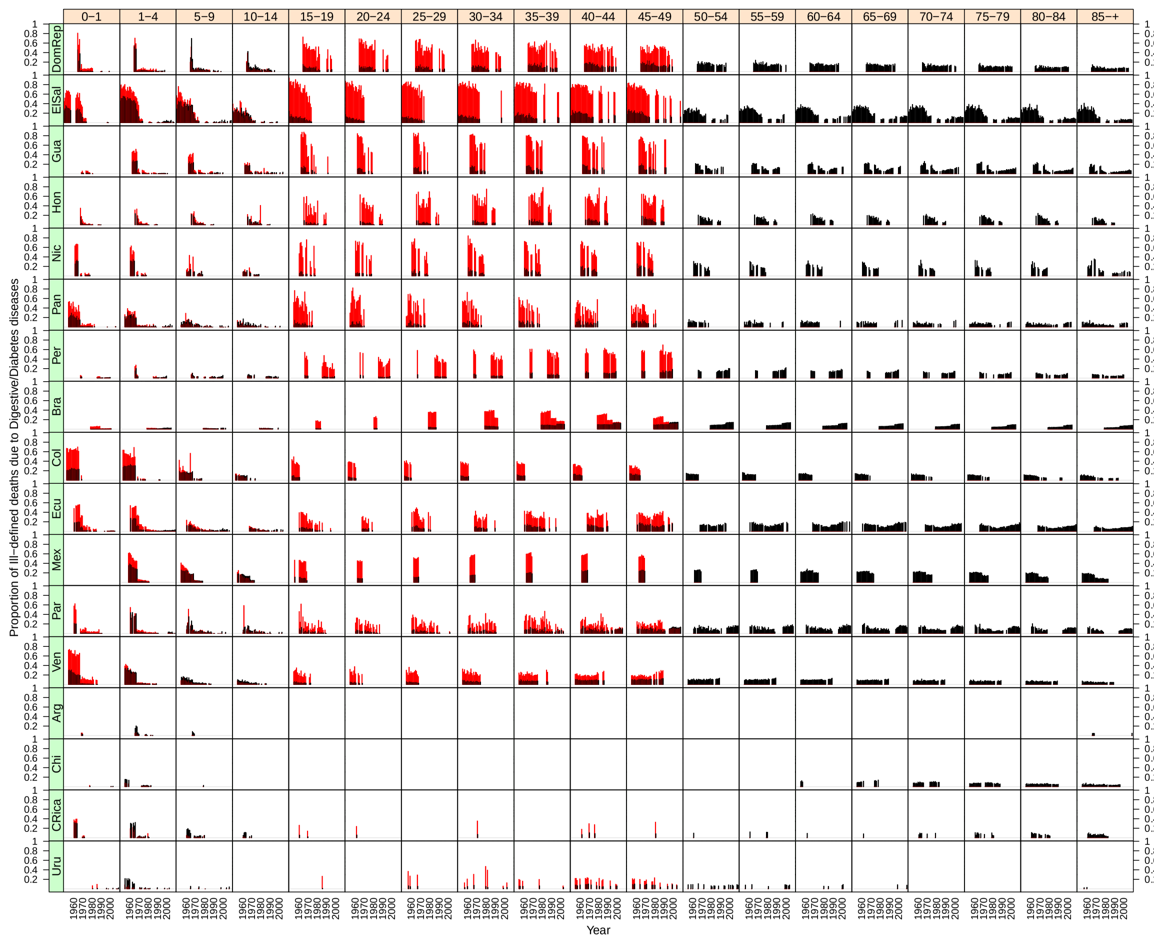


Figure 6.6: Estimated proportion of ill-defined deaths due to digestive/diabetes diseases for Males by country, age, and year. Red values are estimated using our approach and black values come from a proportional distribution



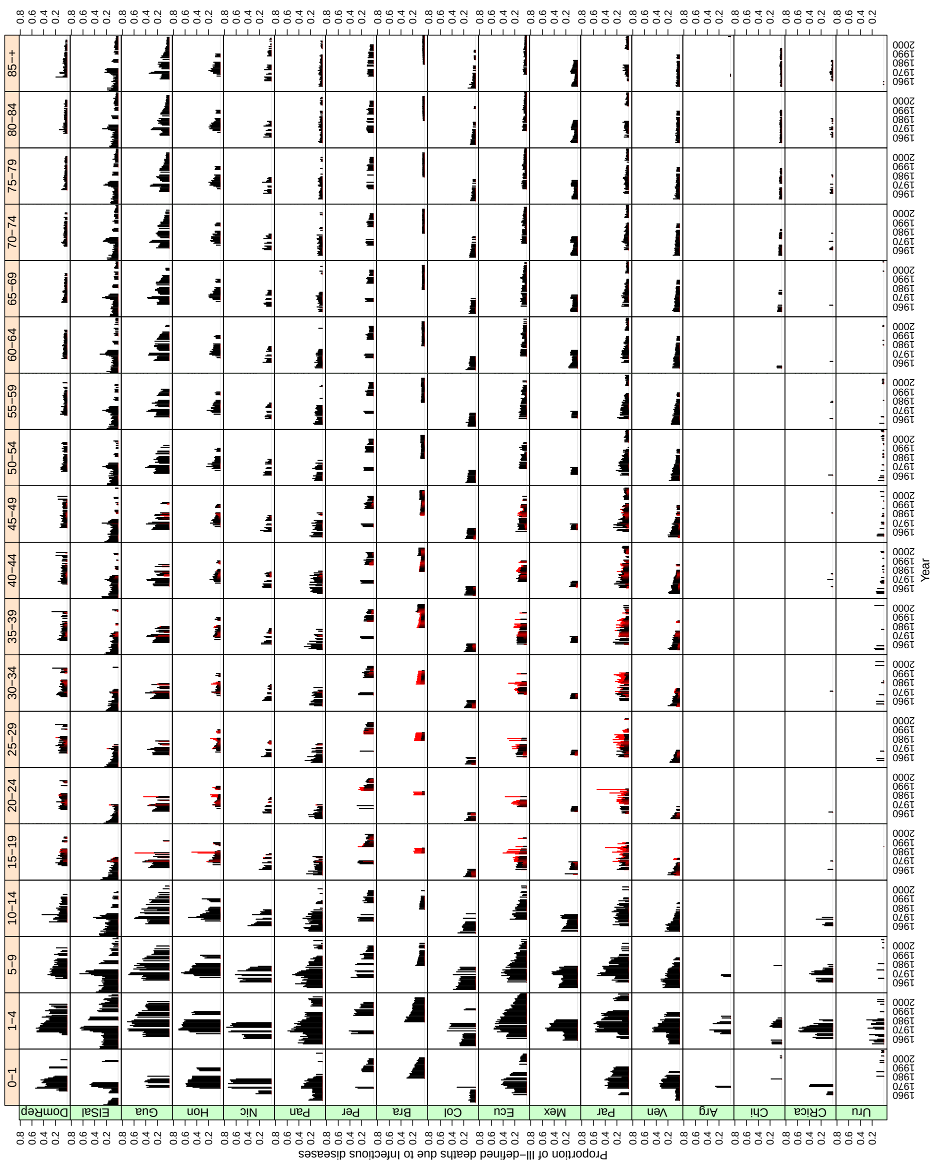



Figure 6.7: Estimated proportion of ill-defined deaths due to infectious diseases for Males by country, age, and year. Red values are estimated using our approach and black values come from a proportional distribution

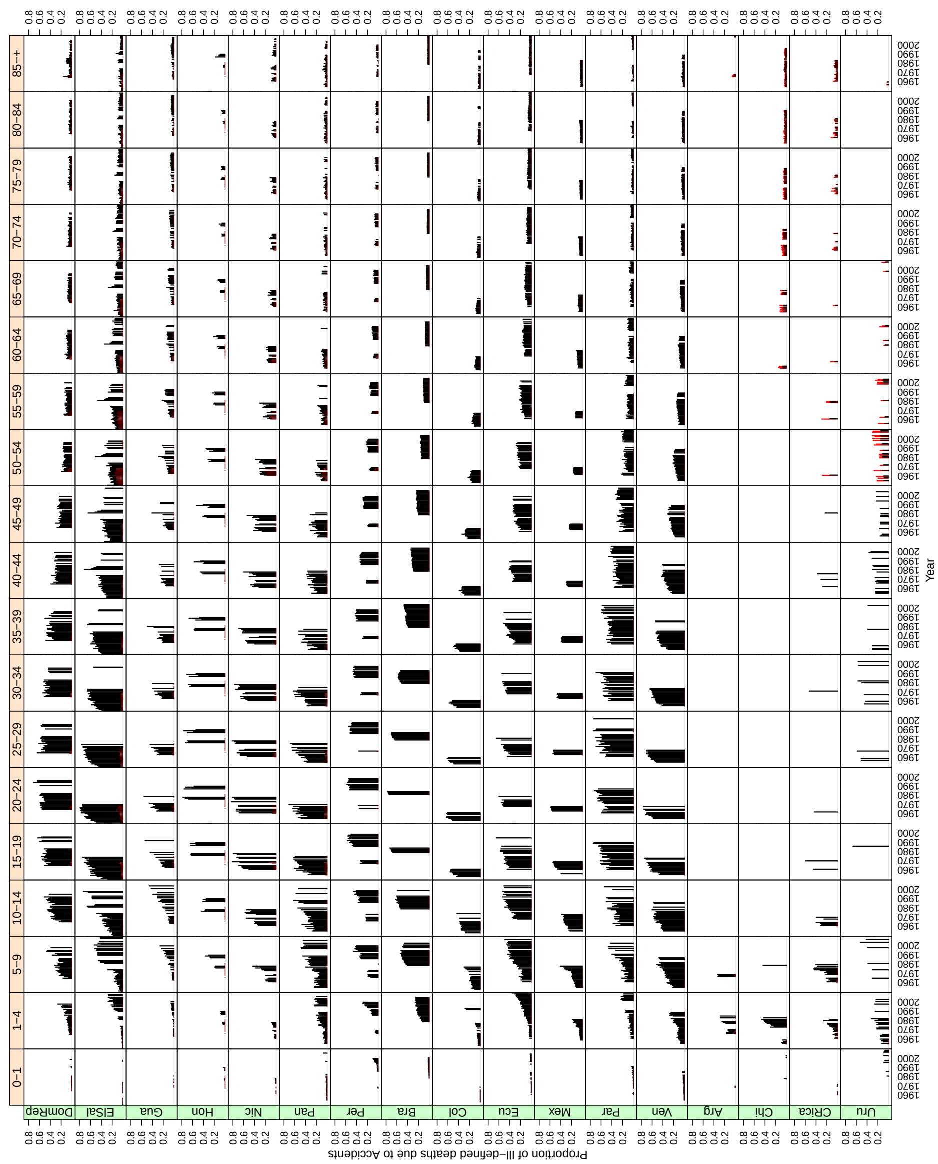



Figure 6.8: Estimated proportion of ill-defined deaths due to accidents for Males by country, age, and year. Red values are estimated using our approach and black values come from a proportional distribution

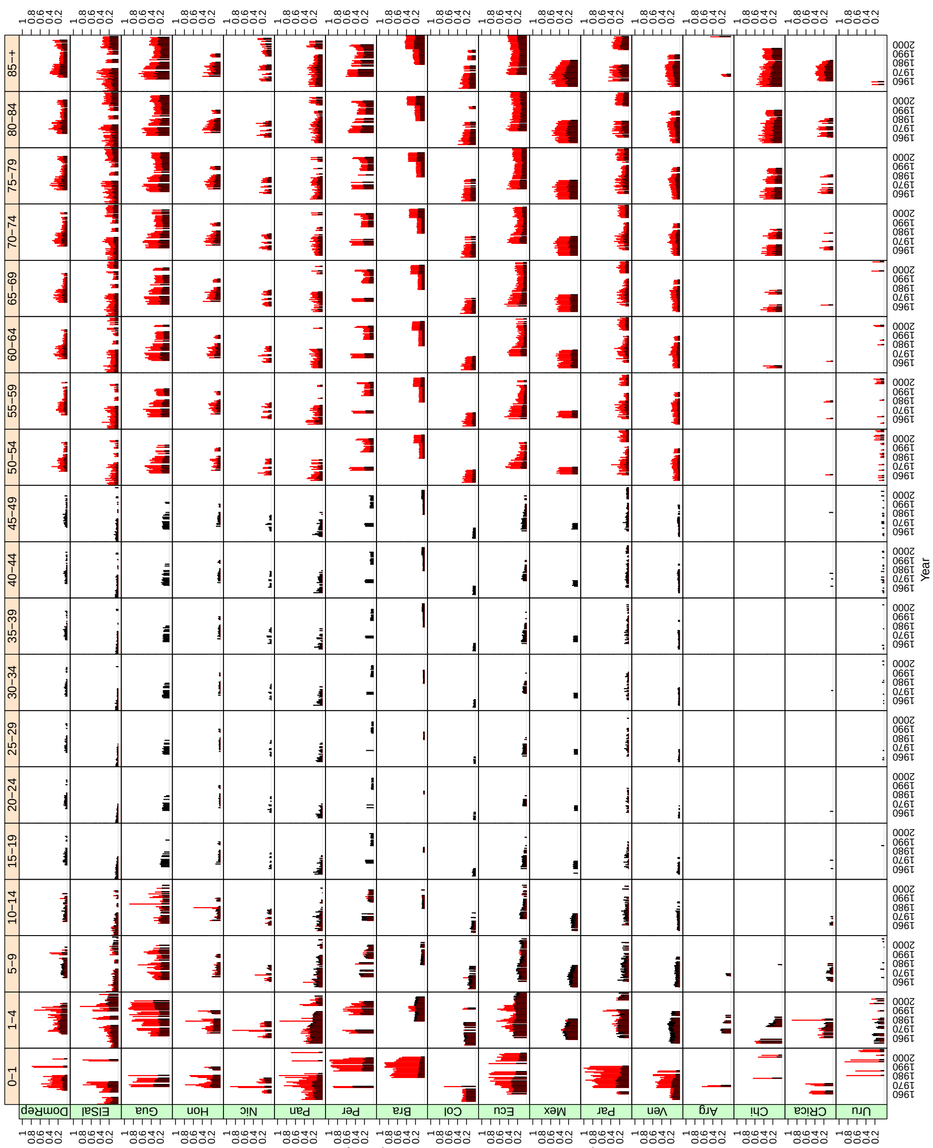



Figure 6.9: Estimated proportion of ill-defined deaths due to residual causes for Males by country, age, and year. Red values are estimated using our approach and black values come from a proportional distribution.

# Chapter 7

# Predicting death counts and mortality rates at older adult ages using a Logistic model

## 7.1 Introduction

All life tables included in LAMBdA end with an open age group 85+. The decision to censor the life tables at age 84 is a consequence of two constraints, one a product of the nature of the data and the other a result of the methodology followed throughout to adjust adult mortality rates. First, it is only after 1980 that census and vital statistics offices have made available, in any shape or form, counts of deaths and or population by single year of ages for populations older than 84 years. Second, as noted before, LAMBdA life tables depend on adjustments for relative completeness of death registration and age misreporting. The former rests on assumptions of age invariance of relative completeness that may be questioned at very old ages. Indeed, assuming that there is identity of relative completeness in the age interval 5-59 and 60+ is already a bit of a stretch, although it might not have discernible consequences. Adjustments for age misreporting, on the other hand, rely on a procedure whose robustness at very old ages is unknown. Both population and death counts contract significantly at very old ages and observed figures are subject to massive noise that could overwhelm any correction or adjustment.

While the decision to censor life tables at age 84 considerably reduces the margin of errors of estimates of old age mortality, it also deprives investigators of the material to perform tests about senescence and longevity that require more granular information at ages above 85. For this reason we undertook a number of tests to assess the behavior of different techniques that expand the age coverage of the life tables by extrapolating rates above age 84.

## 7.2 Methods to estimate death distribution for ages 85 and older

Estimates of the distribution of deaths for ages above 85 by single years of age up to age 100+ could support the study of mortality at very old ages, testing of theories about compression of mortality and, more generally, and improve comparability with mortality statistics in high-income countries. However, the raw data and adjustment procedures used in LAMBdA are not suitable to produce direct estimates of mortality by single years of ages at very old ages. For this reason we must rely on indirect approaches. In this section we report results from consistency checks of mortality rates at older ages using two methods based on a modified version of a standard (logistic) approach for older age mortality. The approach was independently developed by Himes et al. (1994) and Kannisto and colleagues (Kannisto, 1994; Kannisto et al., 1994), and it is commonly referred to as the 'Kannisto model'. The model can be thought of as a linear function of age of the logit of the mortality rates (Himes et al., 1994) or, alternatively, as a logistic function of age-specific mortality rates at ages 80+ (Kannisto, 1994; Kannisto et al., 1994).

In this report we illustrate the use of this approach by applying them to unadjusted (method 1) and adjusted (method 2) older adult mortality data in the Latin American countries included in LAMBdA. The aim of these tests is to assess the performance of each method, to identify, if any are found, singularities that could signal anomalies in observed or adjusted death counts or death rates, and to assess the robustness of estimated death rates at older ages to these anomalies.

Our main take away message is that it is very important *to correct mortality rates at adult ages for both completeness of death registration and age misreporting before doing any analyses with mortality rates at very old ages, regardless of how these may be estimated*. In fact, we find that using unadjusted death rates in the age interval 60-84 to estimate mortality rates at ages 85-99 –as is conventionally done when working with high quality data–leads to downward biases that exceed 30 percent . The result of this is that life expectancy at older ages will contain sizable upward biases. Since in most countries the accuracy of vital statistics and census counts increases over time, the use of unadjusted death rates will produce potentially large distortions of time trends.

## 7.3 Extrapolating mortality rates at ages 85+

Because available data on deaths typically come in five-year age groups we first disaggregate them into single years of age using Sprague multipliers.[1] We then employ methods 1 and 2 to extrapolate mortality rates above age 85, compare the results that obtain, and assess the robustness of our extrapolation method.

[1] see Chapter 5.

### 7.3.1 Method 1: predicting cumulative death counts and death rates using *unadjusted* death count data

The first method estimates counts of deaths for ages 85 to 110+ by fitting a logistic model to the *cumulative* proportion of deaths from some arbitrary baseline age to the last single age for which there is information, namely, the open age group or $\omega = 85$. We define a baseline age $x_0$ to ensure we have at least 20 points to fit the model and compute the (log of) cumulative proportion of deaths in single years of age ($d_x$) as ($D(x) = \Sigma_{x_o}^{x} d_x / \Sigma_{x_o}^{\omega} d_x$). We then fit a two parameter model of the form:

$$\ln(D(x)) = (1/b)\ln\left(\frac{1+a}{1+a\exp(b(x-x_0))}\right), \quad x \in [x_0, 84]. \tag{7.3.1}$$

Parameters $a$ and $b$ are estimated by country-year and then used to predict the number of deaths up to age 100+ by reversing the above process. That is, we predict the values of D(x), $\widehat{D(x)}$, from equation (7.3.1), multiply these quantities by the observed value of the total number of deaths above $x_0$, and derive each of the $\widehat{d_x}$ values from successive differences between predicted accumulated number of deaths. These quantities are then used to compute estimates of the death rate at ages $(x, x+1)$ as $\widehat{\mu_x} = \widehat{d_x}/\Sigma_x^{\omega}\widehat{d_x}$.

An important caveat is in order. With a handful of exceptions, the observed number of deaths in the populations we study is lower than what it should be due to imperfect completeness of death counts in vital statistics. Furthermore, the observed death counts can also be influenced by age misreporting mostly in the form of age overreporting. This implies that the observed values of $D_x$ will be in error and so will the estimated parameters of (7.3.1). However, in cases when age misreporting can be ignored and the fraction of unreported deaths is age-invariant, the observed values of $D(x)$ will be correct and so will the parameter estimates of (7.3.1). Unfortunately, however, this does not ensure a safe escape from errors. This is because the *predicted* values $\widehat{d_x}$ will be incorrect since to obtain them we must multiply the values $\widehat{D(x)}$ by the *observed count* of accumulated deaths above age $x_0$, a quantity distorted by lack of completeness of death registration (and misreporting when this exists). As a consequence, all life table statistics computed from the estimated parameters will be in error even if the parameter estimates in (7.3.1) are unbiased. The bottom line is that we cannot employ this variant of Kannisto's model to estimate mortality at very old ages without engendering substantial errors.

### 7.3.2 Method 2: predicting death rates using *adjusted* death rates

To apply the second method we use adjusted (for relative completeness and age misreporting) death rates up to the maximum available age but excluding the open age group, for most countries in LAMBdA this corresponds to maximum age 84. We then fit the following model (Himes et al., 1994):

$$\texttt{logit}(m(x)) = \ln(a) + bx, \quad x \in [x_0 - 20, 84]. \tag{7.3.2}$$

where $m(x)$ are age-specific mortality rates. Estimates of parameters $a$ and $b$ are then used to compute single age-specific mortality rates for ages up to 100. If the adjustment procedures are correct, these predicted and extrapolated (beyond age 85) rates should reflect well the age patterns and levels of old age mortality. Because the values of $m(x)$ can be adjusted for relative completeness and age misreporting, this method has the potential to produce consistent parameter estimates and the extrapolated values above age $\omega = 85$ should not be influenced either by errors in completeness or age misreporting.[2]

## 7.4 Comparability of methods

### 7.4.1 Comparing mortality rates

We assess differences between predicted values from fitted models to unadjusted death counts (first approach) and to adjusted data (second approach). To compare the two sets of predicted values we compute the absolute value of relative differences by age, starting at age 85, and then average these differences over age:

$$\begin{aligned} \texttt{Abs.rel.diff.m(x)} &= \left| \frac{m(x)_{unadjusted} - m(x)_{adjusted}}{m(x)_{adjusted}} \right|, \quad x \in [85, 99] \qquad (7.4.1) \\ \overline{\texttt{Abs.diff.m}} &= \texttt{mean(Abs.rel.diff.m(x))} \end{aligned}$$

### 7.4.2 Comparing estimates of life expectancy at ages 60, 75 and 85

We use estimates of mortality rates by age to compute life tables for ages 0 to 100 using standard demographic techniques (Preston et al., 2001). We compare life expectancy estimates at ages 60, 75, and 85 to highlight the impact of using uncorrected data when estimating average length of life in older adults in Latin American countries.

## 7.5 Results

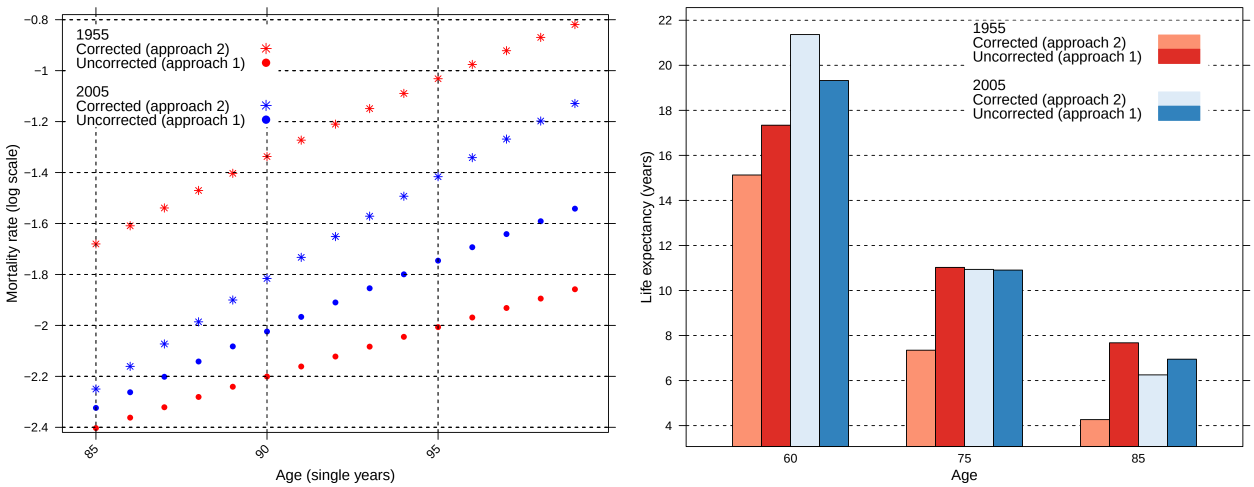
### 7.5.1 Application of extrapolation methods in LAMBdA

In the case of LAMBdA life tables we fitted equations (7.3.1) and (7.3.2) starting at age 60. As an example of this approach, Figure 7.1 shows results for Mexican women and pivotal years in 1955 and 2005. Panel a shows mortality rates and panel b displays life expectancy at ages 60, 75 and 85. Thus, panel a shows underestimation of mortality rates at ages 85+ when using unadjusted data (in red) relative to estimates using fully adjusted death rates (in blue). It is clear that using unadjusted figures will lead to underestimation of life expectancy at younger ages (by about 2 years) and an overestimation at older ages (about 0.5 years).

### 7.5.2 Comparing extrapolated mortality rates

Figures 7.2 and 7.3 show age patterns of absolute proportionate differences of mortality rates starting at age 85 for women and men, respectively. These figures reveal two general patterns of absolute proportionate differences. First, there are clear differences in predicted mortality rates between the two methods and these differences increases over age across

[2] This, of course, assumes adjustments of $m(x)$ are the correct ones.




(a) Mortality rates at ages 85-99 (b) Life expectancy at ages 60, 75, 85

Figure 7.1: Estimates of mortality rates and life expectancies when using unadjusted (method 1) and adjusted (method 2) mortality rates at ages 60-84 to make predictions at ages 85-99 for females in Mexico, 1955 and 2005.

all countries for both men and women. In Honduras before 1980, for example, mortality rates at ages 85+ are at least 40% lower if one uses uncorrected death rates below age 85s. Thus, if one extrapolates older adult mortality using uncorrected data (method 1), the resulting quantities will severely underestimate the true values and life expectancy will be overestimated. Second, proportionate differences between estimates narrow in more recent years. This is the result of improvements in national vital statistic systems. In Uruguay, for example, there is virtually no difference in the extrapolated values originating in methods 1 and 2 in 2007.

### 7.5.3 Comparing estimates of life expectancy at ages 60, 75 and 85 between methods

To simplify presentation of results for all countries we group them into three categories that roughly represent distinct regimes of mortality declines: (a) Forerunners include Argentina, Costa Rica, Cuba, and Uruguay. These are countries with an early onset and very gradual mortality decline; (b) Laggards include Bolivia, Dominican Republic, El Salvador, Guatemala, Nicaragua, Peru, Honduras, and Paraguay or countries with a late and very rapid mortality decline. (c) Intermediate countries include Brazil, Chile, Colombia, Mexico, Panama, and Venezuela. These are countries that experienced a regime of mortality decline intermediate between forerunners and laggards. Figure 7.4 shows time trends in proportionate differences between methods 1 and 2 estimates of life expectancy at ages 60, 75 and 80 for women and men. Figure 7.5 summarizes these patterns using averages of absolute proportionate differences within decades and by sex and regime of mortality decline.

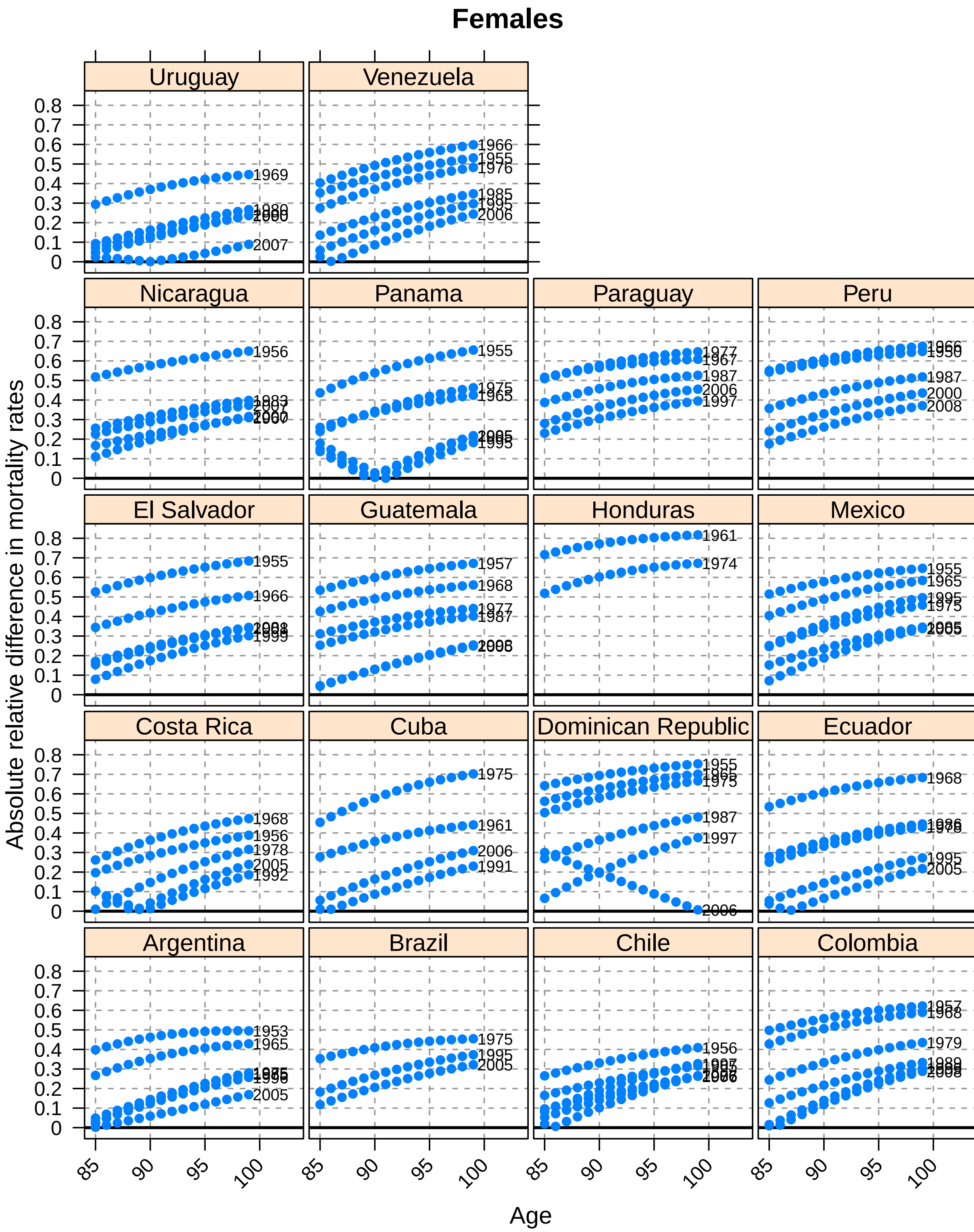



Figure 7.2: Female age patterns of absolute relative differences between method 1 (unadjusted mortality rates) and method 2 2 (adjusted mortality rates) in predicted mortality rates for ages 85-99.

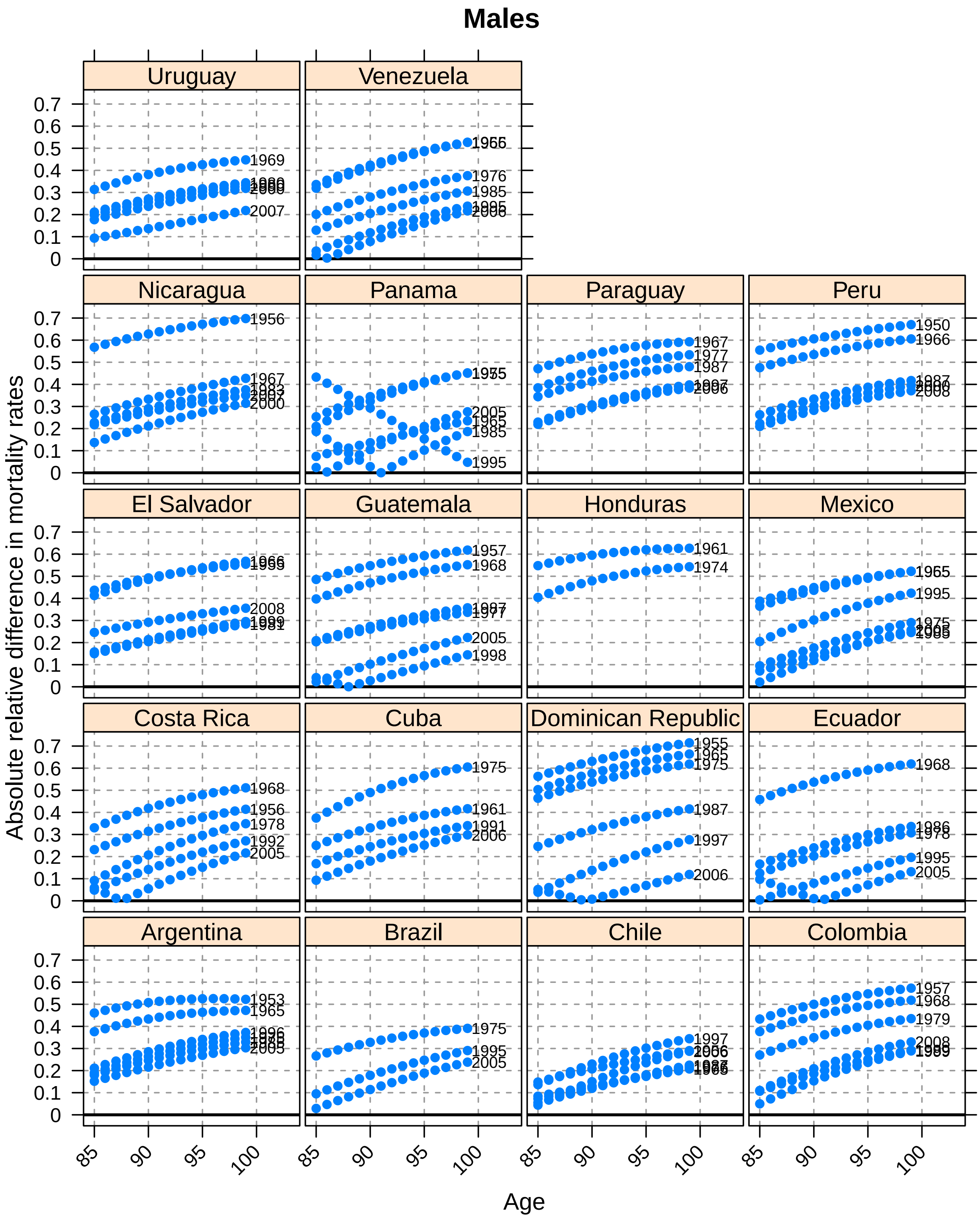



Figure 7.3: Male age patterns of absolute relative differences between method 1 (unadjusted mortality rates) and method 2 (adjusted mortality rates) in predicted mortality rates for ages 85-99.

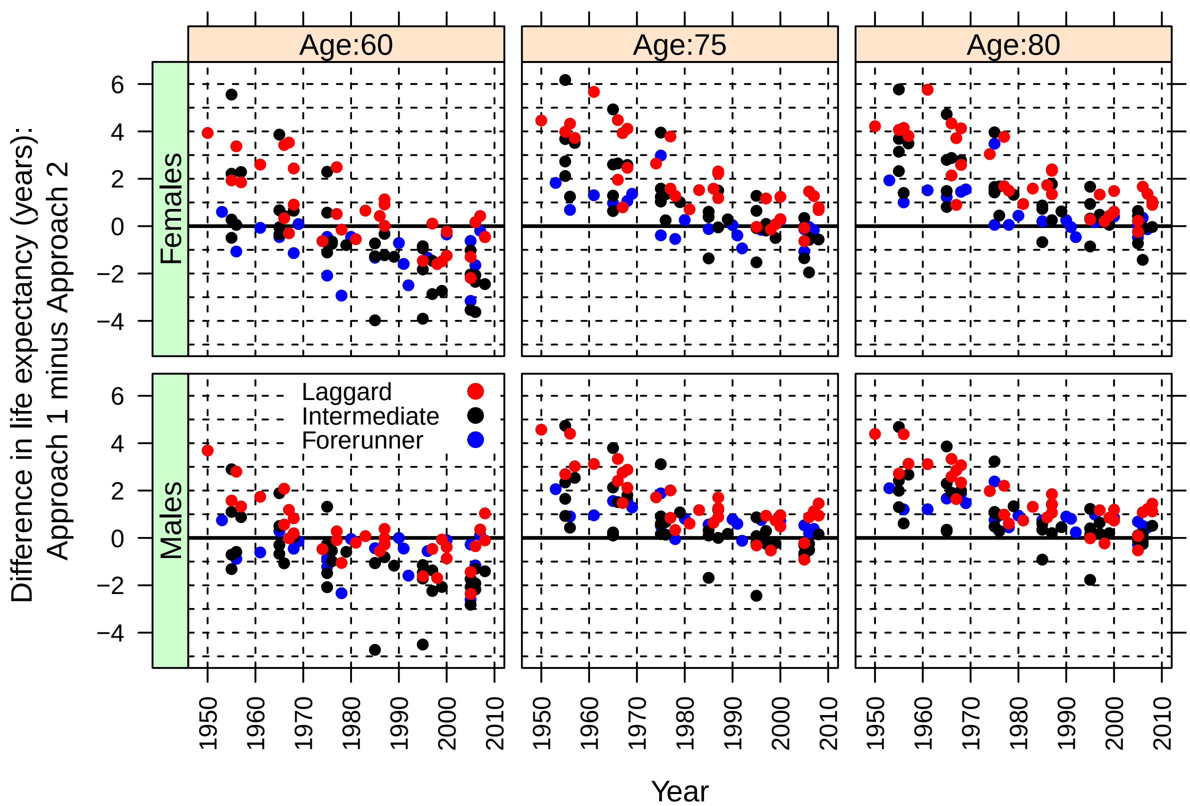



Figure 7.4: Time trends in differences in life expectancy between approaches at ages 60, 70, and 85 by sex in LAMBdA countries, 1950-2010. Methods 1 and 2 are described in section 2. Laggard countries include Bolivia, Dominican Republic, El Salvador, Guatemala, Nicaragua, Peru, Honduras, and Paraguay. Intermediate countries are Brazil, Chile, Colombia, Mexico, Panama, and Venezuela. Forerunners are Argentina, Costa Rica, Cuba, and Uruguay.

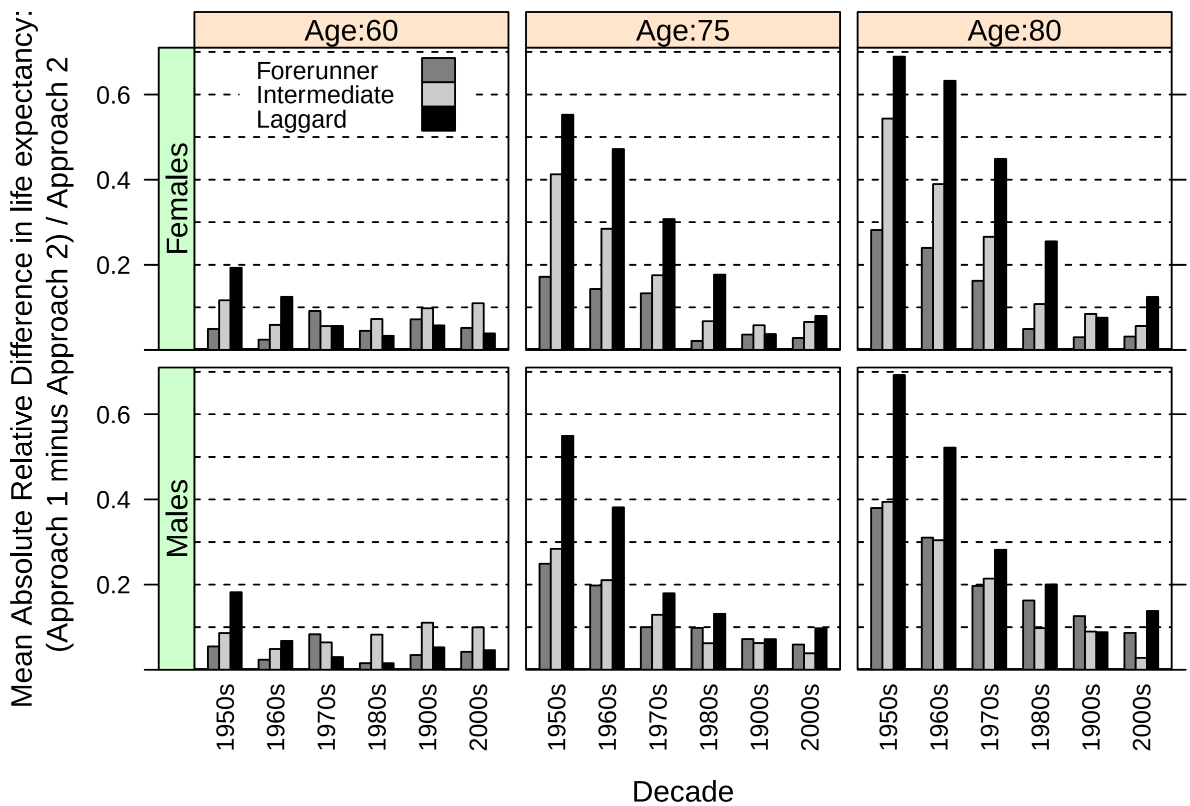



Figure 7.5: Time trends in the mean of absolute relative differences in life expectancy between approaches at ages 60, 70, and 85 by sex and decade for countries in LAMBdA, 1950-2010. Methods 1 and 2 are described in section 2. Laggard countries include Bolivia, Dominican Republic, El Salvador, Guatemala, Nicaragua, Peru, Honduras, and Paraguay. Intermediate countries are Brazil, Chile, Colombia, Mexico, Panama, and Venezuela. Forerunners are Argentina, Costa Rica, Cuba, and Uruguay.

## 7.6 Comparison of life Expectancy at age 85 from fully adjusted rates up to age 85+ and from extrapolated rates with a Logistic model

As described in Chapter 3, LAMBdA life tables are calculated using adjusted (for relative completeness and age misreporting) death rates for all ages, including the open age group. Thus, life expectancy at older ages depend on the precision of the adjustment at very old ages. An alternative approach would have been to follow method 2, fit a logistic model using rates above age 60 , extrapolate mortality rates above age 84 and then calculate the mortality rate n the open age groups that is consistent with the extrapolated rates. As we already mentioned in many places in this documentation, our primary concern was to produce life tables with minimal assumptions about age patterns of mortality. And computing mortality rates at age 85+ using method 2 was not exactly coherent with this strategy. But it is worth asking, what difference would have made? In this section we attempt to answer this question and assess the magnitude and nature of these differences. To do this we focus on $e(85)$ and compare graphically two sets of quantities, the estimated values of $e(85)$ thata result from simple adjustments of observed rates and those that result from application of the second variant (method 2) of Kannisto model. This is done in a series of plots contained in Figures 7.6 to 7.9 that display differences (in years) between estimates in LAMBdA and the estimates from extrapolation. These differences are defined as $(e(85)(LAMBdA) - e(85)(extrapolation))$ .

The following are key results:

1. In all cases the differences are small and nver amounting to more than 10 percent of the values involved.

2. With a handful of exceptions, Kannisto e(85) is always lower than LAMBdA.

3. In most cases the differences between LAMBdA-Kannisto reveal no systematic time patterns but in a few cases we detect increasing (decreasing) trends.

4. There is no clear relation between countries' mortality regime and/or countries' data quality and differences between estimates.. One would assume that vital and census statistics and LAMBdA adjustments in more recent periods get better. If so, differences between LAMBdA and method 2 that are associated with errors would contract, not expand. This is in fact what happens in most countries (differences peak and then decrease rapidly).

In summary, and as could be expected, the direct procedure to compute mortality rates in the open age group does not produce results that are identical to those that require the support of a logistic model. The latter relies on unverifiable assumptions about the age pattern of mortality at ages over 60 whereas the former makes no such assumptions but, in turn, could be influenced by errors derived from adjustments for completeness and age misreporting that are appropriate for younger adult ages but unsatisfactory for ages above 85. The fact that differences between the two are at worst mild, should prop the use of a mortality-pattern free sets of estimates.

Figure 7.6: Time trends of differences in $e(85)$ between LAMBdA and extrapolated (method 2): Argentina-Costa Rica

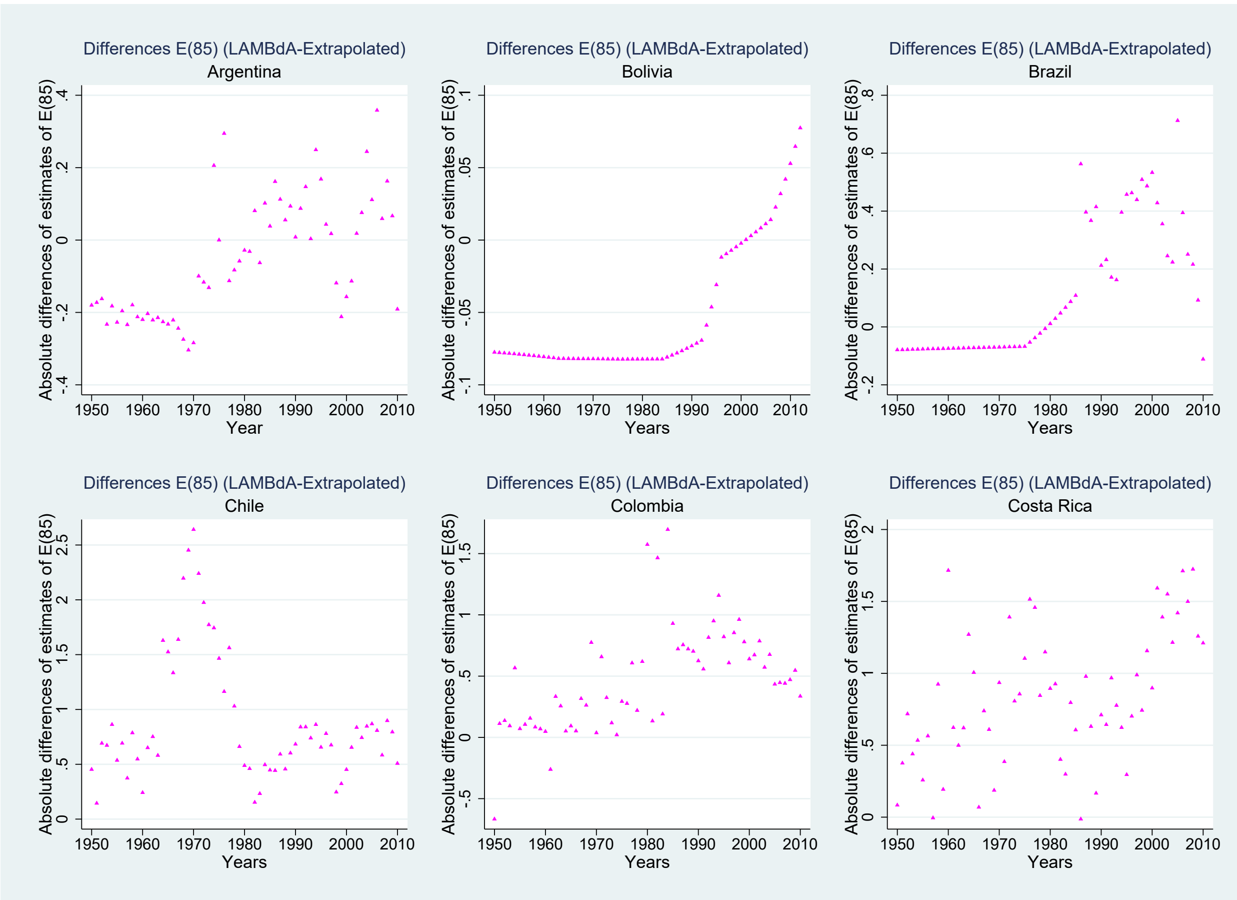


Figure 7.7: Time trends of differences in $e(85)$ between LAMBdA and extrapolated (method 2)(Cont.): Cuba-Honduras

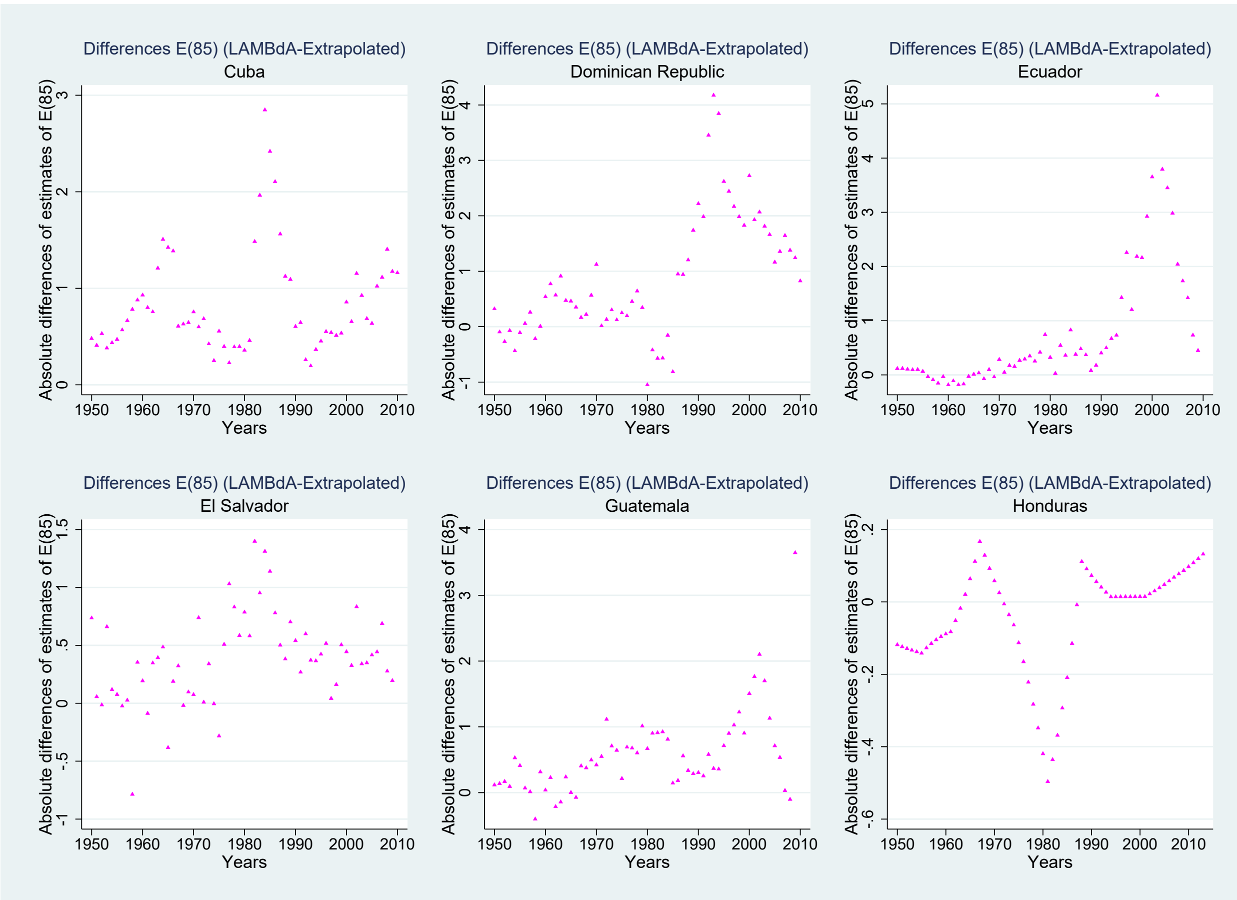


Figure 7.8: Time trends of differences in $e(85)$ between LAMBdA and extrapolated (method 2)(Cont.): Mexico-Uruguay

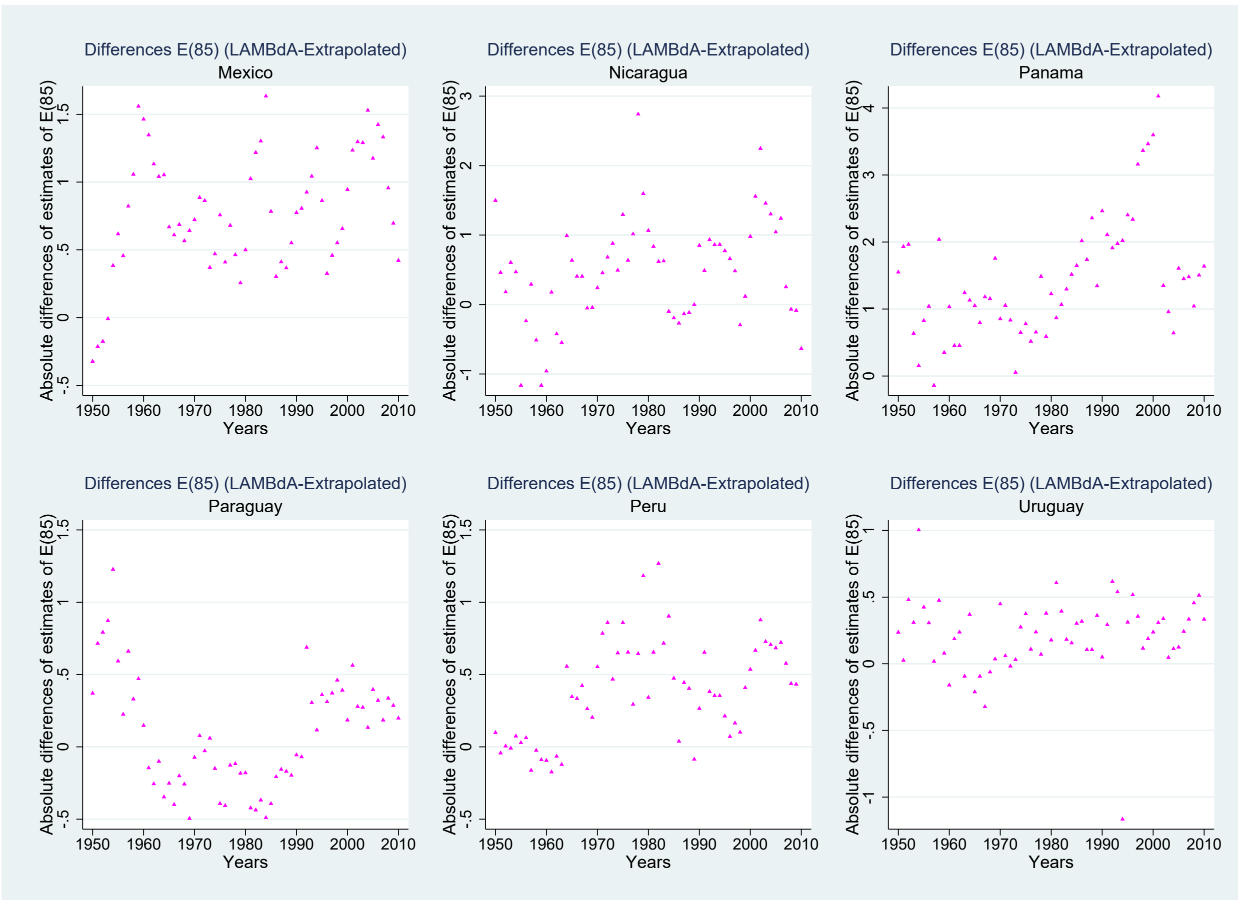
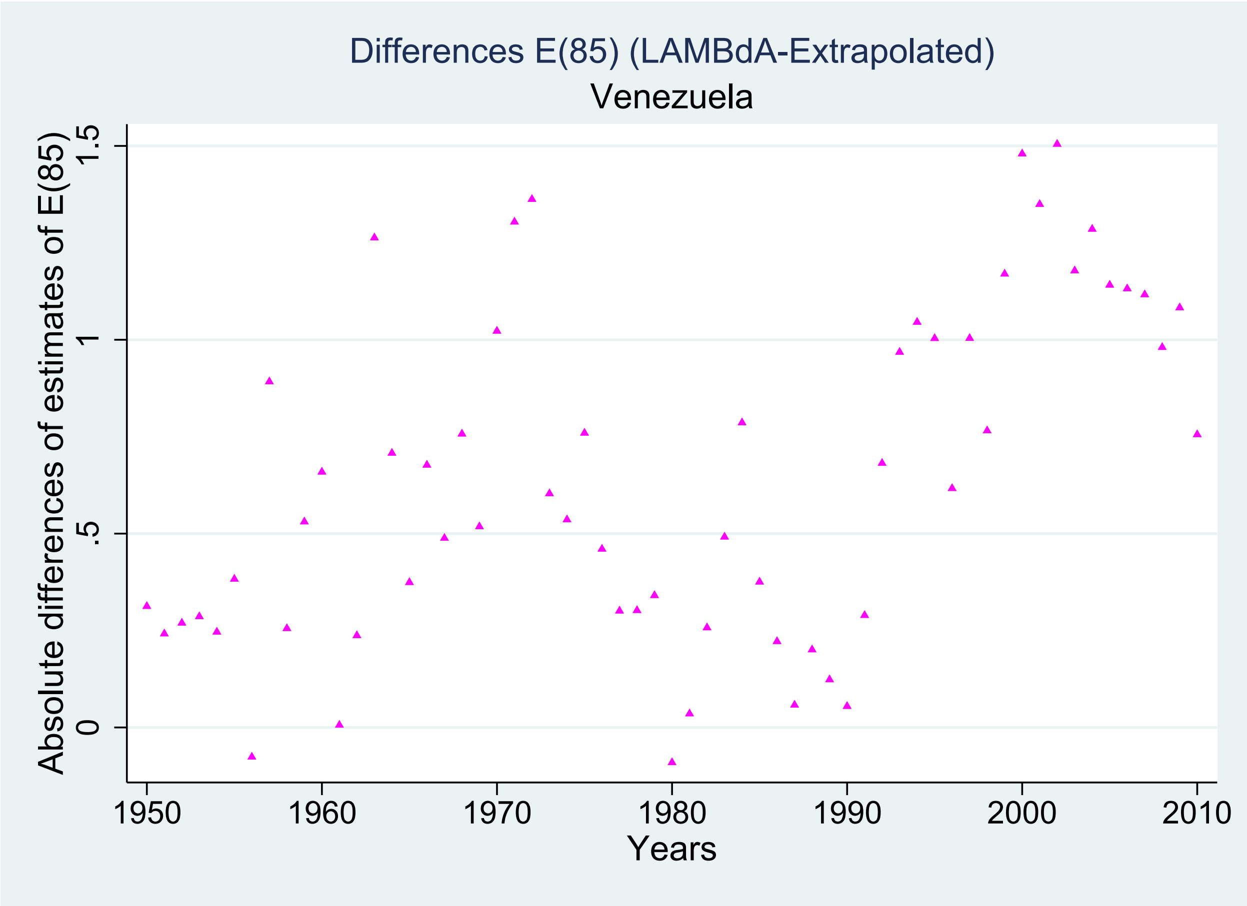


Figure 7.9: Time trends of differences in $e(85)$ between LAMBdA and extrapolated (method 2)(Cont.): Venezuela



# Chapter 8

# Consistency analysis of LAMBdA, HMD and Model Life Tables

## 8.1 Introduction

The Latin American Mortality Database (LAMBdA) was built to document the history of mortality decline in countries of the Latin American and Caribbean region (LAC). LAMBdA was constructed to enable robust empirical testing of conjectures about the nature and determinants of the secular mortality decline in LAC, to highlight unique aspects relative to the mortality decline in countries of North, South and Western Europe and North America, to identify the features of this past experience that are relevant for the future of longevity in the region and, finally, to place the mortality trajectory stretching nearly a century and a half, in a large canvass portraying mortality improvements among humans.

LAMBdA is a large data set consisting of counts of deaths and populations as well as adjusted and undajusted life tables that summarize mortality experiences by year and decade. These statistics are computed for 19 countries and cover the period between 1850 to 2010.[1] Adjusted life tables were computed to correct for errors of coverage of vital statistics and population censuses as well as for systematic age misreporting at adult ages. Aside from the fact that LAMBdA documents a peculiar mortality experience in an entire continent, the data set has two valuable properties. First, adjustments are consistent across countries and time periods, the sources and magnitude of errors are explicitly disclosed, and the methodology for adjustments as well as the codes implementing them are set forth in detail and can be reproduced and altered if the investigator so desires.[2]

Second, LAMBdA contains two modules: one that includes 'optimal' life tables only, e.g. single estimates of a life table that, according to the judgment of the LAMBdA team, best reflects the experience of a particular country-year. In addition, LAMBdA will soon

[1]The raw data are periodically updated as vital statistics and population census counts or estimates become available.

[2]LAMBdA's web site makes available (a) full documentation of adjustment techniques, (b) the detailed sequence of steps that lead from raw statistics to final estimates, and (c) all codes (STATA and R) employed to compute the life tables contained in the data set.

include a second module with alternative estimates of life tables for each country-year and probabilities of "certainty" for each of them. These life tables and associated probabilities can be used by investigators in lieu of a unique ("optimal") estimate in any analysis that requires accounting for uncertainty of measurement.

This chapter reports results from two sets of consistency checks of life table indicators in LAMBdA life tables. The first set of tests consists of contrasting multiple life table statistics in LAMBdA with those in two additional sources of human mortality data, the Human Mortality Database (HMD) and two model patterns (West and South) from the Coale-Demeny model life tables (CDLT) (Coale et al., 1983).

The second set of tests compares adult mortality estimates in some countries included in LAMBdA with alternative ones derived from indirect techniques (orphanhood). These tests aim to show that, at least in countries where the comparison can be carried out, there are no significant discrepancies between direct estimates of adult mortality in LAMBdA (from adjusted vital statistics) and indirect estimates. The differences we observe are almost surely accounted for the indirect estimates dependency on assumed adult patterns of mortality and/or by well-known biases of estimates derived from orphanhood information.

## 8.2 First set of consistency tests: LAMBdA, HMD and CDLT models

We begin describing results of comparisons using a number of life table statistics computed from life tables contained in LAMBdA, HMD and two Coale-Demeny model patterns, West and South (CDLT) These two model life tables correspond to mortality patterns that seemingly represent better the observed LAC mortality experiences contained in the adjusted life tables. In all cases we use life tables in five year age groups (except at the outset where we distinguish age 0 and the age interval 1-4). All life tables are closed at age 85+. Although LAMBdA life tables are adjusted for relative completeness of death registration and for systematic age misreporting at all ages above 45, none of the adjustments relies on narrow assumptions regarding prevailing age patterns of mortality.[3]

The aim of the tests described in this section is not so much to verify or disproof uniformly high levels of concordance across all four life table systems. After all LAMBdA represent a distinct experience and it should surprise no one that the life tables contained in it are indeed different. Rather, we seek to unveil singularities that could signal anomalies in the data instead of reflecting unique but verifiable conditions that generate a particular set of life tables.

Two caveats. First, the CDLT life tables are synthetic summaries of mortality experiences in a number of countries, all of which are included in HMD. Thus, one should not

[3]Although the actual construction of any single LAMBdA life tables does not directly invoke assumptions about model patterns, we relied on model patterns to produce candidate estimates of infant and child mortality. As described in Chapter 3 we chose to use the West and South models from the Coale-Demeny families and the Latin American model from the United Nations families. In this chapter we only discuss results using the West and South to avoid redundancies and cluttering since those obtained using the Latin American model fall in between the other two set of results.

expect that HMD behaves as any single model within the CDLT anymore than LAMBdA life tables will behave as those in HMD or the CDLT. HMD and LAMBdA reflect observed experiences whereas CDLT are synthetic patterns extracted mostly from observed Western mortality trajectories up until 1970.

Second, the representation of life tables in HMD and LAMBdA, but particularly in the latter, during any period of time or for any given mortality level, reflects the experience of different countries. Thus, for example, LAMBdA life tables for the period before 1950 with life expectancy exceeding 60 years reflect the experience of a small number of countries, a result of the fact that the generalized onset of mortality decline in the region was delayed until after 1940. Thus, contrasts of life table statistics for the period before 1950 will be unevenly influenced by a handful of experience and could be the result of country specific idiosyncracies. This is one more reason to interpret our tests as just a device to identify singularities that reveal faulty measurement, poor adjustments, or other artifacts immanent in the procedures we use to construct the tables.

To implement consistency checks we stratify LAMBdA life tables by periods, before 1950 and after 1950. This is done with the aim of distinguishing estimates' behaviors during a period when the quality of vital statistics and censuses is increasing (post 1950) and a period when the quantity and quality of data is lower and mortality statistics must be carefully adjusted before using them (before 1950). It turns out that this stratification also reflects with high fidelity a shift in historical patterns as the two time intervals broadly represent periods before and after major mortality changes took place. In particular, the post 1950 period is one of unprecedented gains in life expectancy, possibly matched only by the post-1950 Japanese experience and post-1970 mortality in China and a handful of North African countries. In addition to stratifying by period, we also compute statistics by levels of mortality indexed either by life expectancy at birth or by the probabilities of dying before age 5. This ensures that all comparisons of life table indicators in LAMBdA, HMD and CDLT, remove differences that are best attributed to mortality levels and not age-pattern disparities.

We use four sets of consistency checks. First, we compute summary measures of life tables statistics at different ages and compare them to those observed in HMD and CDLT (Figures 8.1–8.6). Second, we estimate relations between child and young adult mortality, on one hand, and older adult mortality, on the other, and compare those in LAMBdA with those found in HMD and CDLT (Figures 8.7–8.9). Third, we estimate differences between the values of mortality indicators observed in LAMBdA and those expected if LAMBdA life tables followed the CDLT patterns (Figures 8.10–8.15). Finally, we assess patterns of gender differences and again compare them to those in HMD and CDLT (Figures 8.17–8.20).

The main take away message from these assessments is that the statistics whose behavior we chose to study exhibit similar patterns to and behave much like those observed in HMD or those contained in CDLT. We find that when departures from the CDLT are detected they resemble those found in HMD. The checks reveal no aberrant behavior that could suggest the presence of systematic biases or distortions.

## 8.3 Years of life lived in closed and opened age intervals

In this section we use box-plots to examine patterns of average number of years of life lived in different age groups among women (top panels of figures) and men (bottom panels of figures) (Figures 8.1–8.6). The values plotted in the figures are, respectively, the average number of years of life lived in the following age groups: between 0-59, 65-79, and 80 and above. In addition we focus on the difference between average years lived in the age group 0-64 and average years lived after age 80. In all cases we plot the statistics by mortality levels (indexed by life expectancy at birth) to highlight changes that take place as overall mortality levels improve (represented in the figures along the x-axis). For comparison we include the same indicators from HMD and CDLT models West and South (displayed along different columns in all figures).

### 8.3.1 Average years of life lived in the interval 0-64 ($_{65}e_0$)

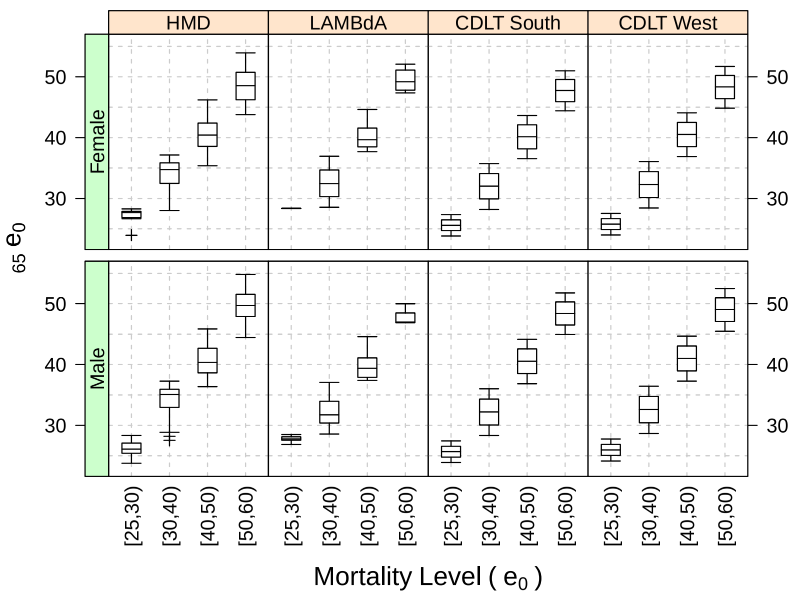
Human life tables with life expectancy at birth below 65 years or so, should display similar values of mean years lived in the interval 0-64 since both statistics are strongly dominated by mortality experienced early in life. Thus, an increase in life expectancy at birth in populations experiencing a life expectancy between 30 and 65 years should be reflected in corresponding increases in the mean years lived in the interval 0-64 and less so in the mean number of years lived above that age. The results we obtain are consistent with this expectation. Figure 8.1 displays boxplots of the distribution of $_{65}e_0$ at different mortality levels in LAMBdA, HMD and models West and South in the CDLT among males and females. The left panel shows values in LAMBdA life tables available for the period before 1950 whereas the right panel includes the values for the period 1950-2010. Although the figures show strong similarities in the distributions of $_{65}e_0$ at different mortality levels, there are some differences across data sources. We describe these below.

First, among females the median values of $_{65}e_0$ before 1950 are slightly lower in LAMBdA than in HMD, at least when $e_0$ is within the range [30,40) or [50,60). When $e_0$ is within the range [40,50)(see tables in section 8.7.1) LAMBdA life tables show slightly higher values. For example, when life expectancy at birth is in the range [30-40) before 1950 the median value of $_{65}e_0$ in LAMBdA is about 32.4 years, quite similar to that in model life table West (32.3) and South (32) but about 2.3 years lower than in HMD (34.7). When $e_0$ is within the range [40,50), $_{65}e_0$ values of the statistics are approximately .5 year lower in LAMBdA than in HMD but quite close to the values in the CDLT system. In the uppermost bracket of life expectancy at birth the observed values in LAMBdA are again smaller than those in HMD by about 1.5 - 2 years. For the period after 1950, LAMBdA shows higher median values in $_{65}e_0$ than the other sets of life tables, particularly so at lower levels of life expectancy. These differences, however, are all quite modest and never exceed 2 years.

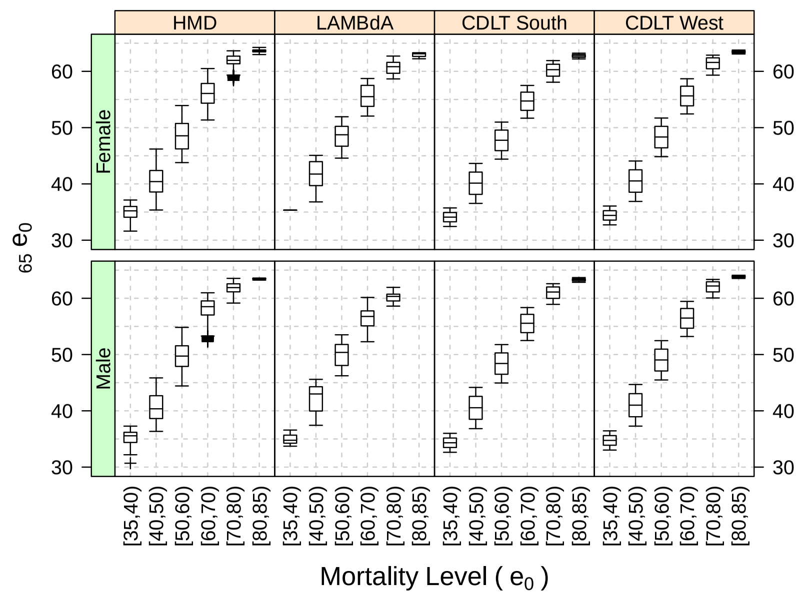
The interquartile range (IQR) of $_{65}e_0$, suggests consistently similar distributions of values of the target statistics across levels of life expectancy and life table sets. LAMBdA's $_{65}e_0$ IQR for females is of the order of 10-12 years in the period 1900-1950, consistent with life tables at similar levels of mortality in HMD and CDLT. For the period after 1950, the female IQR for ${}_{65}e_0$ in LAMBdA life tables shows a slightly more compressed distribution when $e_0 \in [40,50)$ but a more disperse distribution when $e_0 \geq 50$.

Second, results for males are consistent with those of females. LAMBdA includes life tables with lower median values in ${}_{65}e_0$ for all mortality levels before 1950, somewhat higher mean values when $e_0 \leq 60$ after 1950 (see tables in section 8.7.1), and few differences in the IQR ranges in each period across all levels of life expectancy in the four life table systems.

The lower levels of mean years of life lived in 0-64 in LAMBdA life tables before 1950 possibly reflect higher child mortality relative to adult mortality in the LAC countries during that period. These values are then reduced during the massive mortality decline experienced after 1950 and concordance between life tables in LAMBdA and the others strengthens. Because the dispersion of values around the median for each mortality level is quite narrow across life table systems, the life tables can be hardly distinguished.




(a) LAMBdA 1900-1950




(b) LAMBdA 1950-2010

Figure 8.1: Box-plots of the average years of life lived between age 0 and 65 ($_{65}e_0$) vs. mortality level, indexed by life expectancy at birth ($e_0$).

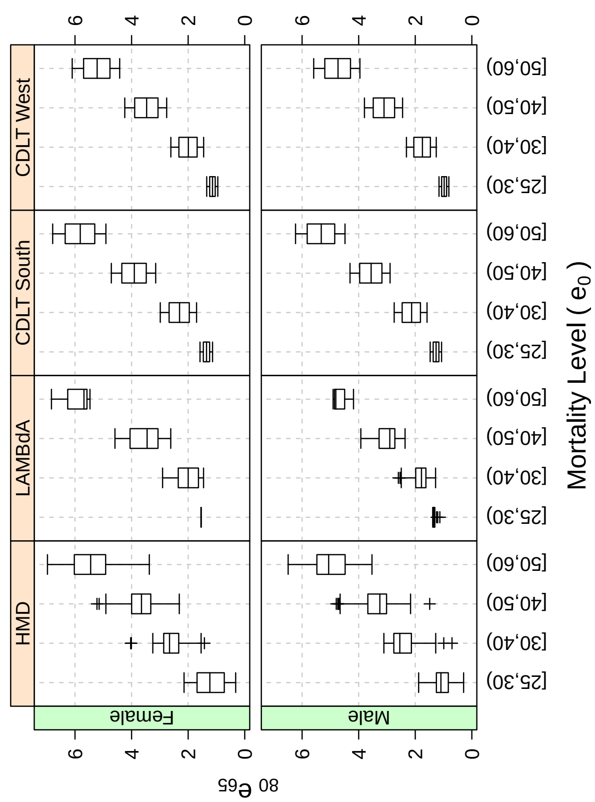
### 8.3.2 Average years of life lived in the interval 65-84 (${}_{80}e_{65}$)

We chose the age interval 65-84 to reflect older age mortality excluding ages older than 85+ (these are examined below). Figure 8.2 displays indicators of the distributions of ${}_{80}e_{65}$ at different mortality levels for LAMBdA, HMD and models West and South in the CDLT for males and females. As was the case before, there is strong overall concordance between LAMBdA and the other life table systems

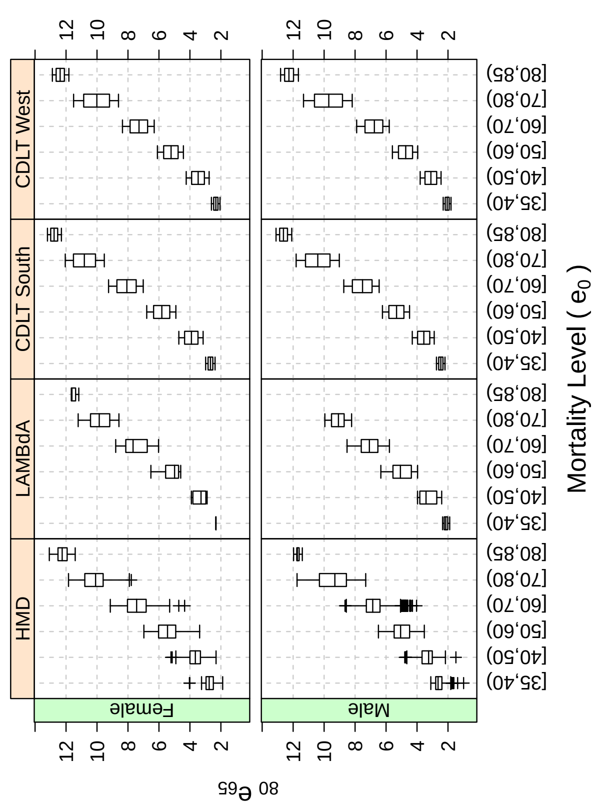
For the period before 1950, and among both males and females, median values in ${}_{80}e_{65}$ are consistently lower in LAMBdA than in HMD and model South (but not model West) across all mortality levels (see tables in section 8.7.2). Interestingly, male and female life tables in model South exhibit higher values of the target statistics than life tables in LAMBdA and differences can be as large as 2 years. This occurs irrespective of level of mortality or time period. It should be noted that model South in the CDLT reflects age patterns of mortality dominated by life tables from Southern European countries between 1915 and 1955 approximately. The pattern is characterized by relatively higher mortality before age 5, lower mortality between ages 30 and 64 and higher mortality at older ages. Thus, the old age pattern embedded in LAMBdA throughout the entire period appears to be less beneficial at older ages than the one embedded in model South. The magnitude of discrepancies between HMD and Models South are in the same direction as those between LAMBdA and model South but of much reduced magnitude.

Although differences identified above never exceed 2 years, they represent relative differences of the order of 8 to 12 percent. For the period after 1950, LAMBdA median ${}_{80}e_{65}$ values continue to be lower relative to the other life table sets but differences are now much smaller and, since the values of the target statistics increase, the proportionate differences are much smaller than those observed during the period before 1950. By and large the IQR ranges are more compressed in LAMBdA before 1950 but reflect similar spread of values for the more recent period. This is due to the composition of countries and higher inter-country heterogeneity during the period before 1950.

A final remark is in order. As life expectancy at birth increases, the strength of the contribution of years of life lived at older ages rises and one should observe higher rates of increase in the mean years lived in 65-84 at higher levels of life expectancy. This expected convexity of the relation between the target statistics and levels of life expectancy is quite plain in all four life table sets.




(a) LAMBdA 1900-1950

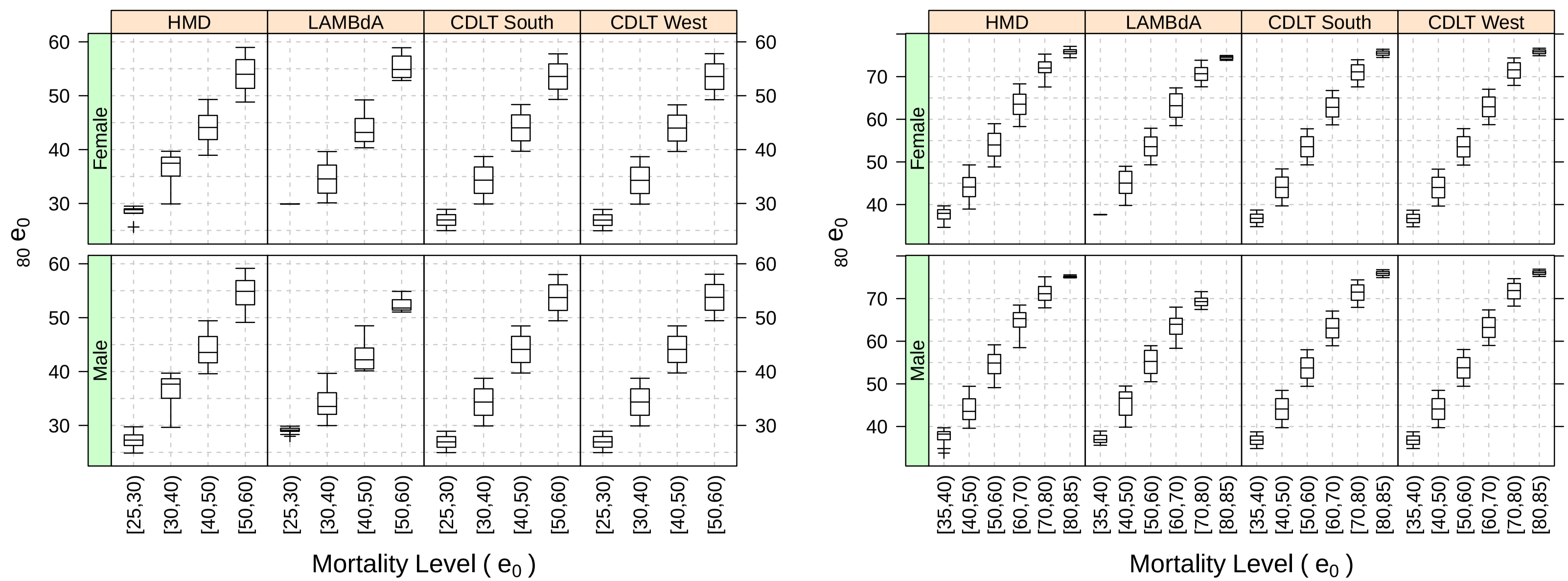



(b) LAMBdA 1950-2010

Figure 8.2: Box-plots of the average years of life lived between ages 65 and 80 ($_{80}e_{65}$) vs. mortality level, indexed by life expectancy at birth ($e_0$).

### 8.3.3 Average years of life lived in the interval 0-79 ($_{80}e_0$)

The statistic we study in this section, $_{80}e_0$, is the average years of lived from birth up to age 79 and should be very similar to the life expectancy at birth, at least when its value is lower than 65 or 70 years. Thus, the plot of one against the other should result in a straight line going through the origin with an increase in the slope at very high levels of life expectancy. Figure 8.3 shows that this is the case. The figure also shows that the medians of the statistic for women in LAMBdA during the period before 1950 are consistently lower than those in the HMD across all mortality levels but less so than those in the model life tables (see tables in section 8.7.3). Although these differences contract during the post-1950 period, they persist during the twentieth first century , specially at higher values of life expectancy at birth. Because at these high levels of life expectancy child mortality must be very similar across life tables, differences must be attributable to lower mortality rates in HMD at young adult ages (20-45) or, more likely, in the age group 45-79. This is consistent with conjectures that suggest that old age mortality levels in LAC are relatively higher than in North America or Western Europe. Despite their persistence, the differences are small. Thus, results for males point to differences of less than 1.5 years before 1950 and less than 0.4 years after 1950. The IQR ranges, however, reveal a wider distribution of $_{80}e_0$ in LAMBdA when $e_0 \in [30, 40)$ and a more compressed one when $e_0 > 40$. Despite these differences, HMD and LAMBdA exhibit strongly concordant features and there is no clear evidence of singularities or aberrant deviations from known patterns of mortality.




(a) LAMBdA 1900-1950

(b) LAMBdA 1950-2010

Figure 8.3: Box-plots for the average years of life lived between 0 and 80 ($_{80}e_0$) vs. mortality level, indexed by life expectancy at birth ($e_0$).

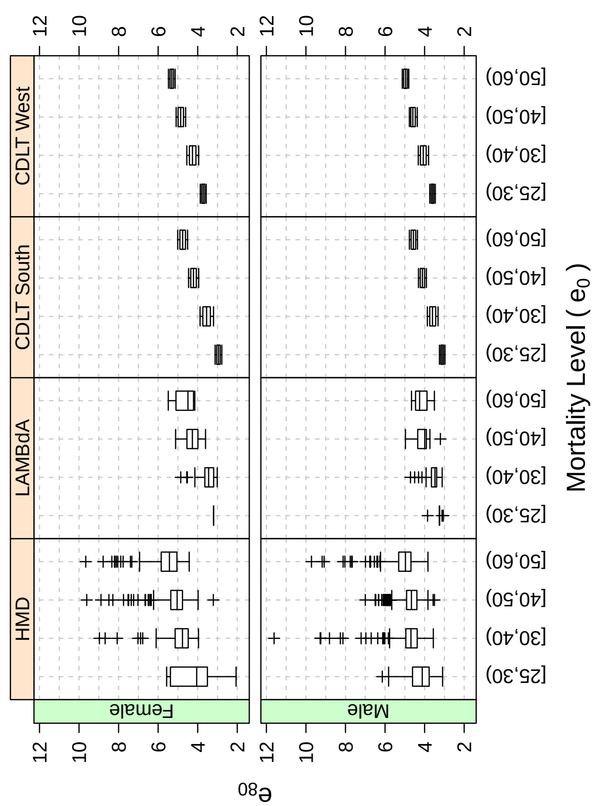
### 8.3.4 Life Expectancy at age 80 ($e_{80}$)

Life expectancy at age 80 is an indicator of the intensity of mortality risks among the oldest old. It is an important statistic in countries with high levels of life expectancy at birth as the fraction of their population attaining their 80th birthday has grown rapidly in the last two decades or so. The statistics may be influenced by the strength of mortality at younger ages as this shapes the composition by sturdiness of individuals who survive to age 80.

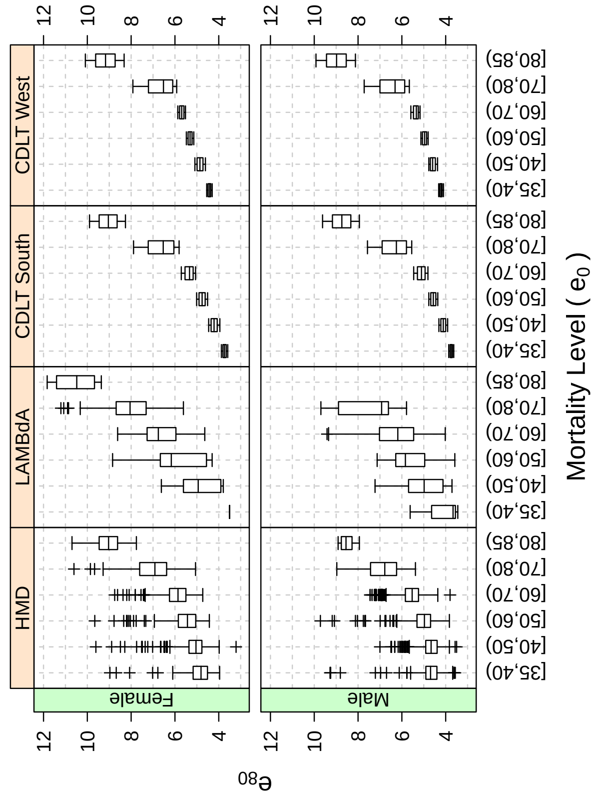
Figure 8.4 displays boxplots of the statistic. Not surprisingly, at lower levels of life expectancy at birth the values of $e_{80}$ are quite small and in the models of mortality the variability by levels of life expectancy is minimal. Not so in HMD where the influence of outliers is very strong, and in LAMBdA as well where the impact of outliers' is smaller but still quite visible. Measurement of mortality at these very old ages is more fragile, particularly at lower levels of life expectancy, and the influence of systematic errors is stronger. Thus, the observed heterogeneity may be a result of measurement problems. Although less likely, it could also reveal divergent patterns of mortality at very old ages caused by different frailty composition or different patterns of senescence.

LAMBdA and HMD life tables during the post-1950 period represent a broader range of mortality experiences and includes populations with higher levels of life expectancy at birth. This is manifested in massive heterogeneity in the values of $e_{80}$. The distributions of the statistic in LAMBdA are more compressed and less marked by outliers and, at least among females, have slightly higher medians. Thus, for example, females with the highest life expectancy at birth have a median value of $e_{80}$ of about 1.3 years higher than the median in HMD. Differences are lower when life expectancy decreases and are much less salient among males. An important caveat is that the distribution of the statistic for high levels of life expectancy at birth in LAMBdA is computed with a sample of just a handful of country-years whereas those from HMD are based on a much larger sample that reflects more diverse mortality patterns.

In summary, the empirical behavior of $e_{80}$ is less patterned in HMD and LAMBdA than in models of mortality and may reflect heavier influences of heterogeneity in mortality experiences and possible idyosincracies of measurement. With regards to heterogeneity of values, LAMBdA and HMD are not much different from each other as the distribution of the statistic is quite dispersed in both sets of life tables. Contrasts do exist but they implicate the statistic's median level. Thus, LAMBdA median values $e_{80}$ for females with high life expectancy are 1.0 to 1.5 years higher than the medians for similar levels of life expectancy in HMD. These differences characterize females only and they contract at lower levels of life expectancy. Among males the differences are reversed and LAMBdA's medians of life expectancy at age 80 are lower by as much a 1 year than those for similar levels of life expectancy in HMD. It is quite unlikely that systematic errors of misreporting could be inducing contrasts across gender.




(a) LAMBdA 1900-1950



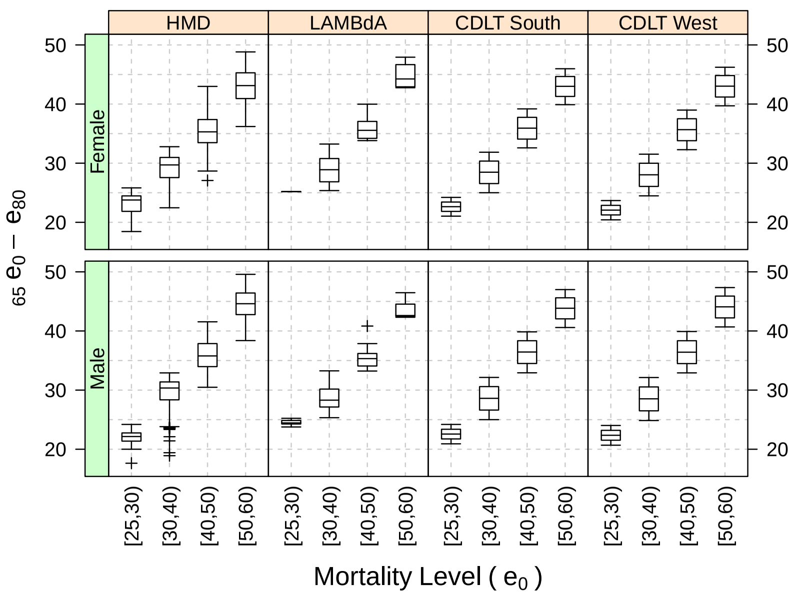

(b) LAMBdA 1950-2010

Figure 8.4: Box-plots of life expectancy at age 80 ($e_{80}$) vs. mortality level, indexed by life expectancy at birth ($e_0$).

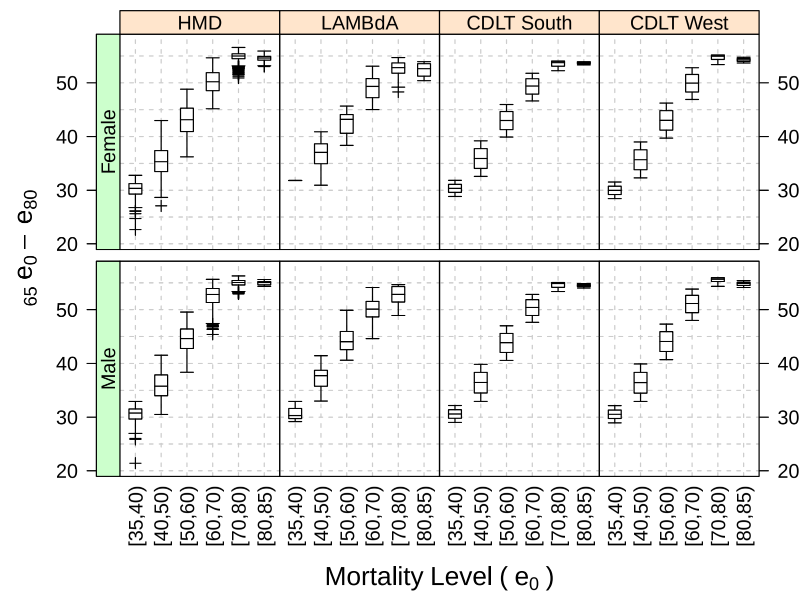
### 8.3.5 Differences between average years of life lived in the interval 0-64 and life expectancy at age 80 (${}_{65}e_0 - e_{80}$)

The numerical value of this statistic will tend to increase as life expectancy attains higher levels. This reflects the impact of mortality improvements at very young ages that tend to raise ${}_{65}e_0$ much more than $e_{80}$. The growth of the statistic, however, should stall as populations reach very high levels of life expectancy and improvements in mortality at older ages begin to surpass improvements at very young ages.

The patterns in figure 8.5 are consistent with this expectation. During the pre-1950 period, when life expectancy at birth was at lower levels, the median values in LAMBdA and HMD are quite similar and differences do not follow a consistent pattern. The same holds for the distribution of the statistic ${}_{65}e_0 - e_{80}$. At higher levels of life expectancy, during the post 1950 period, differences contract and virtually disappear. Both HMD and LAMBdA follow very closely patterns observed in models South and West of the CDLT.




(a) LAMBdA 1900-1950

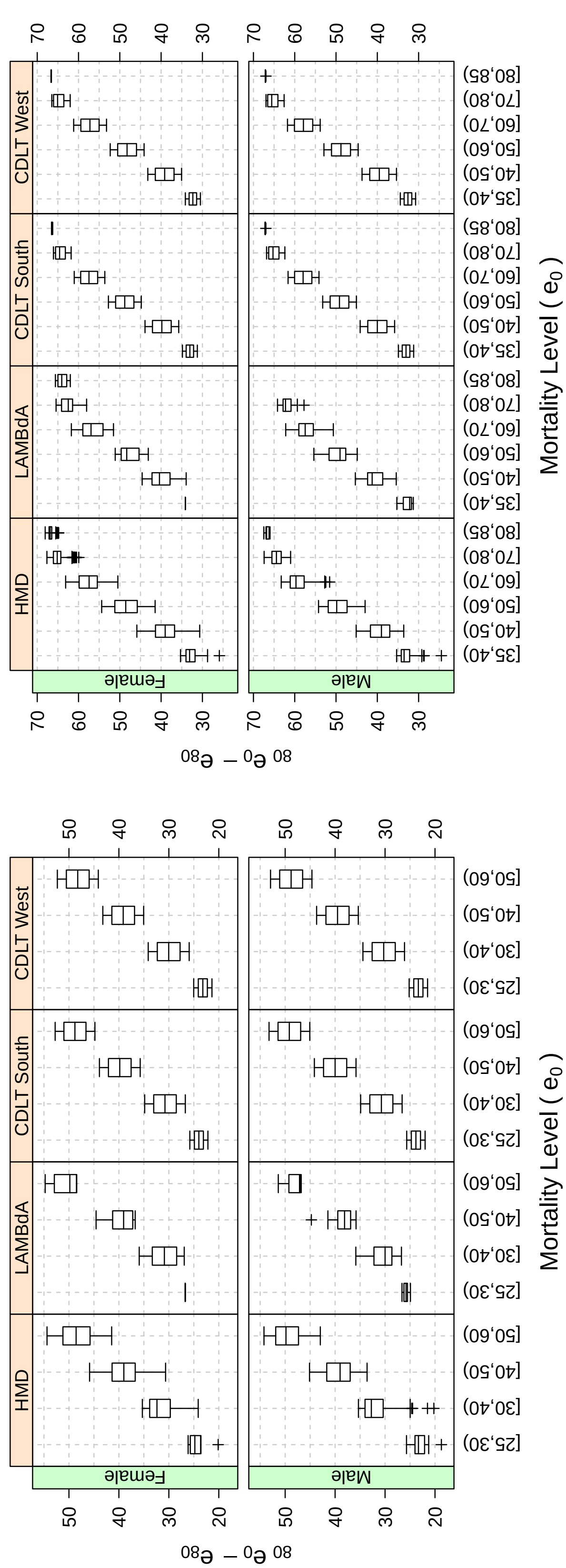



(b) LAMBdA 1950-2010

Figure 8.5: Box-plots showing average years of life lived between ages 0 and 65 minus life expectancy at age 80 ($_{65}e_0 - e_{80}$) vs. mortality level, indexed by life expectancy at birth ($e_0$).

### 8.3.6 Average years of life lived in the interval 0-79 minus life expectancy at age 80 (${}_{80}e_0 - e_{80}$)

The statistic we examine in this section is similar to the previous one but uses an older age as a boundary. Thus, trends of ${}_{80}e_0 - e_{80}$ by levels of life expectancy reflect the relative size of gains in years of life lived in the age intervals 0-79 and 80+. One should expect the same pattern observed for the statistic ${}_{80}e_0 - e_{80}$ except that the slope of the curve must taper off at higher levels of life expectancy, that is, when improvements in mortality at ages below 80 approach a ceiling or threshold. The box-plots in figure 8.6 do bear the expected pattern in all cases. Further, as the previous figure did, the plots reveal a high degree of concordance across life table systems as the difference we detect hardly exceed a few percentage points. There are no salient contrasts by gender nor across time periods. If anything, it is the dispersion of values that manifests some heterogeneity, at least when comparing HMD and LAMBdA against the two model patterns.




(a) LAMBdA 1900-1950

(b) LAMBdA 1950-2010

Figure 8.6: Box-plots showing average years of life lived between ages 0 and 80 minus life expectancy at age 80 ($_{80}e_0 - e_{80}$) vs. mortality level, indexed by life expectancy at birth ($e_0$).

## 8.4 The relation between infant, early childhood and adult mortality

In this section we examine the relation between indicators of mortality at key stages in the life span. The most important of these are the relations between infant and early child mortality, child and adult mortality, and child older age mortality. The original CDLT mortality patterns emerged from contrasts that revealed singularities in relations between mortality rates in these age groups. Although the original analyses to detect patterns included all five year age group, in this assessment we only consider four age groups that, however, summarize well the mortality experience embedded in any life table.

### 8.4.1 Infant and early childhood mortality

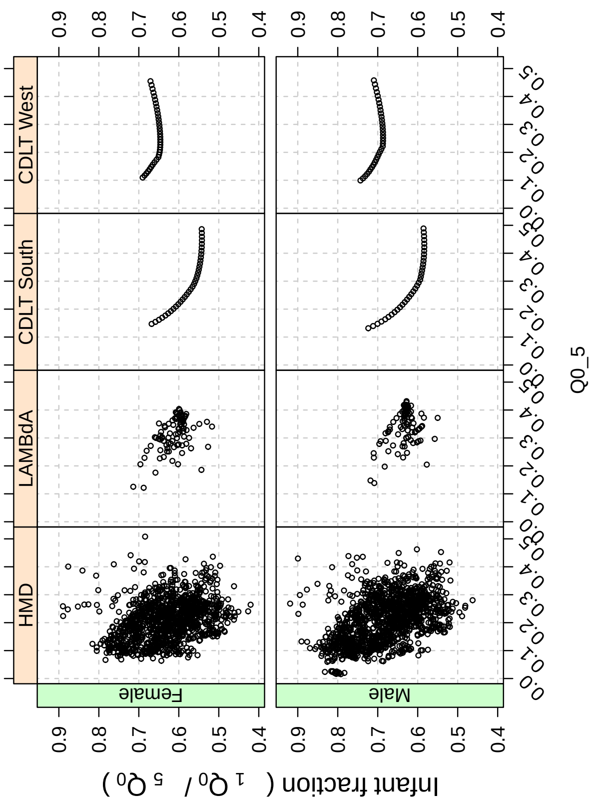
We will use the ratio of infant mortality to the probability of dying before age 5 or ${}_1Q_0/{}_5Q_0$ and ${}_5Q_0$. This ratio reflects the relative contribution to mortality below age five of infant mortality.[4] One of the most important distinctions in the CDLT models is rooted in the relation between mortality in these two age groups. Thus, for example, the North model is unique in that it is characterized by low infant mortality relative to mortality in the age group 1-4 thus producing low values of the statistic we examine. The ratios in Model South are similar to, though slightly higher, than those in Model North but result from simultaneously elevated infant mortality and mortality rates in the age group 1-4. Instead, model West has higher ratios at all levels of mortality implying that any mortality level below age 5 is attained from higher contributions of infant mortality than from contributions of mortality in the age group 1-4. It goes without saying that these diverging patterns reflect underlying epidemiological regimes characterized by the dominance of different causes of death.

Figure 8.7 displays the relation between the target statistic and the value of ${}_5Q_0$, which we use as an index for the level of mortality. First, note the profile of the relations in the two models from CDLT: the curves are similarly concave but the West model curve is shifted upward relative to the South model curve. Second, both HMD and LAMBdA contain patterns that are quite similar to those of the CDLT models but concordance is sharper at lower levels of mortality (compare plots before and after 1950 with those of the models). The heterogeneity in HMD at higher mortality levels is massive and one can hardly identify a clear relation. This occurs because HMD contains representation of life tables belonging to four mortality models in the CDLT, all of which have different patterns of relations between infant and early childhood mortality. Instead the relation in LAMBdA is compact and less heterogeneous, specially so at lower levels of mortality. Further, the pattern of relations in LAMBdA resemble more closely relations in model South than those in the model West. Third, both HMD and LAMBdA contain a few outliers though most of these are located in life tables with high mortality (compare panels for periods before and after 1950). The final inference from these comparisons is that neither in LAMBdA nor in HMD do life tables reveal gender contrasts of note as all three results summarized above apply equally well to

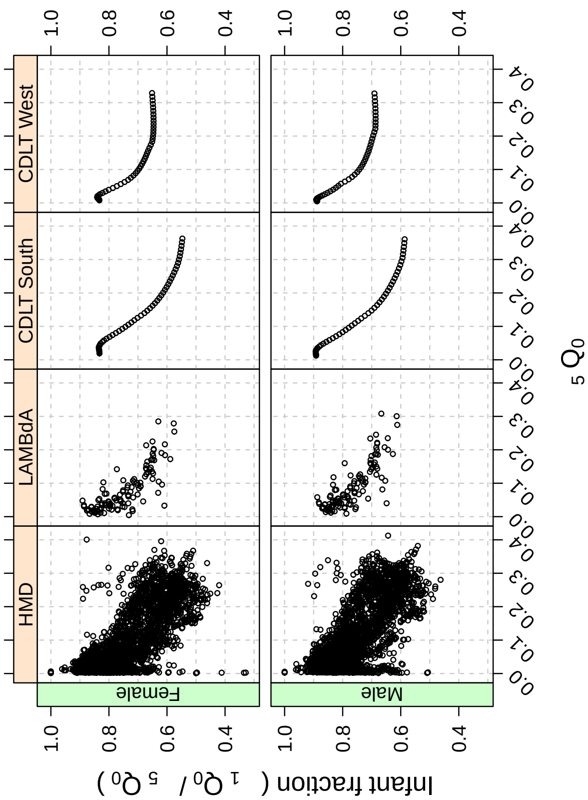
[4] When the integrated force of mortality below age 5 is small this ratio is approximately equal to the ratio of infant mortality to the integrated force of mortality below age 5.

males and females.

In summary, other that the presence of a few outlying observations in high mortality life tables, the behavior of the statistic in LAMBdA bears strong similarities to those in HMD as well as to those embedded in the CDLT South model.




(a) LAMBdA 1900-1950




(b) LAMBdA 1950-2010

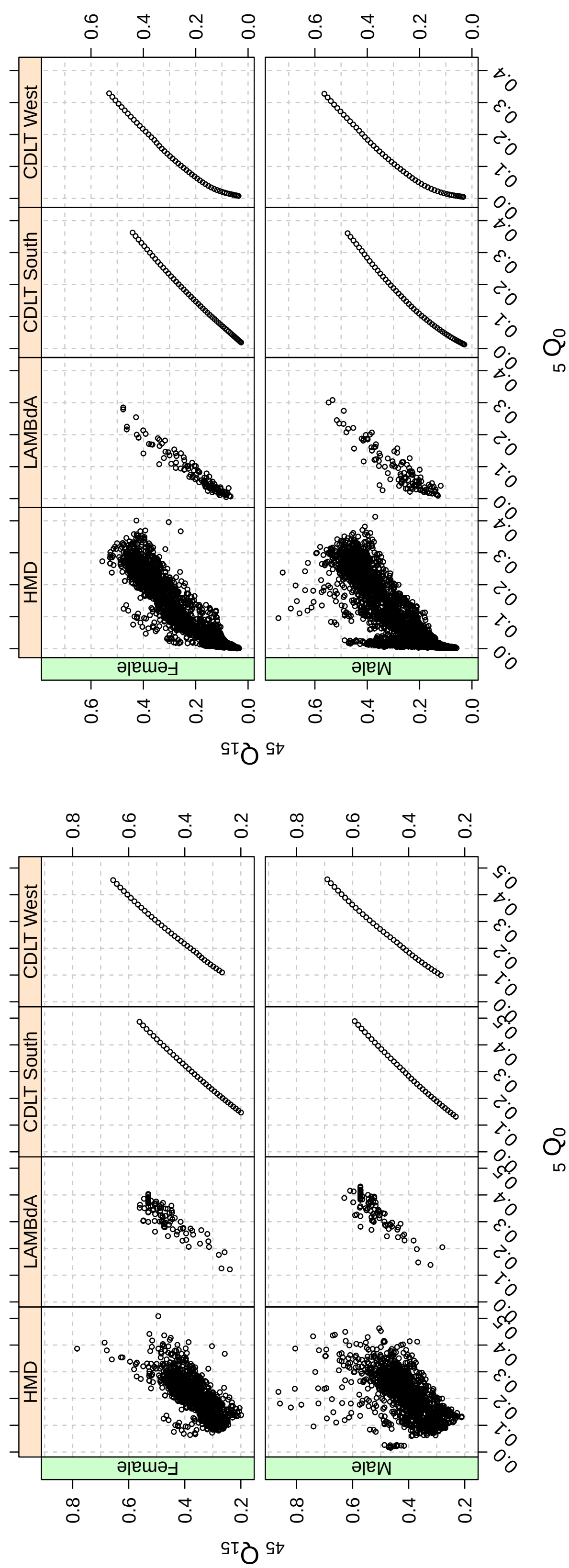
Figure 8.7: Relation between the infant fraction ($_1Q_0/_5Q_0$) and child mortality ($_5Q_0$).

### 8.4.2 Adult and child mortality

In this section we focus on the relation between mortality in the age group 15-44 (${}_{45}Q_{15}$) and below age 5 (${}_{5}Q_{0}$). The CDLT models contain sharp distinctions in the relation between mortality among children and young adults. Perhaps the best known is the pattern contained in models North, South and East. In models North and East mortality in childhood, particularly in the age interval 1-4 in Model North and infant mortality in Model East, is high relative to mortality in adulthood. The contrast is even stronger in Model South in part because in this model mortality both in infancy and early childhood are quite elevated.

The relation between these two indicator is driven by a number of factors. First, the force of selection is stronger when child mortality is high and this may alter the composition by frailty of those surviving to adulthood. Jointly with, and in addition to, selective pressures we should consider the nature of mortality regimes heavily influenced by early childhood infectious diseases characterized by high lethality that also confer later life immunity. Second, epidemiological regimes heavily affected by violence and accidents will be characterized by higher levels of mortality in young adult ages relative to early childhood. Finally, the pattern of relation could vary by gender as mortality among young adult females is highly dependent on maternal mortality in high fertility regimes.

Figure 8.8 shows the relation between the two mortality indicators. The first feature in these plots is that the slopes of the relations in LAMBdA and HMD are virtually identical to those in the two CDLT models. The second feature is that the relations differ mostly on levels. Thus, for example, the curves for HMD and LAMBdA are shifted upwards, and by considerable margins, relative to CDLT regardless of mortality levels. The comparison between HMD and LAMBdA shows that the relation among females in HMD is shifted upward relative to that in LAMBdA, independently of mortality levels, and appears in both time periods: young adult females experience slightly lower average mortality in LAMBdA than they do in HMD at a given level of child mortality. The reverse is true among males: young adult mortality in LAMBdA is somewhat higher relative to young adult mortality in HMD. This is consistent with the idea that LAC in general but some countries within it in particular, experience relative high levels of young adult mortality due to violence and accidents.




(a) LAMBdA 1900-1950

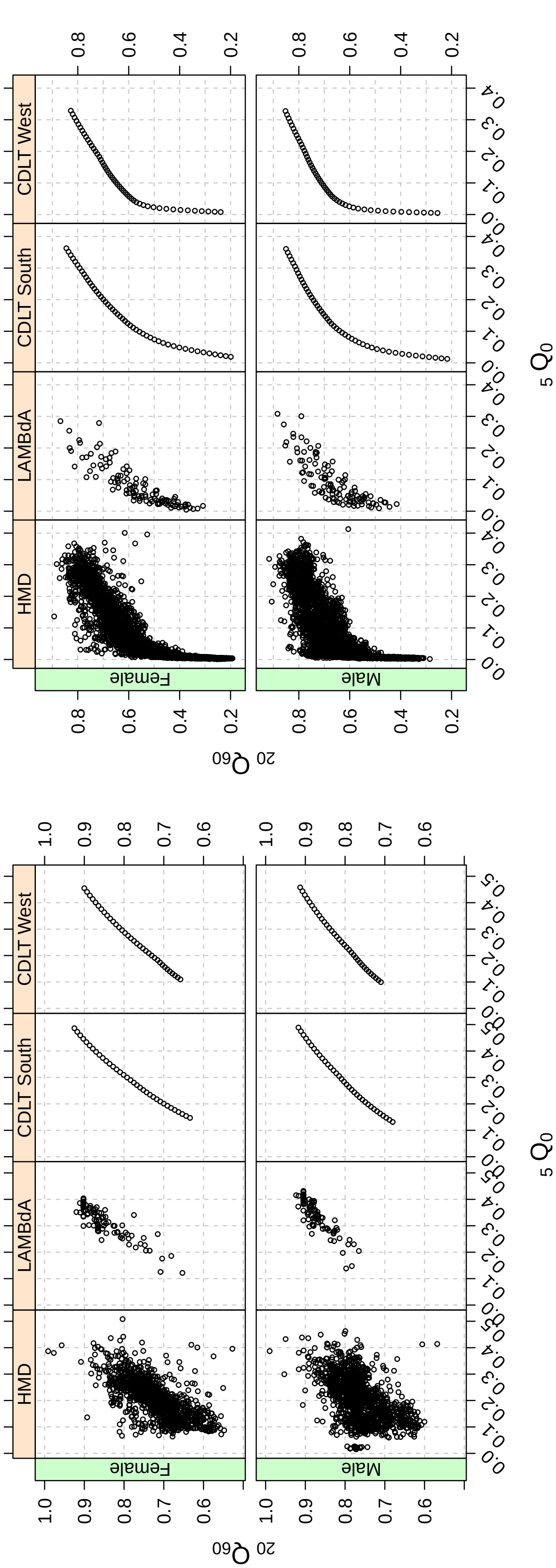
(b) LAMBdA 1950-2010

Figure 8.8: Relation between adult mortality (${}_{45}Q_{15}$) and child mortality (${}_{5}Q_{0}$).

### 8.4.3 Older adult mortality and child mortality

What are the relations between child and late adult mortality? The CDLT models contain distinct patterns of relations between measures of late adult mortality such as ${}_{20}Q_{60}$ and the probability of dying before age 5, (${}_{5}Q_{0}$). In particular, model North has lower late adult mortality than early childhood mortality whereas the reverse is true in models South and East. Figure 8.9 reveals that HMD and LAMBdA share the profile of the relation with models South and West but their levels differ. Thus, LAMBdA's late adult mortality levels are somewhat higher than those in the other three life table systems, particularly at higher levels of child mortality among both males and females. LAMBdA life tables for both females and males contain important outliers with excess late adult mortality at low levels of child mortality. Apart from the outliers just noted, and if judged solely in terms of levels of late adult mortality conditional on child mortality, LAMBdA and HMD patterns are more similar to each other than to those in models South and West.

Thus, although on the whole there are no distinct features setting LAMBdA life tables apart from HMD or from model mortality patterns, the system includes life tables reflecting experiences of systematically higher late adult mortality than expected (given levels of child mortality). This is particularly so in the more recent period at lower levels of child mortality declines.




(a) LAMBdA 1900-1950

(b) LAMBdA 1950-2010

Figure 8.9: Relation between older adult mortality ($_{20}Q_{60}$) and child mortality ($_5Q_0$).

### 8.4.4 Observed and model-pattern age-specific mortality

The final test consists of comparing observed values of $({}_nQ_x)$ for selected age groups and those predicted using relations estimated in Models South and West. The procedure we follow is simplified but succeeds in capturing the main features of the life tables. We first use the parameters estimated by Coale-Demeny (Coale et al., 1983) that relate $({}_nQ_x)$ and life expectancy at age 10, $E_{10}$. Two equations (linear and logarithmic) predict values of ${}_1q_0$, ${}_4q_1$, and ${}_5q_x$ $(0 \leq x \leq 75)$ for both genders. The equations in the CDLT are as follows:

$$ {}_nq_0 = A_x + B_x e_{10} \tag{8.4.1} $$

$$ \log_{10}(10{,}000{}_nq_0) = A_x^{'} + B_x^{'} e_{10} \tag{8.4.2} $$

Coale and Demeny use a blend of these two equations to compute the values of the conditional probabilities corresponding to the various life tables relations. In their words:

> The values of ${}_nq_x$ estimated from the logarithmic regression are always above those from the regression of untransformed mortality rates at the high and low extremes of observed life expectancy, and the logarithmic regression values are always lower in the middle range. In other words, the two regression lines always intersect twice within the range of observations. In constructing the model life tables, ${}_nq_x$ values were taken from the simple regression at all points to the left (i.e., at points with lower life expectancy) of the first intersection of the regression lines; and to the right of the intersection, ${}_nq_x$ value were taken from the logarithmic regression. Between the two intersections, the average of the ${}_nq_x$ values from the regression was used (Coale et al., 1983, p.26).

It should be noted that this methodology to generate estimates of the desired parameters, ${}_nq_x$, creates an opportunity for the emergence of inconsistencies. In fact, while a fixed index value of $e_{10}$ results in predicted values of the elements of the sequence of ${}_nq_x, x = 0, 1, 5, \ldots, 80$, the latter will in general not produce a life table with a value of $e_{10}$ identical to the value used to predict the sequence. Because of this, the comparisons we carried out here should not be interpreted as exact comparisons with life tables patterns embedded in the final Coale-Demeny life tables but rather with those implied by the set of relations linking ${}_nq_x, x = 0, 1, 5, \ldots, 80$ and $e_{10}$. In what follows we compute predicted values of ${}_nq_x$ from equations in Models West and South for each observed value of $e_{10}$ in all country years contained in LAMBdA and then compare them with those included in LAMBdA life tables.

To facilitate the assessment of relations between observed and expected values of the conditional probabilities we chose the following coarse age groups: 0-4, 5-19, 20-44, 45-64 and 65-84. We compute the predicted values for all individual age groups $(x, x+n)$ from age 0 to 80-84 (0, 1-4, 5-9, …, 80-84) and transformed these into measures of integrated hazards $({}_nh_x = -\ln(1 -_n q_x))$ in each age group. The integrated hazards were then added over suitable age intervals $(x, x+n)$ to construct integrated hazards for each coarse age group, ${}_nh_x$ , for age groups 0-4, 5-19, …, 65-84. A similar procedure was used with the observed values on LAMBdA. Finally, we compute relative errors as:

$$RD(x,i,t) = \frac{\texttt{Obs}({}_k h_{x,i,t}) - \texttt{Pred}({}_k h_{x,i,t})}{\texttt{Pred}({}_k h_{x,i,t})} * 100 \qquad (8.4.3)$$

where $x$ is age, $k$ is the length of the age interval, $i$ is country, $t$ is year. $\texttt{Obs}(\cdot)$ and $\texttt{Pred}(\cdot)$ are the observed and predicted values, respectively.

Figures 8.10–8.14 plot observed and predicted integrated hazard values for each of the coarse age groups we select as target.

**Child mortality**

Figure 8.10 displays observed and predicted values of the integrated hazards in the age group 0-5. Male and female LAMBdA life tables for both periods reflect levels of child mortality similar to those in the West pattern (red circles) and lower than those in the South pattern (blue circles). In LAMBdA the contrasts associated with the South model are considerably more marked at higher levels of life expectancy (period 1950-2010) than during the period of higher mortality. The shift probably reflects the large contribution of public health interventions and medical innovations in post-1950 LAC region which could not possibly be reflected in the life tables that are the empirical reference for the South models in the CDLT. The latter are known to represent the highest levels of child mortality encountered in the set of tables that form the basis of the CDLT. Unsurprisingly, the life tables in HMD are much more heterogeneous and the corresponding figures are less compact. Yet, it is plain that the contrast between HMD and Models South is as sharp as the one between models South and LAMBdA.

**Adolescent and young adult mortality.**

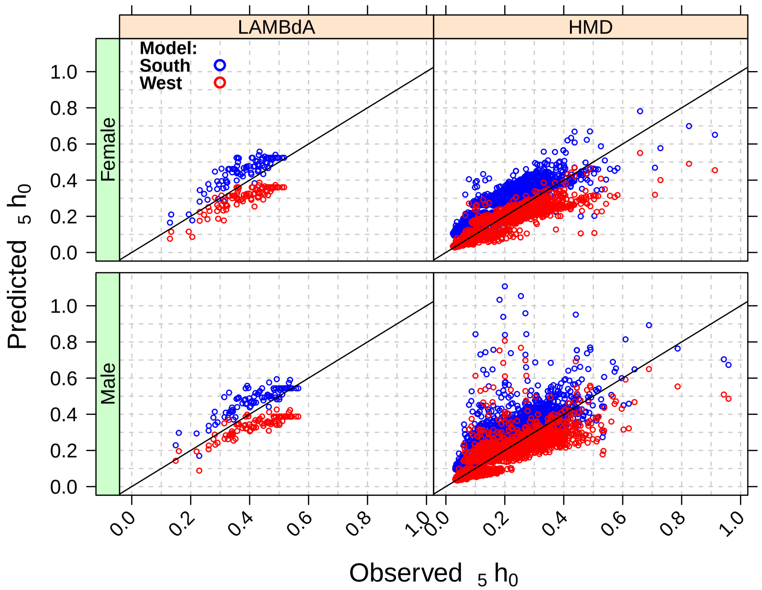
As shown in Figures 8.11 and 8.12, the mortality experience in the age groups 5-14 and 20-44 tightly follows the profiles found in models West and South. There are very few differences irrespective of mortality levels, time period, or gender. If anything, mortality rates in life tables for the pre-1950 period are somewhat lower than expected but if so only by a scarce margin. More heterogeneity is embedded in the HMD female life tables for higher mortality levels where one finds lower than expected mortality.

**Adult mortality.**

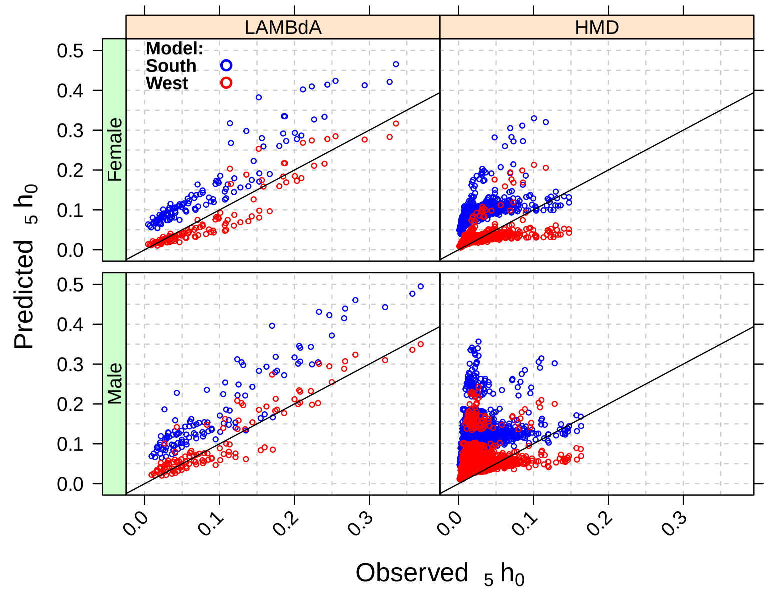
Figure 8.13 shows that LAMBdA mortality in the older adult age group 45-64 follows closely both the West and South patterns, particularly in life tables with higher life expectancy. When life expectancy at age 10 is lower for the pre-1950 period there are slight deviations revealing slightly lower levels than expected. And, here again, the concordance is less in the HMD life tables, particularly so among females.

**Old age mortality.**

The contrasts between observed and expected mortality levels at very old ages (65-84)in LAMBdA are somewhat muted and contain mild gender patterns. Thus, figure 8.14 shows that although observed mortality is higher than expected in more recent life tables (lower mortality levels in post 1950 period), the differences are slightly more marked among males than among females. In life tables for the earliest period, LAMBdA older age mortality is lower than expected, irrespective of the mortality model used to compute expected values. Life tables in HMD, on the other hand, tend to have considerable lower levels of older age mortality than those expected by the South and East model, particularly in life tables for more recent periods with lower levels of mortality. Finally, differences are much sharper among females than among males.

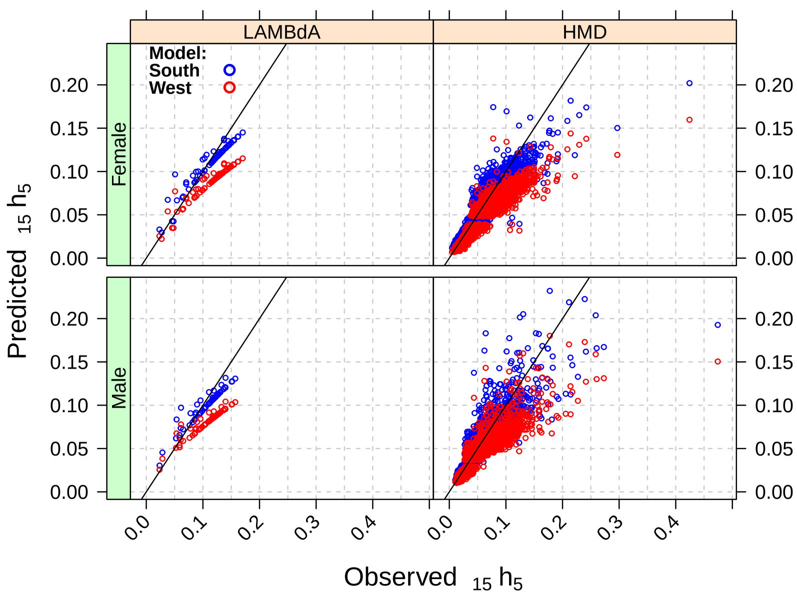



(a) LAMBdA 1900-1950

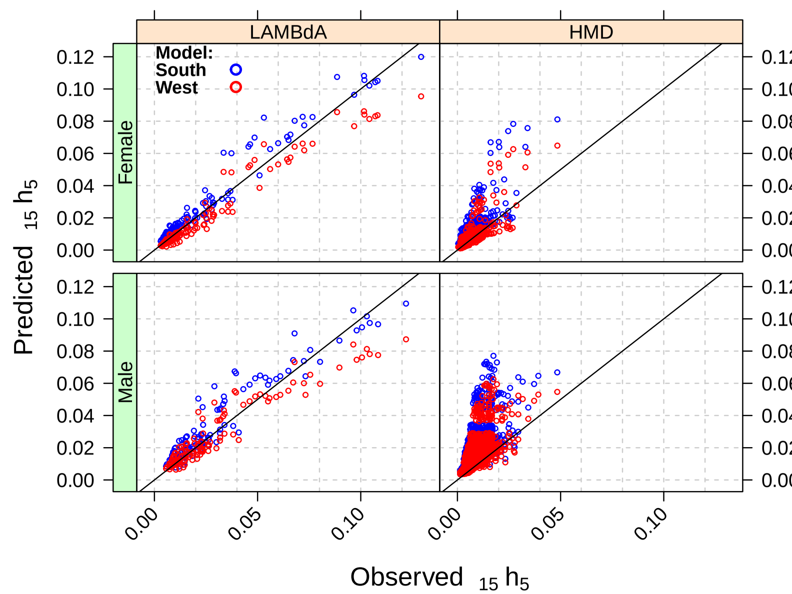



(b) LAMBdA 1950-2010

Figure 8.10: Observed vs. predicted integrated hazard ($_5r_0$). Note: Predicted values are estimated with the use of regression equations from Coale-Demeny Models West and South using $e_{10}$ as input.

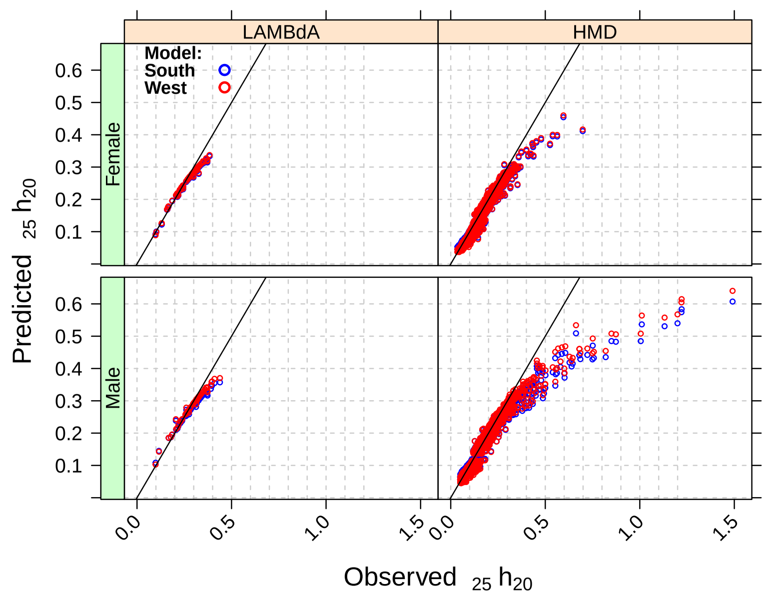



(a) LAMBdA 1900-1950

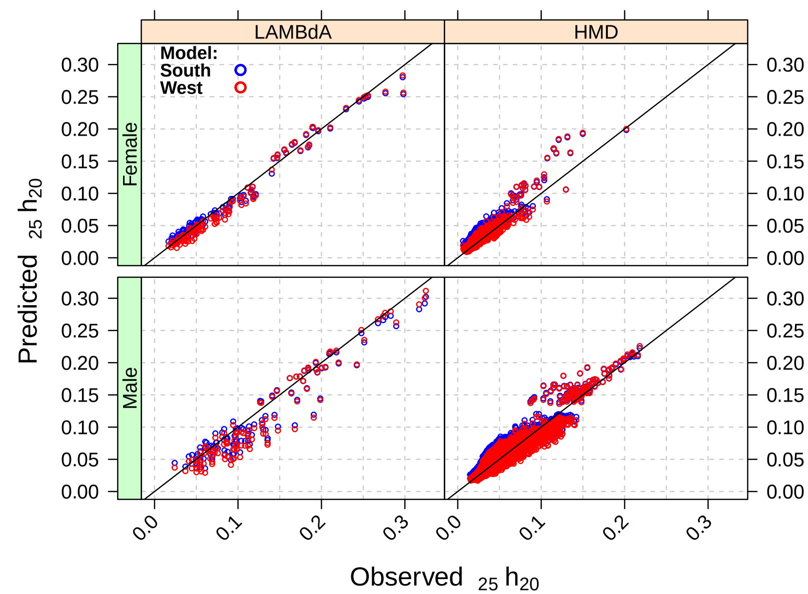



(b) LAMBdA 1950-2010

Figure 8.11: Observed vs. predicted integrated hazard ($_{15}r_5$). Note: Predicted values are estimated with the use of regression equations from Coale-Demeny Models West and South using $e_{10}$ as input.

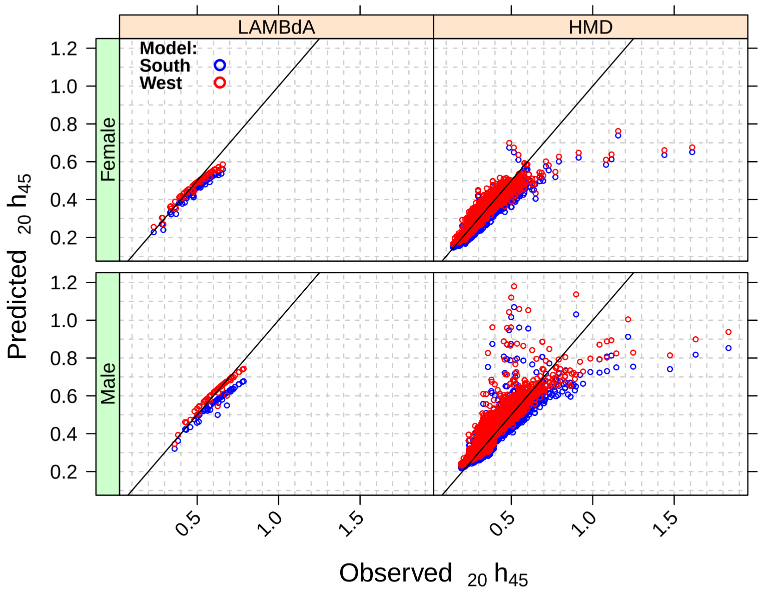



(a) LAMBdA 1900-1950

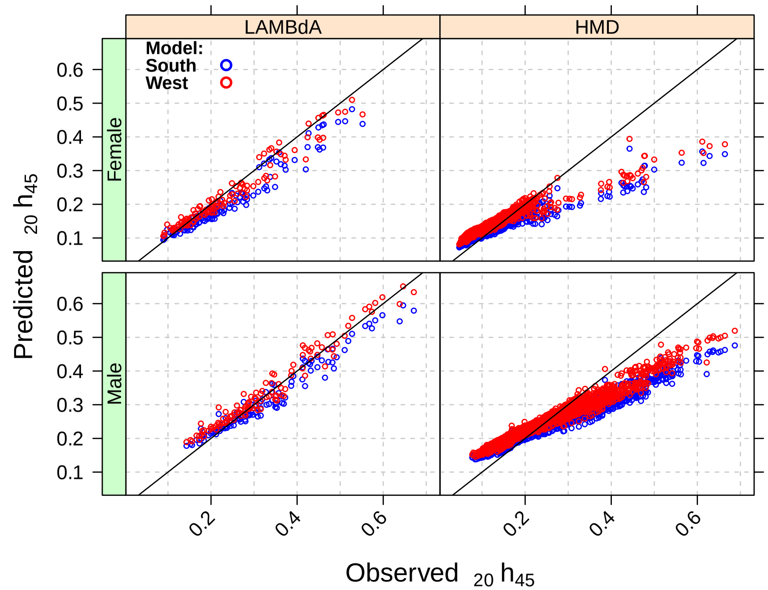



(b) LAMBdA 1950-2010

Figure 8.12: Observed vs. predicted integrated hazard ($_{25}r_{20}$). Note: Predicted values are estimated with the use of regression equations from Coale-Demeny Models West and South using $e_{10}$ as input.

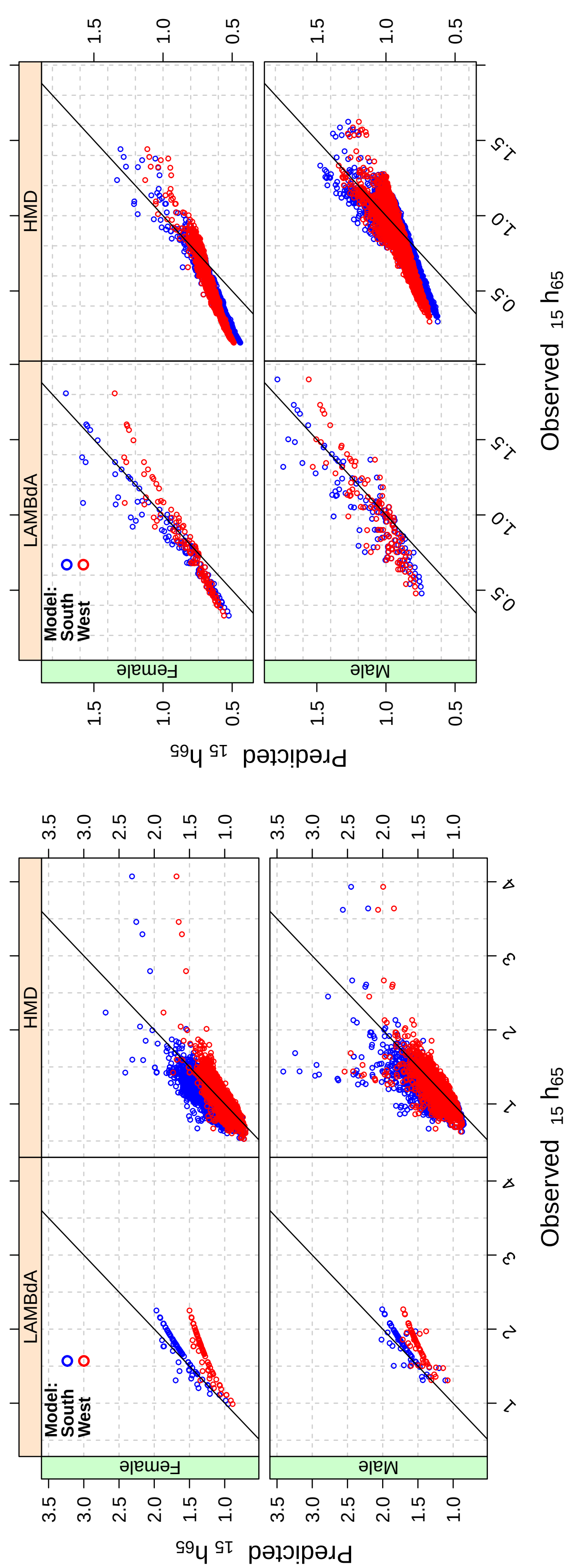



(a) LAMBdA 1900-1950




(b) LAMBdA 1950-2010

Figure 8.13: Observed vs. predicted integrated hazard ($_{20}r_{45}$). Note: Predicted values are estimated with the use of regression equations from Coale-Demeny Models West and South using $e_{10}$ as input.




(a) LAMBdA 1900-1950

(b) LAMBdA 1950-2010

Figure 8.14: Observed vs. predicted integrated hazard (${}_{15}\tau_{65}$). Note: Predicted values are estimated with the use of regression equations from Coale-Demeny Models West and South using $e_{10}$ as input.

### 8.4.5 Detecting outliers

To facilitate detection of systematic patterns of relative differences between observed and "expected" values computed in the previous section, we now estimate regression equations for each coarse age-group (and gender). In these models the dependent variable is the relative difference of the integrated hazards computed before. The independent variables are dummies for countries and years. The reference category for country are six countries with the highest quality vital statistics (Argentina, Chile, Cuba, Costa Rica, Mexico, Uruguay). Dummies for and the year 2010. We use dummies for years with the reference category being 2010.

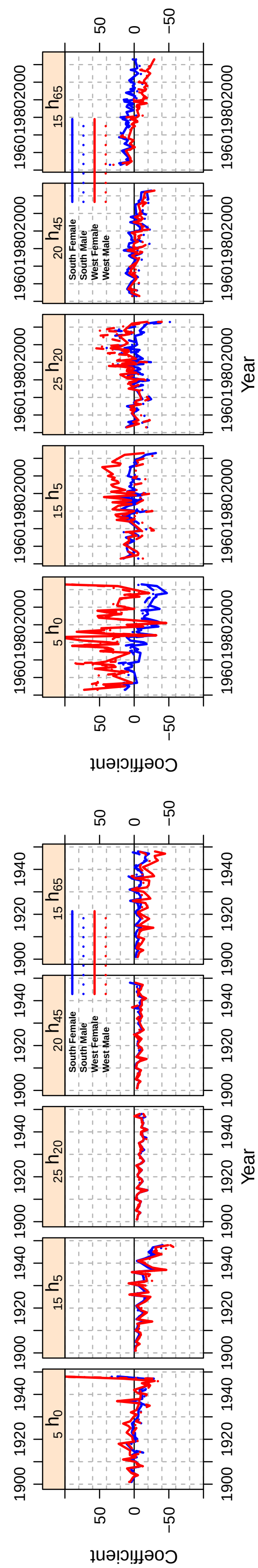
The model is as follows:

$$RD(x,i,t) = \beta_1 \mathtt{Country}_i + \beta_2 \mathtt{Year}_t + \epsilon_{x,i,t} \tag{8.4.4}$$

where $x$ is age, $i$ is country, and $t$ is year.

Figure 8.15 displays the estimated coefficient for the variable year for each coarse age group, ${}_nh_x$ or 0-4, 5-19, ..., 65-84, and in each of two time periods. Figure 8.16 shows the effects of the country-dummy variable.

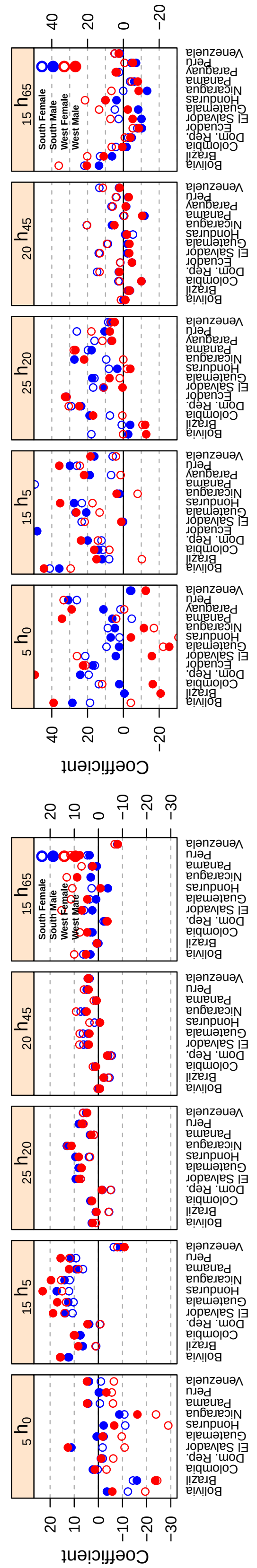
If the differences contained no systematic patterns we would expect to see values of coefficients for the dummy variables to concentrate around 0. With two exceptions this is what happens in the figure displaying estimated effects of year. The exceptions are the first two age groups in the pre-1950 period. There we observe a time trend, e.g. the proportionate differences turn negative as observed values are systematically smaller than expected. A more attenuated pattern is observed in the post 1950 period and then only for mortality below age 5. The figure displaying effects of dummies for countries shows an analogous pattern: there are systematic cross country differences only for mortality during childhood and adolescence but none at older ages. However, even in this case the effects are not associated with a single country but are characteristic of all those not contained in reference category. In all cases the observed values are larger than expected in both the pre and post-1950 period. Because the methodologies used to estimate child mortality in these two periods are quite different, it is improbable that the observed patterns are the result of systematic biases. It is more likely that the observed deviations are an outcome of genuine difference in age patterns of mortality.




(a) LAMBdA 1900-1950

(b) LAMBdA 1950-2010

Figure 8.15: Regression coefficients by age group for the variable Year.




(a) LAMBdA 1900-1950

(b) LAMBdA 1950-2010

Figure 8.16: Regression coefficients by age group for the variable Country.

## 8.5 Gender patterns

In this section we compute statistics to asses relations and contrasts between male and female mortality. Human mortality patterns are characterized by fairly constrained associations between male and female mortality across ages. Departures from these associations occur under extreme conditions (exogenous shocks such as wars and armed conflicts in general, and epidemics (e.g., HIV-AIDS)) or across cultural settings where social protection, nutritional status and/or access to health care are gender biased. Under more or less standard conditions males tend to have excess mortality at all ages, including infancy and early childhood, at young adult ages and then at older ages. A mortality data base for LAC countries should reproduce expected patterns even though one should expect deviations in populations subjected to abnormal conditions.

We choose a handful of statistics used in Section 2, namely, ${}_5Q_0$, ${}_{65}e_0$, ${}_{80}e_{65}$, and $e_{80}$. As a benchmark we compute the regression lines relating these statistics in Models West and South. Figures 8.17-8.20 display the main results.

The statistics for the age groups 0-64 display no distinct patterns. The relations are regular, follow the West and South patterns, there are no singular outliers and, finally, HMD and LAMBdA behave quite similarly. Differences emerge at ages older than 65. The two statistics we use for this age group show a distinct pattern of lower than expected female mortality that is evident in both LAMBdA and HMD and is particularly salient at higher levels of life expectancy and in the most recent periods. As before, these results do not reveal any evidence of obvious errors or aberrations but may reflect genuine differences contained in HMD and LAMBdA life tables when compared to those in the CDLT system.

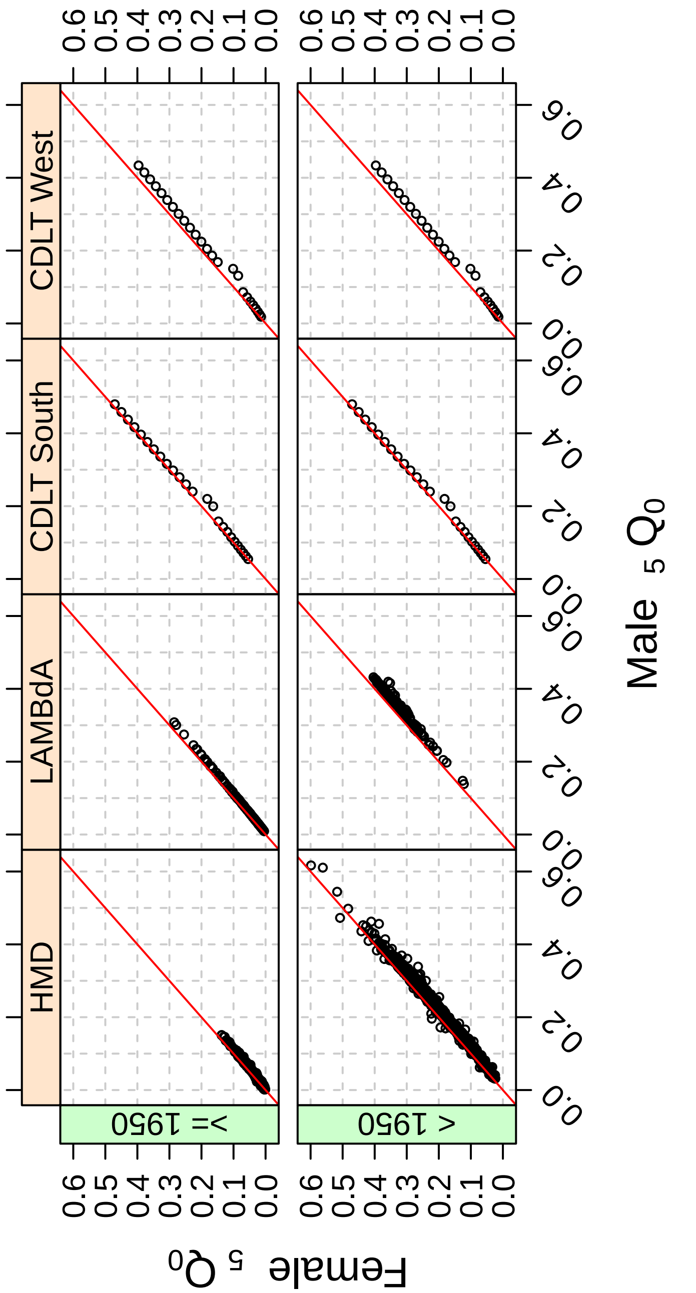



Figure 8.17: Male vs. female child mortality ($_5Q_0$)

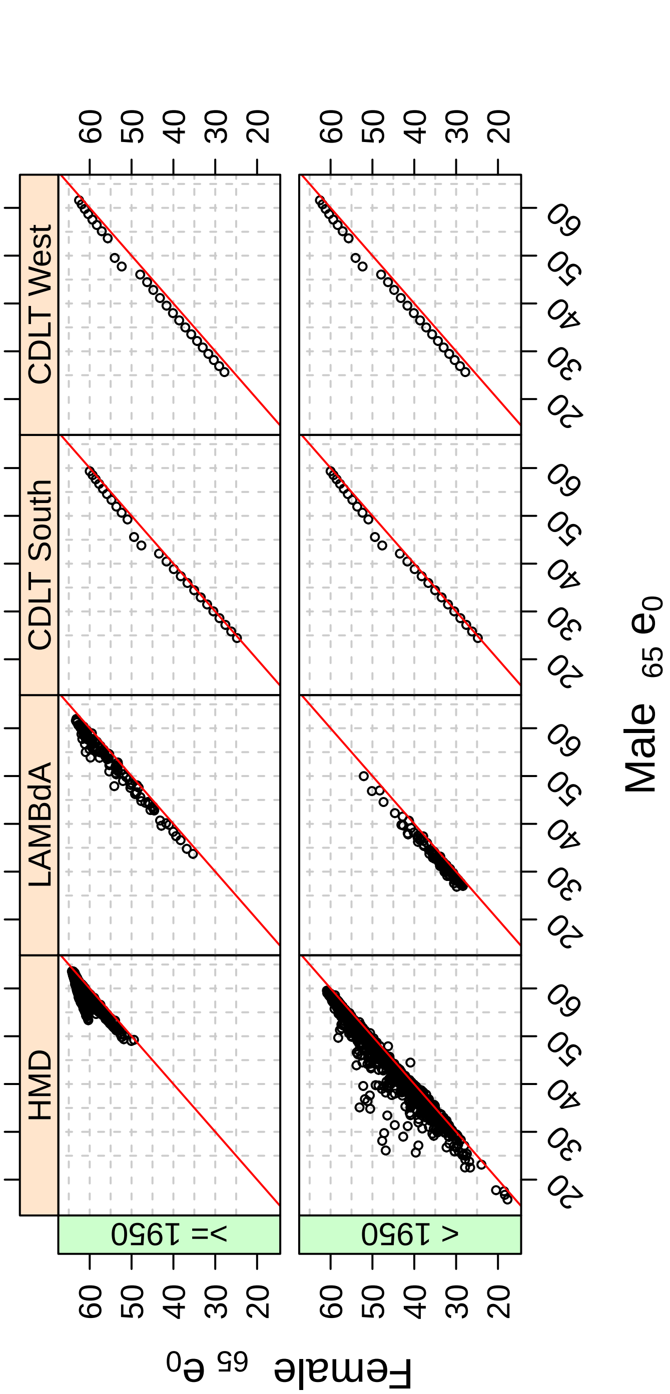



Figure 8.18: Male vs. female average years of life lived between ages 0 and 65 ($_{65}e_0$)

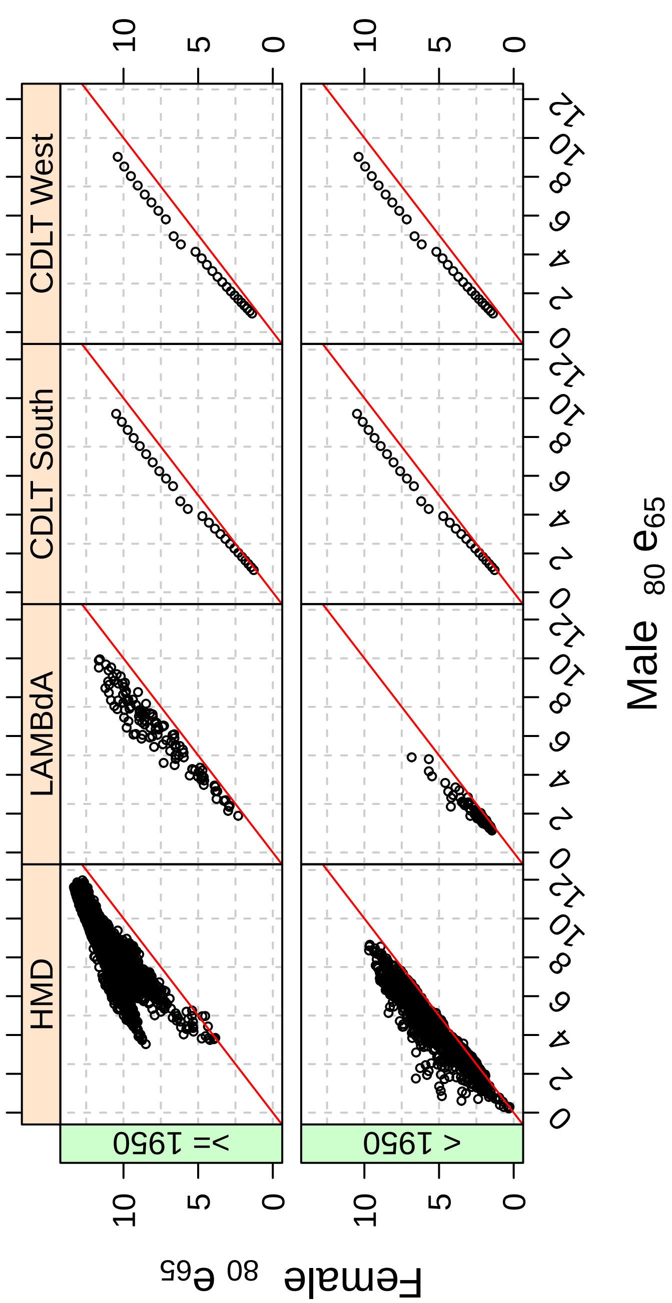



Figure 8.19: Male vs. female average years of life lived between ages 65 and 80 ($_{80}e_{65}$)

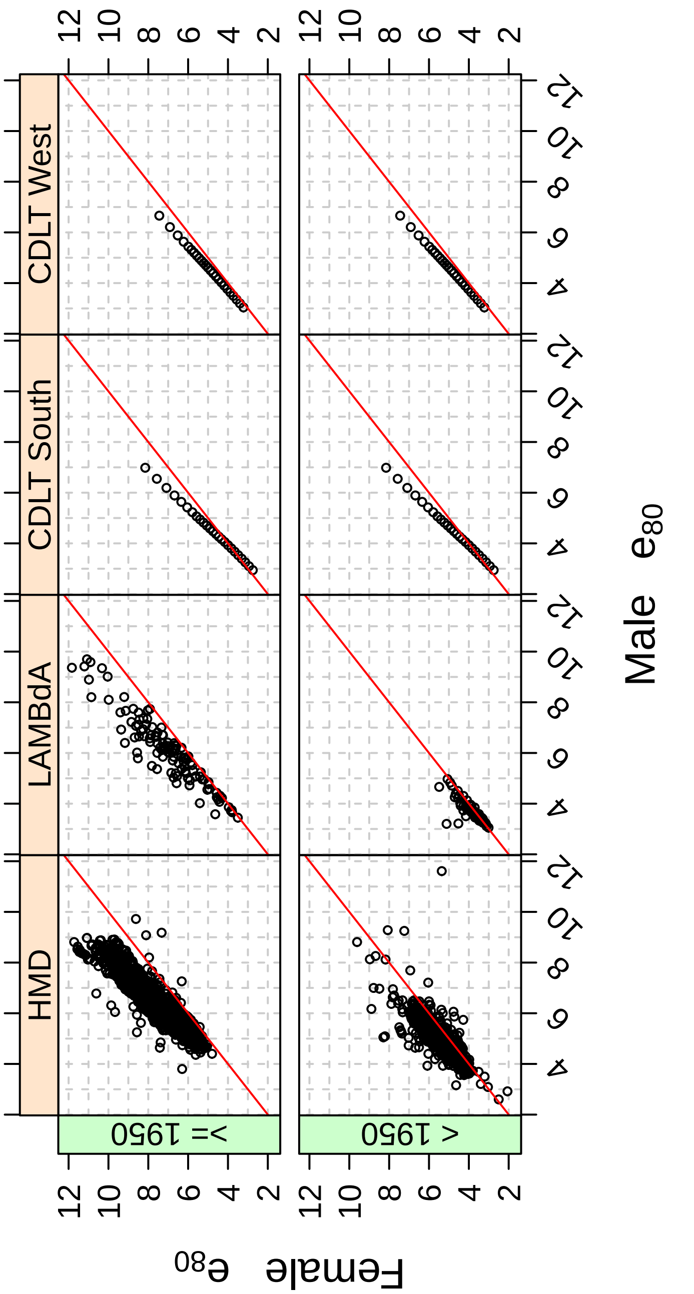



Figure 8.20: Male vs. female Life Expectancy at age 80 ($e_{80}$)

## 8.6 Direct and indirect estimation: LAMBdA compared with estimates from orphanhood statistics

### 8.6.1 Introduction

LAMBdA adult pivotal life tables for the period after 1950 were almost entirely estimated using direct methods, that is, adjusted vital statistics (LAMBdA team, 2020) However, LAMBdA also contains alternative estimates of mortality that users can, if desired, employ either alone or in combination with the pivotal life tables (and derivatives such as the yearly life tables). This is in keeping with the idea that there is no true value for mortality statistics in the region but rather a set of estimates, some better than others, all subjected to uncertainty. Included among the alternative estimates are those obtained via application of orphanhood, widowhood or sisterhood methods designed to estimate female and/or male adult mortality from raw data collected in surveys or census that elicit information on orphanhood, widowhood, and sister survival from eligible respondents (Hill and Trussell, 1977; Brass and Bamgboye, 1981; Palloni et al., 1984; Timaeus, 1991; Timaeus et al., 1996; Danel et al., 1996; Gakidou et al., 2004; Gakidou and King, 2006; Hill, 2000).

In this report we compare LAMBdA's estimates with those obtained via maternal orphanhood and assess concordance between the two sets of quantities. Estimates obtained with widowhood and sisterhood methods are more sensitive to violations of assumptions and to errors of reporting and less desirable as benchmarks. Instead, (maternal) orphanhood methods are more widely used than alternative indirect techniques to estimate adult mortality. And, even if not as robust as Brass type of methods to estimate childhood mortality, in the absence of vital registration they surely can be used with less qualifications than those derived from paternal orphanood, widowhood or sisterhood.

### 8.6.2 Brief summary of orphanhood methods

Orphanhood methods are designed to estimate male and female adult mortality from data on the survival status of respondents' parents. The information is collected via answers to questions such as: Is your mother alive? and Is your father alive? There is no need to elicit the actual dates of death or even the ages at death of the deceased parent. The information is then arranged by age group of respondents and, through a transformation of the fraction of parents reported dead, one obtains estimates of conditional survivorship ratios or probabilities of dying between some standard ages, such as ${}_{45}q_{15}$, ${}_{15}q_{35}$, or ${}_{30}q_{30}$. If one assumes an underlying (model) mortality pattern, these indexes can be transformed into estimates of life expectancy at birth or at adult ages, say 60.

As all indirect demographic procedures, orphanhood methods have limitations. We highlight the two most important ones. First, information of parental survival can only be collected from offspring who are alive at the time of the survey, that is, who survive the high mortality risks to which they are exposed during infancy and early childhood. When parental and child death are correlated, either because both are responsive to common, shared conditions, or because the former directly affects the latter, the data on orphanhood will generate estimates contaminated by selection biases. For the most part these biases will lead to underestimate adult mortality.

The second, lesser, flaw is rooted in the fact that parents with larger family size will be overrepresented in orphanhood reports relative to parents with smaller family size. This would not be of importance were it not for the case that, in most populations where the method is applied, larger family sizes are associated with higher parental (and child) mortality. This will lead to overestimates of adult mortality and offset somewhat the biases due to selection.

### 8.6.3 Methods

To produce a database of adult mortality orphanhood estimates for LAC countries we collected information from surveys that include suitable items on maternal orphanhood and applied the methodology proposed by the United Nations and associated software (United-Nations, 1988, 1983). When it was not possible to access the raw information we used instead estimates of conditional cumulative probabilities of dying (or surviving) produced and published by researchers who did have access to the original survey information (Timaeus et al., 1996). In all cases we converted the estimates of cumulative conditional survival into life expectancies at ages 0, 5 and 60. To carry out this operation we utilize the South Model in the Coale-Demeny life table system. Finally, we plotted LAMBdA estimates of life expectancies obtained via multiple procedures and the values obtained via maternal orphanhood methods. The graphs below display plots with life expectancy at birth, E(0), at age 5, E(5) and at age 60, E(60).

Two caveats are important. First, the comparisons are based on maternal, not paternal, orphanhood since the former are considered to be more reliable and less vulnerable to errors associated with union disruption, remarriage, and parental abandonment. Despite this, paternal orphanhood estimates are correlated with maternal orphanhood estimates and inferences we draw from comparisons with LAMBdA and maternal orphanhood estimates are unlikely to be severely affected.

Second, the transformation of survival ratios estimated by orphanhood methods into life expectancy indicators is done for convenience since researchers are likely to be more inclined to use them in lieu of the actual conditional cumulative probabilities of dying (surviving) directly produced by the application of orphanhood methods. The transformation has a cost, however, in that comparisons between LAMBdA and orphanhood indicators will be sensitive to the choice of the mortality model (in our case, the South female Coale-Demeny model (Coale et al., 1983)). This is definitely the case for E(0) but less so for E(5) and E(60). The graphs we show below use E(0), E(5), and E(60) and thus provide "bounds" for the comparisons in the following sense: if inferences using E(0) are the same as those using E(5) and E(60) then the choice of model is irrelevant. If inferences are inconsistent, the choice of model plays a role in the discordance and the comparison using E(0) is flawed but much less so the comparison that uses either E(5) or E(60).

There is, however, an important drawback to the comparisons of synthetic indicators. It is possible that the estimated values of E(0) (or E(5) or E(60)) in LAMBdA and those associated with orphanhood methods are very similar even though the *conditional probabilities estimated from the data on orphanhood* are different from those in the LAMBdA life tables. This can occur if the mortality patterns (relation between early and adult mortality) embedded in the South model and in LAMBdA are different. Thus, consistency of estimates gauged by using synthetic indicators does not always mean consistency of adult mortality levels. The opposite situation is also possible: even if levels of adult mortality in LAMBdA are similar to those implied by orphanhood estimates, the associated values of E(0) (and E(5) and E(60)) could be different. Thus, lack of consistency between estimates of E(0) (and E(5) and E(60)) from orphanhood and those from LAMBdA does not necessarily reveal the existence of discrepancies in the levels of adult mortality implied by orphanhood information and those contained in LAMBdA life tables.

### 8.6.4 Results

Figures 8.21 through 8.29 include plots of life expectancies at birth and at ages 5 and 60 for all 9 countries with information on maternal orphanhood. These countries are Bolivia (Fig 8.21), Brazil (Fig 8.22), Colombia (Fig 8.23), Dominican Republic (Fig 8.24), Guatemala (Fig 8.25), Honduras (Fig 8.26), Mexico (Fig 8.27), Nicaragua (Fig 8.28), and Peru (Fig 8.29). These data span an interval of time stretching from 1960 to 1987. We highlight three relevant results. First, in all cases LAMBdA's and alternative orphanhood estimates of E(0) establish highly comparable time trends and, with two exceptions (Brazil and Dominican Republic), they point to very similar levels of mortality.

Second, discrepancies between estimates of E(0) are very small and in the two cases, Brazil and Dominican Republic, where they are more noticeable they are in opposite directions. Thus, in Brazil the orphanhood estimates are higher than those in LAMBdA, as it should be if the former are contaminated by selection biases. In the case of Dominican Republic, however, the situation is just the opposite, as would be expected if the second source of bias identified above were more important.

Third, as was the case for E(0), the appearance of time trends embedded in estimates of E(5) and E(60) is highly consistent across types of estimates. However, discrepancies in magnitudes are slightly more marked for E(60) than they are for E(0) (except in Brazil and Dominican Republic). To facilitate interpretation of these results Table 8.1 classifies countries according to the type of agreement between estimates of E(0) and E(60). The key finding is that the discrepancies are mostly concentrated in E(60), not in E(0), and these are always the result of E(60) from LAMBdA life tables being smaller than E(60) from orphanhood methods. We interpret this regularity as a result of differences in the patterns of adult mortality embedded in the two sets of estimates. In fact, LAMBdA mortality rates at older adult ages (older than 45) are adjusted upwards for age misreporting (see Chapter 3). These corrections shift upward mortality rates at older adult ages *relative to those at younger adult ages*. The South model of mortality is also known to have higher excess adult mortality (at age over 55) relative to younger mortality than the other three mortality patterns of the Coale-Demeny system. It follows that the age pattern of adult mortality implied by LAMBdA estimates follows the South age pattern model while adding excess older adult mortality. That is, the older adult mortality patterns in LAC countries is "worse" in LAMBdA than in the South model in the sense that older age mortality rates are higher than would be expected in the South (or the other) model patterns given the same levels of younger adult mortality.

It would be foolhardy, however, to exaggerate the observed differences between the two sets of estimates instead of being startled by the overall level of consistency and agreement between them. Indeed, with the exception of Brazil and Dominican Republic, the estimates are very close in magnitude and, without exceptions, they point to remarkably similar slopes of the time trend of improvements in adult survival.

Table 8.1: Classification of countries according to behaviors of estimates of E(0) and E(60)

| Table 1: Classification of countries according to behaviors of estimates of E(0) and E(60) | |
|---|---|
| | |
| Some discrepancies with:<br>$E(0)^{LAMBdA} < E(0)^{orph}$<br>$E(60)^{LAMBdA} < E(60)^{orph}$ | Brazil, Dominican Republic |
| Consistent with:<br>$E(0)^{LAMBdA} < E0)^{orph}$<br>$E(60)^{LAMBdA} \sim E(60)^{orph}$ | ----- |
| Consistent with:<br>$E(0)^{LAMBdA} \sim E0)^{orph}$<br>$E(60)^{LAMBdA} < E(60)^{orph}$ | Bolivia, Colombia, Honduras |
| Highly Consistent with<br>$E(0)^{LAMBdA} \sim E(0)^{orph}$<br>$E(60)^{LAMBdA} < E(60)^{orph}$ | Guatemala, Mexico, Nicaragua, Peru |

Figure 8.21: Bolivia: Plots of life expectancy at birth (E(0)), and ages 5(E5) and 60(E60) from orphanhood methods and from LAMBdA (multiple methods)

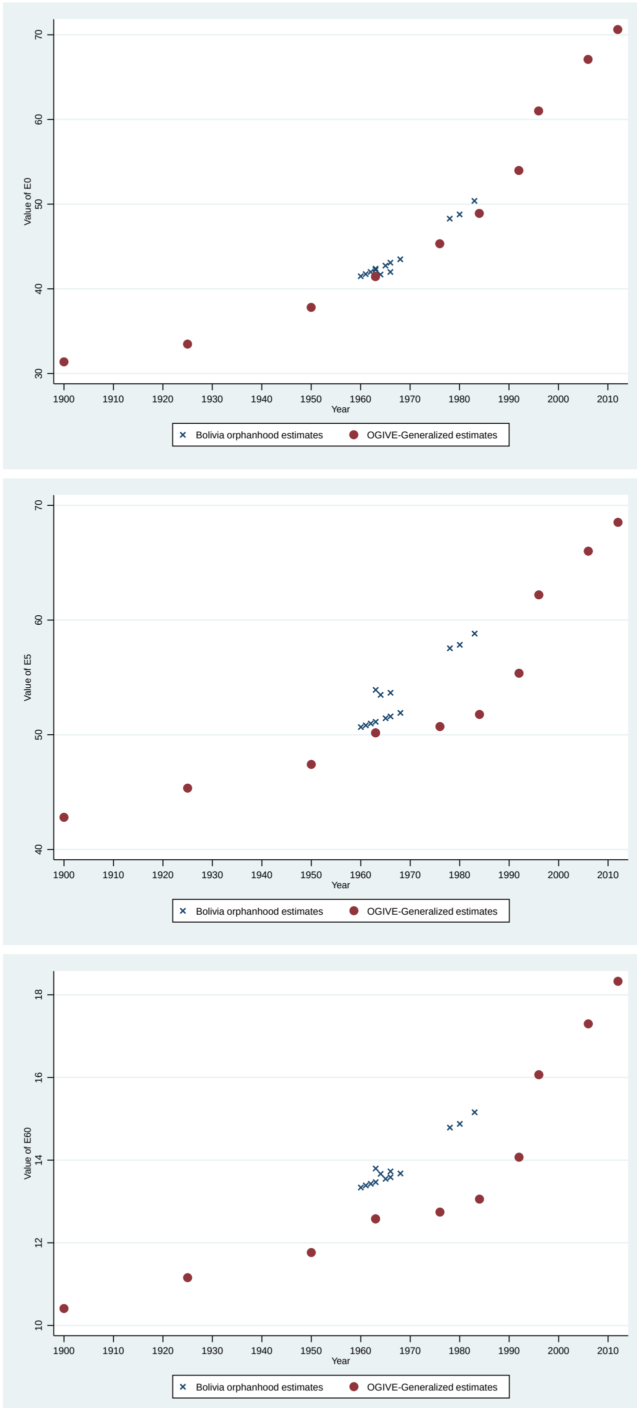


Figure 8.22: Brazil: Plots of life expectancy at birth (E(0)), and ages 5(E5) and 60(E60) from orphanhood methods and from LAMBdA (multiple methods)

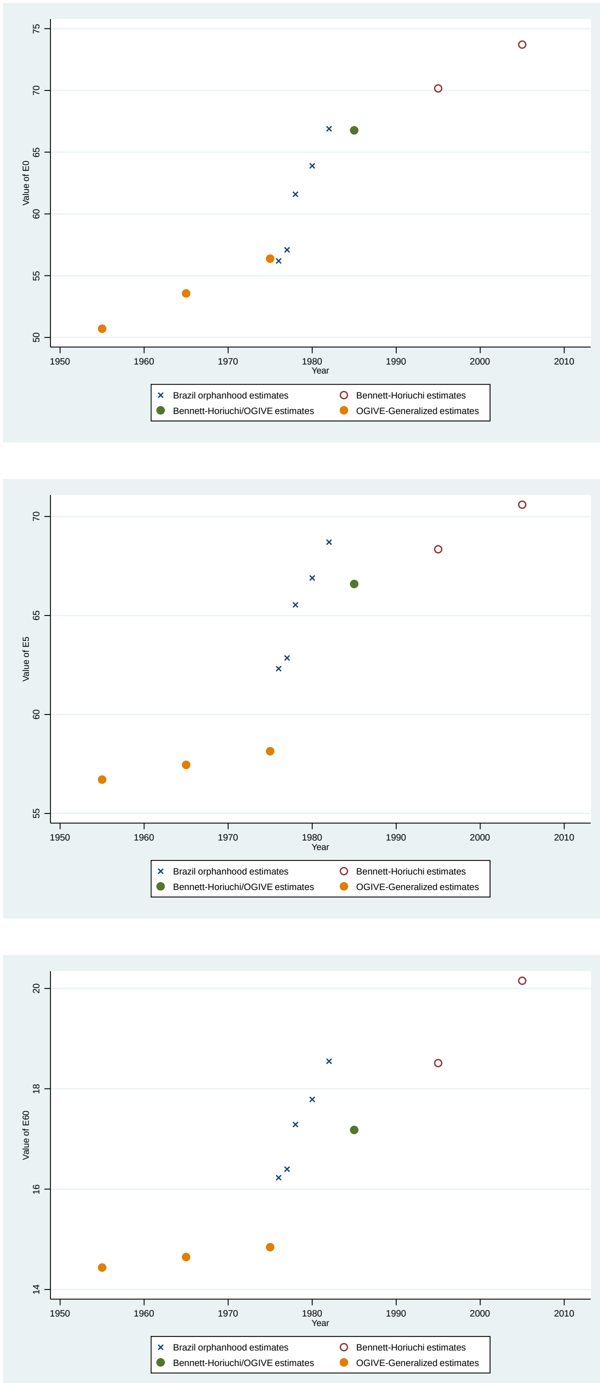


Figure 8.23: Colombia: Plots of life expectancy at birth (E(0)), and ages 5(E5) and 60(E60) from orphanhood methods and from LAMBdA (multiple methods)

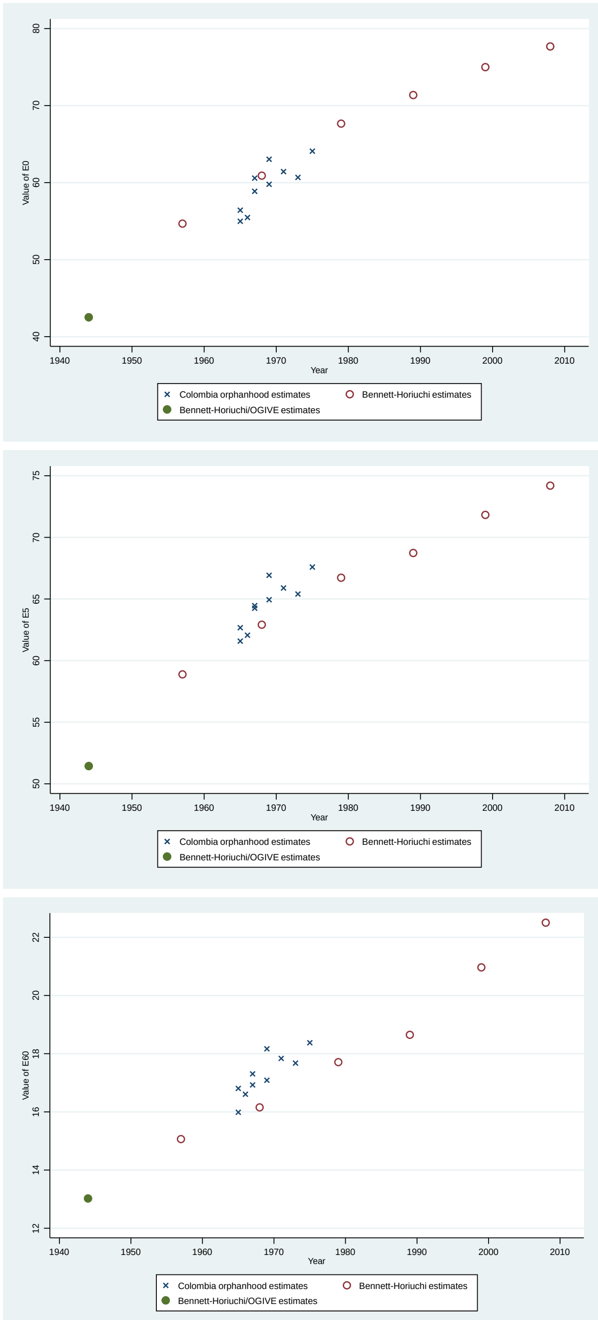


Figure 8.24: Dominican Republic: Plots of life expectancy at birth (E(0)), and ages 5(E5) and 60(E60) from orphanhood methods and from LAMBdA (multiple methods)

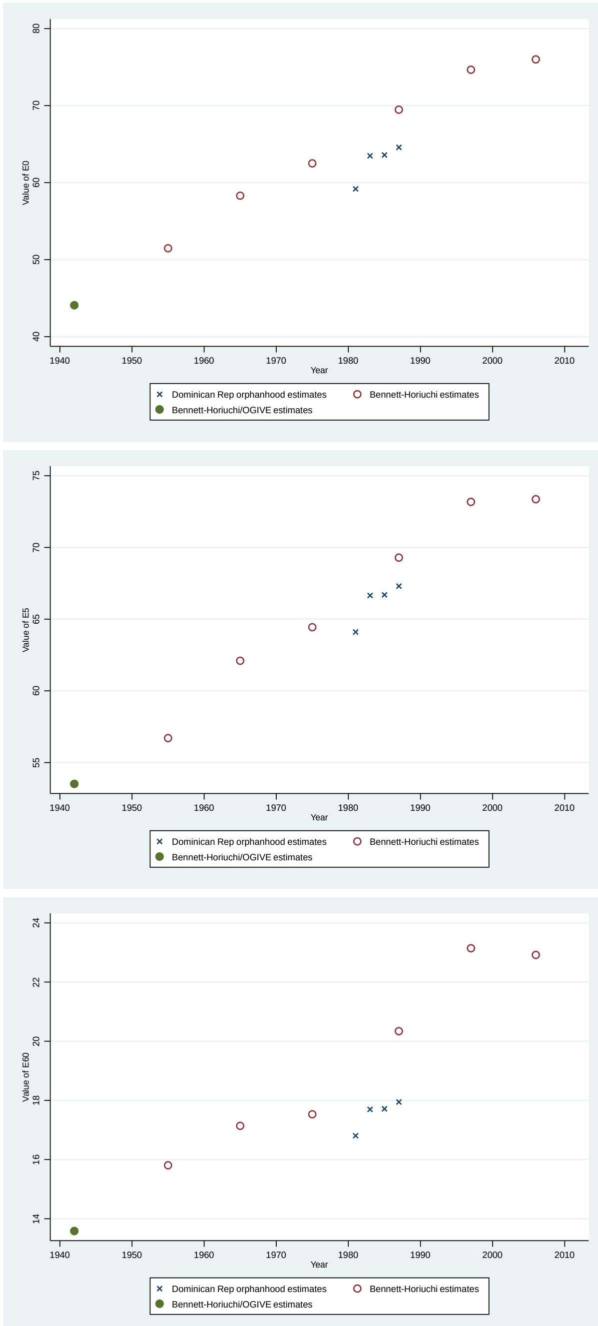


Figure 8.25: Guatemala: Plots of life expectancy at birth (E(0)), and ages 5(E5) and 60(E60) from orphanhood methods and from LAMBdA (multiple methods)

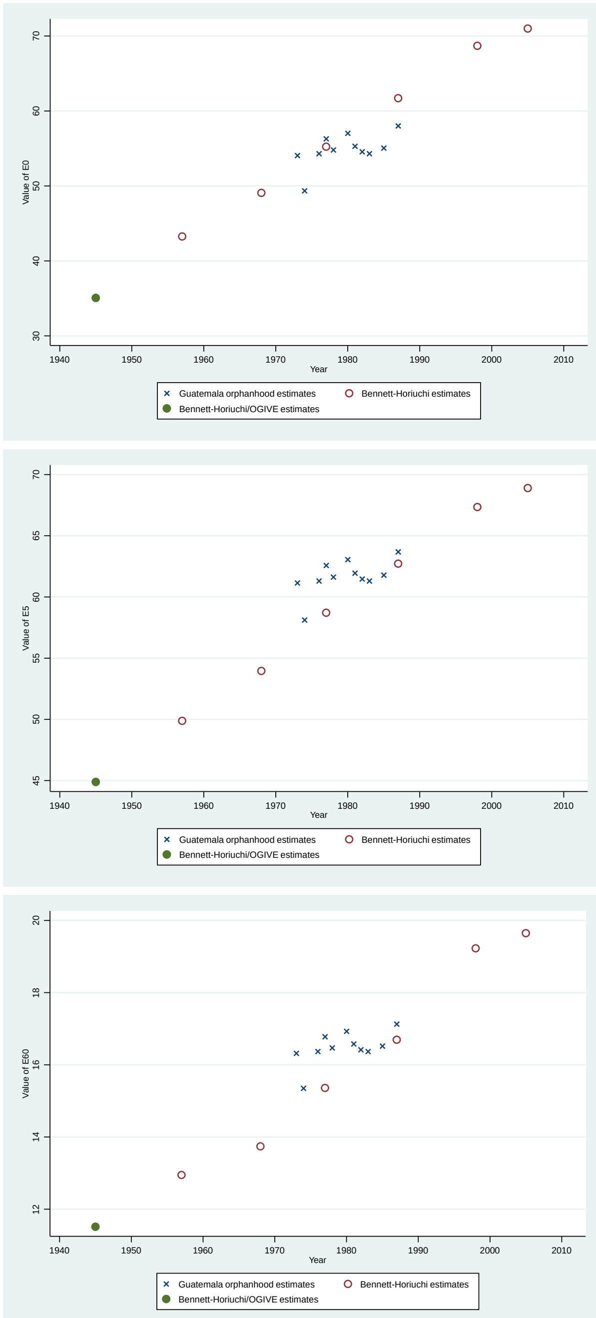


Figure 8.26: Honduras: Plots of life expectancy at birth (E(0)), and ages 5(E5) and 60(E60) from orphanhood methods and from LAMBdA (multiple methods)

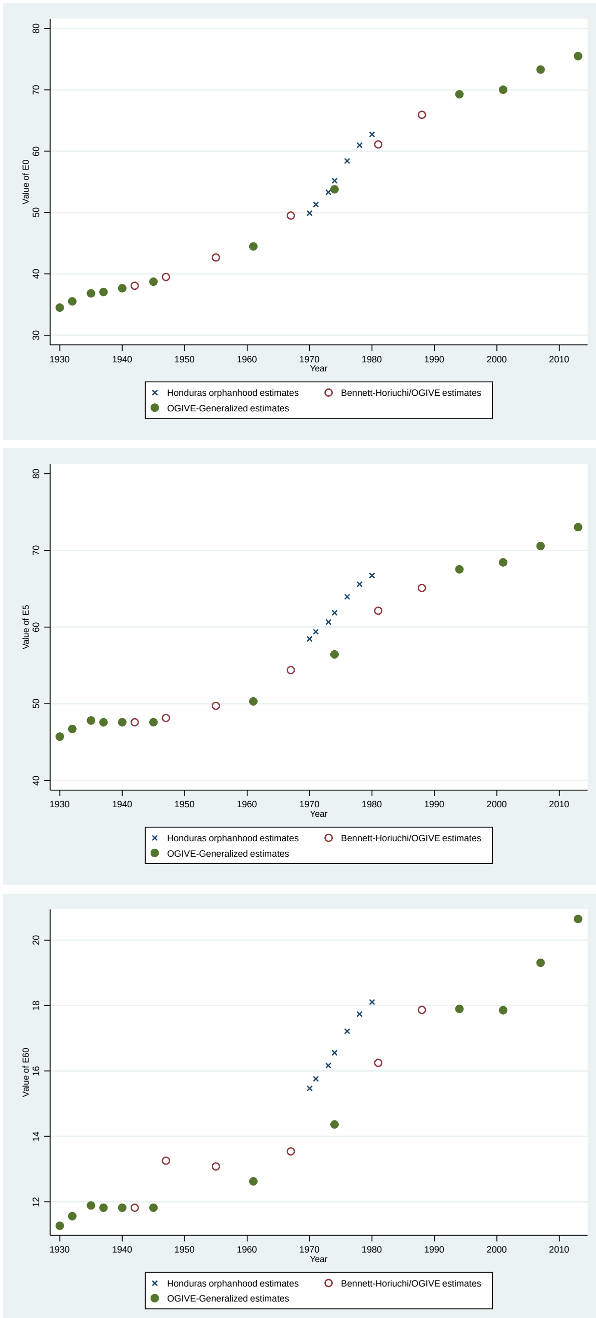


Figure 8.27: Mexico: Plots of life expectancy at birth (E(0)), and ages 5(E5) and 60(E60) from orphanhood methods and from LAMBdA (multiple methods)

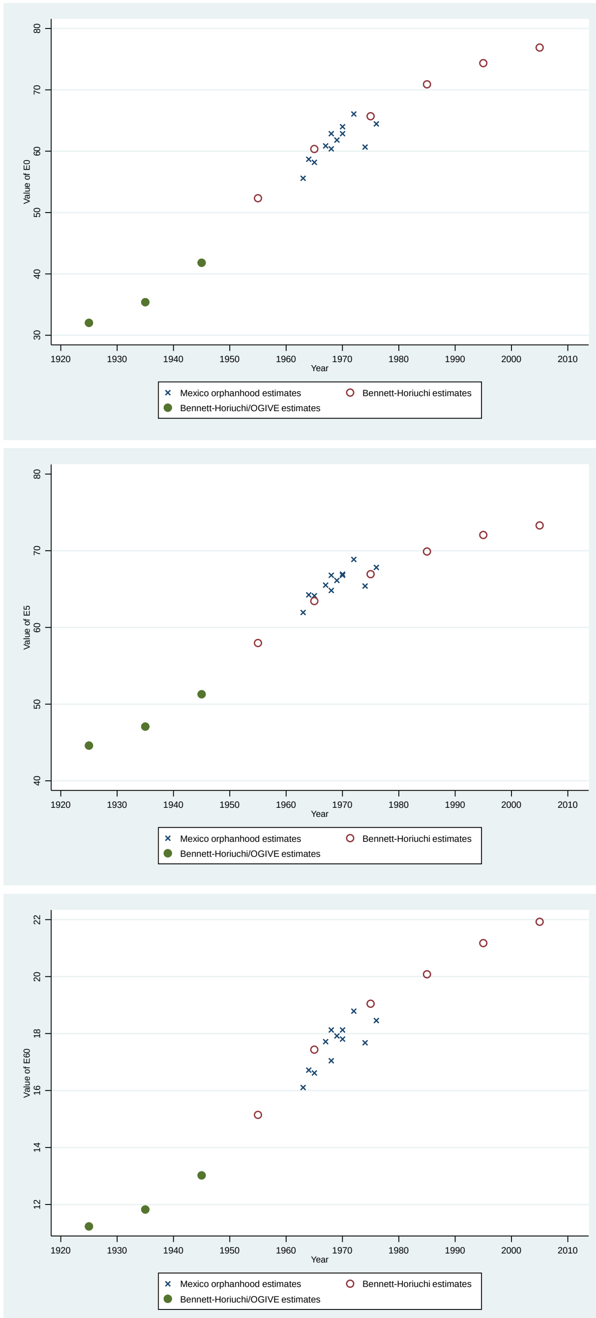


Figure 8.28: Nicaragua: Plots of life expectancy at birth (E(0), and ages 5(E5) and 60(E60) from orphanhood methods and from LAMBdA (multiple methods)

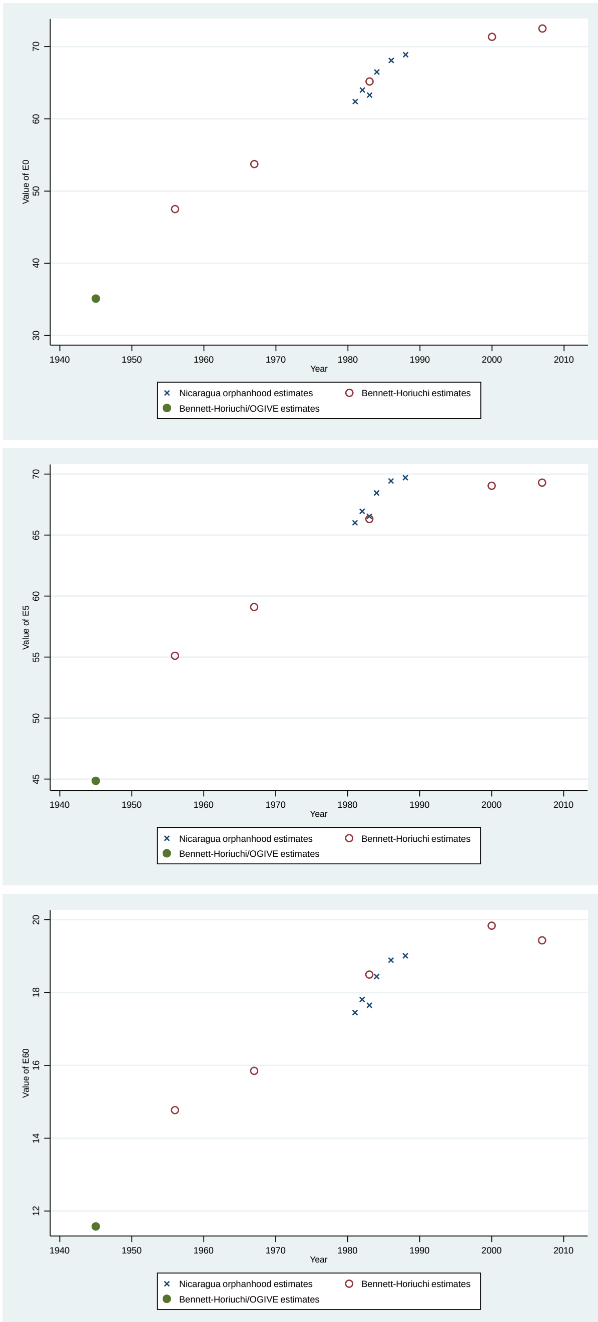
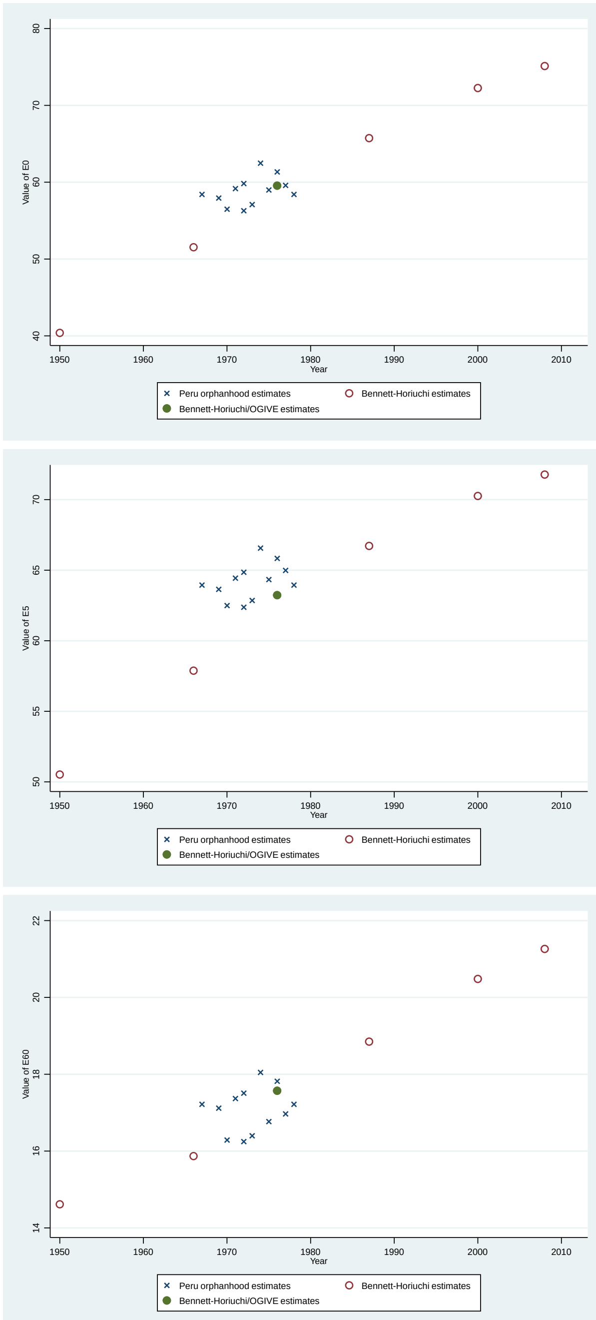


Figure 8.29: Peru: Plots of life expectancy at birth (E(0)), and ages 5(E5) and 60(E60) from orphanhood methods and from LAMBdA (multiple methods)



## 8.7 Summary statistics

### 8.7.1 Average years of life lived between ages 0 and 65

Table 8.2: Median values of the average years of life lived between ages 0 and 65 by mortality level, sex, and dataset: LAMBdA < 1950

| | | Mortality level, $e_0$ | | | |
|---|---|---|---|---|---|
| sex | data | [25,30) | [30,40) | [40,50) | [50,60) |
| Male | HMD | 26.1 | 35.1 | 40.4 | 49.7 |
| Male | LAMBdA | 27.8 | 31.7 | 39.4 | 47.0 |
| Male | CDLT South | 25.7 | 32.2 | 40.6 | 48.4 |
| Male | CDLT West | 26.0 | 32.6 | 41.0 | 49.0 |
| Female | HMD | 27.7 | 34.7 | 40.4 | 48.5 |
| Female | LAMBdA | 28.4 | 32.4 | 39.7 | 49.2 |
| Female | CDLT South | 25.6 | 32.0 | 40.1 | 47.8 |
| Female | CDLT West | 25.8 | 32.3 | 40.5 | 48.3 |

Table 8.3: Interquartile range of the average years of life lived between ages 0 and 65 by mortality level, sex, and dataset: LAMBdA < 1950

| | | Mortality level, $e_0$ | | | |
|---|---|---|---|---|---|
| sex | data | [25,30) | [30,40) | [40,50) | [50,60) |
| Male | HMD | 1.7 | 3.0 | 4.1 | 3.7 |
| Male | LAMBdA | 0.6 | 3.6 | 2.9 | 1.6 |
| Male | CDLT South | 1.8 | 3.8 | 3.7 | 3.4 |
| Male | CDLT West | 1.8 | 3.9 | 3.7 | 3.5 |
| Female | HMD | 1.0 | 3.3 | 3.8 | 4.5 |
| Female | LAMBdA | 0.0 | 4.2 | 3.0 | 2.6 |
| Female | CDLT South | 1.8 | 3.8 | 3.6 | 3.3 |
| Female | CDLT West | 1.8 | 3.8 | 3.6 | 3.4 |

Table 8.4: Median values of the average years of life lived between ages 0 and 65 by mortality level, sex, and dataset: LAMBdA > 1950

| | | Mortality level, $e_0$ | | | | | |
|---|---|---|---|---|---|---|---|
| sex | data | [35,40) | [40,50) | [50,60) | [60,70) | [70,80) | [80,85) |
| Male | HMD | 35.6 | 40.4 | 49.7 | 58.5 | 61.9 | 63.4 |
| Male | LAMBdA | 34.8 | 43.0 | 50.4 | 56.8 | 60.3 | |
| Male | CDLT South | 34.3 | 40.6 | 48.4 | 55.6 | 61.1 | 63.3 |
| Male | CDLT West | 34.7 | 41.0 | 49.0 | 56.5 | 62.2 | 63.8 |
| Female | HMD | 35.2 | 40.4 | 48.5 | 56.1 | 61.9 | 63.6 |
| Female | LAMBdA | 35.3 | 41.8 | 48.7 | 55.5 | 60.8 | 63.1 |
| Female | CDLT South | 34.1 | 40.1 | 47.8 | 54.7 | 60.3 | 62.8 |
| Female | CDLT West | 34.4 | 40.5 | 48.3 | 55.6 | 61.6 | 63.5 |

Table 8.5: Interquartile range of the average years of life lived between ages 0 and 65 by mortality level, sex, and dataset: LAMBdA > 1950

| | | Mortality level, $e_0$ | | | | | |
|---|---|---|---|---|---|---|---|
| sex | data | [35,40) | [40,50) | [50,60) | [60,70) | [70,80) | [80,85) |
| Male | HMD | 1.8 | 4.1 | 3.7 | 2.5 | 1.4 | 0.1 |
| Male | LAMBdA | 1.4 | 4.2 | 3.6 | 2.7 | 1.2 | |
| Male | CDLT South | 1.7 | 3.7 | 3.4 | 2.9 | 1.8 | 0.4 |
| Male | CDLT West | 1.7 | 3.7 | 3.5 | 3.2 | 1.6 | 0.3 |
| Female | HMD | 1.9 | 3.8 | 4.5 | 3.5 | 1.4 | 0.3 |
| Female | LAMBdA | 0.0 | 4.3 | 3.2 | 3.7 | 1.9 | 0.4 |
| Female | CDLT South | 1.6 | 3.6 | 3.3 | 2.9 | 1.9 | 0.5 |
| Female | CDLT West | 1.7 | 3.6 | 3.4 | 3.1 | 1.7 | 0.3 |

### 8.7.2 Average years of life lived between ages 65 and 80

Table 8.6: Median values of the average years of life lived between ages 65 and 80 by mortality level, sex, and dataset: LAMBdA < 1950

| | | Mortality level, $e_0$ | | | |
|---|---|---|---|---|---|
| sex | data | [25,30) | [30,40) | [40,50) | [50,60) |
| Male | HMD | 1.1 | 2.5 | 3.3 | 5.1 |
| Male | LAMBdA | 1.3 | 1.8 | 2.9 | 4.8 |
| Male | CDLT South | 1.3 | 2.1 | 3.6 | 5.3 |
| Male | CDLT West | 1.0 | 1.7 | 3.1 | 4.7 |
| Female | HMD | 1.2 | 2.7 | 3.6 | 5.4 |
| Female | LAMBdA | 1.5 | 2.0 | 3.5 | 5.7 |
| Female | CDLT South | 1.4 | 2.3 | 3.9 | 5.8 |
| Female | CDLT West | 1.1 | 2.0 | 3.5 | 5.2 |

Table 8.7: Interquartile range of the average years of life lived between ages 65 and 80 by mortality level, sex, and dataset: LAMBdA < 1950

| | | Mortality level, $e_0$ | | | |
|---|---|---|---|---|---|
| sex | data | [25,30) | [30,40) | [40,50) | [50,60) |
| Male | HMD | 0.4 | 0.6 | 0.7 | 1.0 |
| Male | LAMBdA | 0.0 | 0.4 | 0.5 | 0.4 |
| Male | CDLT South | 0.2 | 0.6 | 0.7 | 0.9 |
| Male | CDLT West | 0.2 | 0.5 | 0.7 | 0.8 |
| Female | HMD | 1.0 | 0.5 | 0.7 | 1.1 |
| Female | LAMBdA | 0.0 | 0.7 | 0.9 | 0.3 |
| Female | CDLT South | 0.2 | 0.6 | 0.8 | 0.9 |
| Female | CDLT West | 0.2 | 0.6 | 0.7 | 0.8 |

Table 8.8: Median values of the average years of life lived between ages 65 and 80 by mortality level, sex, and dataset: LAMBdA > 1950

| | | Mortality level, $e_0$ | | | | | |
|---|---|---|---|---|---|---|---|
| sex | data | [35,40) | [40,50) | [50,60) | [60,70) | [70,80) | [80,85) |
| Male | HMD | 2.6 | 3.3 | 5.1 | 6.9 | 9.3 | 11.7 |
| Male | LAMBdA | 2.1 | 3.4 | 5.1 | 7.1 | 9.1 | |
| Male | CDLT South | 2.5 | 3.6 | 5.3 | 7.5 | 10.4 | 12.6 |
| Male | CDLT West | 2.1 | 3.1 | 4.7 | 6.8 | 9.7 | 12.3 |
| Female | HMD | 2.8 | 3.6 | 5.4 | 7.4 | 10.1 | 12.2 |
| Female | LAMBdA | 2.3 | 3.3 | 5.0 | 7.7 | 9.9 | 11.6 |
| Female | CDLT South | 2.7 | 3.9 | 5.8 | 8.1 | 10.8 | 12.8 |
| Female | CDLT West | 2.3 | 3.5 | 5.2 | 7.3 | 10.0 | 12.4 |

Table 8.9: Interquartile range of the average years of life lived between ages 65 and 80 by mortality level, sex, and dataset: LAMBdA > 1950

| | | Mortality level, $e_0$ | | | | | |
|---|---|---|---|---|---|---|---|
| sex | data | [35,40) | [40,50) | [50,60) | [60,70) | [70,80) | [80,85) |
| Male | HMD | 0.4 | 0.7 | 1.0 | 0.9 | 1.7 | 0.2 |
| Male | LAMBdA | 0.2 | 1.0 | 1.1 | 1.1 | 0.8 | |
| Male | CDLT South | 0.3 | 0.7 | 0.9 | 1.1 | 1.4 | 0.5 |
| Male | CDLT West | 0.3 | 0.7 | 0.8 | 1.0 | 1.6 | 0.6 |
| Female | HMD | 0.5 | 0.7 | 1.1 | 1.2 | 1.2 | 0.6 |
| Female | LAMBdA | 0.0 | 0.8 | 0.8 | 1.3 | 1.2 | 0.2 |
| Female | CDLT South | 0.3 | 0.8 | 0.9 | 1.1 | 1.3 | 0.5 |
| Female | CDLT West | 0.3 | 0.7 | 0.8 | 1.0 | 1.5 | 0.5 |

### 8.7.3 Average years of life lived between ages 0 and 80

Table 8.10: Median values of the average years of life lived between ages 0 and 80 by mortality level, sex, and dataset: LAMBdA < 1950

| | | Mortality level, $e_0$ | | | |
|---|---|---|---|---|---|
| sex | data | [25,30) | [30,40) | [40,50) | [50,60) |
| Male | HMD | 27.3 | 37.7 | 43.5 | 54.9 |
| Male | LAMBdA | 29.1 | 33.5 | 42.2 | 51.8 |
| Male | CDLT South | 26.9 | 34.3 | 44.1 | 53.7 |
| Male | CDLT West | 26.9 | 34.3 | 44.1 | 53.8 |
| Female | HMD | 28.8 | 37.5 | 44.1 | 54.0 |
| Female | LAMBdA | 29.9 | 34.6 | 43.2 | 54.9 |
| Female | CDLT South | 26.9 | 34.3 | 44.0 | 53.6 |
| Female | CDLT West | 26.9 | 34.3 | 44.0 | 53.6 |

Table 8.11: Interquartile range of the average years of life lived between ages 0 and 80 by mortality level, sex, and dataset: LAMBdA < 1950

| | | Mortality level, $e_0$ | | | |
|---|---|---|---|---|---|
| sex | data | [25,30) | [30,40) | [40,50) | [50,60) |
| Male | HMD | 2.0 | 3.6 | 4.9 | 4.5 |
| Male | LAMBdA | 0.6 | 4.0 | 3.6 | 1.9 |
| Male | CDLT South | 2.0 | 4.4 | 4.4 | 4.3 |
| Male | CDLT West | 2.0 | 4.4 | 4.4 | 4.3 |
| Female | HMD | 0.8 | 3.5 | 4.5 | 5.3 |
| Female | LAMBdA | 0.0 | 5.0 | 4.1 | 2.9 |
| Female | CDLT South | 2.0 | 4.4 | 4.3 | 4.2 |
| Female | CDLT West | 2.0 | 4.4 | 4.3 | 4.3 |

Table 8.12: Median values of the average years of life lived between ages 0 and 80 by mortality level, sex, and dataset: LAMBdA > 1950

| | | Mortality level, $e_0$ | | | | | |
|---|---|---|---|---|---|---|---|
| sex | data | [35,40) | [40,50) | [50,60) | [60,70) | [70,80) | [80,85) |
| Male | HMD | 38.2 | 43.5 | 54.9 | 65.3 | 71.1 | 75.1 |
| Male | LAMBdA | 36.9 | 46.6 | 55.3 | 64.0 | 69.3 | |
| Male | CDLT South | 36.8 | 44.1 | 53.7 | 63.1 | 71.5 | 75.9 |
| Male | CDLT West | 36.8 | 44.1 | 53.8 | 63.2 | 71.9 | 76.1 |
| Female | HMD | 37.9 | 44.1 | 54.0 | 63.5 | 72.0 | 75.9 |
| Female | LAMBdA | 37.7 | 45.0 | 53.6 | 63.2 | 70.7 | 74.5 |
| Female | CDLT South | 36.8 | 44.0 | 53.6 | 62.8 | 71.1 | 75.5 |
| Female | CDLT West | 36.7 | 44.0 | 53.6 | 62.9 | 71.6 | 75.8 |

Table 8.13: Interquartile range of the average years of life lived between ages 0 and 80 by mortality level, sex, and dataset: LAMBdA > 1950

| | | Mortality level, $e_0$ | | | | | |
|---|---|---|---|---|---|---|---|
| sex | data | [35,40) | [40,50) | [50,60) | [60,70) | [70,80) | [80,85) |
| Male | HMD | 1.9 | 4.9 | 4.5 | 3.4 | 3.3 | 0.3 |
| Male | LAMBdA | 1.7 | 5.3 | 5.2 | 3.7 | 1.8 | |
| Male | CDLT South | 2.0 | 4.4 | 4.3 | 4.1 | 3.2 | 0.9 |
| Male | CDLT West | 2.0 | 4.4 | 4.3 | 4.2 | 3.2 | 0.9 |
| Female | HMD | 2.2 | 4.5 | 5.3 | 4.7 | 2.5 | 0.9 |
| Female | LAMBdA | 0.0 | 5.2 | 4.2 | 5.4 | 3.0 | 0.6 |
| Female | CDLT South | 2.0 | 4.3 | 4.2 | 4.0 | 3.2 | 1.0 |
| Female | CDLT West | 1.9 | 4.3 | 4.3 | 4.2 | 3.2 | 0.9 |

### 8.7.4 Life Expectancy at age 80

Table 8.14: Median values of the life expectancy at 80 by mortality level, sex, and dataset: LAMBdA < 1950

| | | Mortality level, $e_0$ | | | |
|---|---|---|---|---|---|
| sex | data | [25,30) | [30,40) | [40,50) | [50,60) |
| Male | HMD | 4.1 | 4.7 | 4.7 | 5.0 |
| Male | LAMBdA | 3.2 | 3.5 | 4.0 | 4.3 |
| Male | CDLT South | 3.1 | 3.6 | 4.1 | 4.6 |
| Male | CDLT West | 3.6 | 4.1 | 4.6 | 5.0 |
| Female | HMD | 4.0 | 4.8 | 5.0 | 5.4 |
| Female | LAMBdA | 3.2 | 3.4 | 4.3 | 4.5 |
| Female | CDLT South | 3.0 | 3.5 | 4.2 | 4.8 |
| Female | CDLT West | 3.7 | 4.3 | 4.9 | 5.3 |

Table 8.15: Interquartile range of the life expectancy at 80 by mortality level, sex, and dataset: LAMBdA < 1950

| | | Mortality level, $e_0$ | | | |
|---|---|---|---|---|---|
| sex | data | [25,30) | [30,40) | [40,50) | [50,60) |
| Male | HMD | 0.8 | 0.6 | 0.5 | 0.6 |
| Male | LAMBdA | 0.0 | 0.3 | 0.4 | 0.6 |
| Male | CDLT South | 0.1 | 0.3 | 0.2 | 0.2 |
| Male | CDLT West | 0.1 | 0.3 | 0.2 | 0.2 |
| Female | HMD | 1.9 | 0.6 | 0.6 | 0.8 |
| Female | LAMBdA | 0.0 | 0.4 | 0.5 | 0.7 |
| Female | CDLT South | 0.2 | 0.3 | 0.2 | 0.2 |
| Female | CDLT West | 0.2 | 0.3 | 0.2 | 0.2 |

Table 8.16: Median values of the life expectancy at 80 by mortality level, sex, and dataset: LAMBdA > 1950

| | | Mortality level, $e_0$ | | | | | |
|---|---|---|---|---|---|---|---|
| sex | data | [35,40) | [40,50) | [50,60) | [60,70) | [70,80) | [80,85) |
| Male | HMD | 4.7 | 4.7 | 5.0 | 5.5 | 6.8 | 8.6 |
| Male | LAMBdA | 3.7 | 5.0 | 5.8 | 6.2 | 6.9 | |
| Male | CDLT South | 3.7 | 4.1 | 4.6 | 5.1 | 6.2 | 8.7 |
| Male | CDLT West | 4.2 | 4.6 | 5.0 | 5.3 | 6.3 | 9.0 |
| Female | HMD | 4.8 | 5.0 | 5.4 | 5.9 | 6.9 | 9.0 |
| Female | LAMBdA | 3.5 | 4.9 | 6.2 | 6.8 | 8.0 | 10.5 |
| Female | CDLT South | 3.7 | 4.2 | 4.8 | 5.3 | 6.5 | 9.0 |
| Female | CDLT West | 4.4 | 4.9 | 5.3 | 5.7 | 6.5 | 9.2 |

Table 8.17: Interquartile range of the life expectancy at 80 by mortality level, sex, and dataset: LAMBdA > 1950

| | | Mortality level, $e_0$ | | | | | |
|---|---|---|---|---|---|---|---|
| sex | data | [35,40) | [40,50) | [50,60) | [60,70) | [70,80) | [80,85) |
| Male | HMD | 0.5 | 0.5 | 0.6 | 0.6 | 1.2 | 0.5 |
| Male | LAMBdA | 1.1 | 1.5 | 1.3 | 1.6 | 2.3 | |
| Male | CDLT South | 0.1 | 0.2 | 0.2 | 0.3 | 1.0 | 0.8 |
| Male | CDLT West | 0.1 | 0.2 | 0.2 | 0.2 | 1.0 | 0.9 |
| Female | HMD | 0.7 | 0.6 | 0.8 | 0.8 | 1.2 | 0.8 |
| Female | LAMBdA | 0.0 | 1.7 | 1.8 | 1.3 | 1.4 | 1.4 |
| Female | CDLT South | 0.1 | 0.2 | 0.2 | 0.3 | 1.1 | 0.8 |
| Female | CDLT West | 0.1 | 0.2 | 0.2 | 0.2 | 1.0 | 0.9 |

# Chapter 9

# Country data reports

## 9.1 Introduction

The following is a country-by-country descriptions of date sources and methods employed to produce estimates throughout the period under observation. For additional clarification about the nature of methods and final estimates, please refer to Chapters 2 and 3.

Because in some countries we were able to produce multiple estimates for time intervals that overlap or were adjacent to each other, all were considered candidate estimates in one or more pivotal years. These multiple estimates may originate in the generalized OGIVE, Brass or Bennett-Horiuchi methods. In most of these cases the estimates lined up and overlapped substantially, led to smooth time trends and were not inconsistent among themselves nor with pivotal estimates for adjacent periods. In a handful of cases, however, the estimates were inconsistent in that one or more of them implied implausible shifts of mortality levels in short periods of time relative to (a) the remaining alternative estimates and/or (b) a time trend established by pivotal estimates for periods before or immediately following the interval of time with multiple estimates. In these cases we used a **conciliation** procedure based on three rules:

1. Selected Deletions: Always discard any of the multiple estimates that imply implausible shifts in adult mortality relative to time trends established by the set of estimates for the period immediately preceding and following the time interval with competing estimates. In these cases we chose the median of estimates that remain under consideration.

2. Preserve by choosing median: If two or more competing estimates clustered around a plausible mortality level, e.g. did not depart from a time trend, we choose the median value.

3. Preserve according to method: In all other cases, choose among competing estimates assigning priority to those obtained with the following methods (see Chapter 3): Bennett-Horiuchi, Martin's modification of Brass(BMartin),, and generalized OGIVE.

## 9.2 Argentina

### 9.2.1 Vital statistics

**Deaths**

- 1912, 1913, 1914, 1915 (ages: 0, 1-5, 6-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-69, 70-79, 80-89, 90+)
- 1947-1979 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-29, 30-39, 40-49, 50-59, 60+)
- 1980- 1981 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+)
- 1982-1996 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+)
- 1997-2016 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85-89, 90-94, 95+)

**Causes of Death**

- ICD7: 1966-1968
- ICD8: 1969, 1970, 1977, 1978
- ICD9: 1979-1996
- ICD10: 1997-2016

**Births**

- 1910-2012

**Vital statistics sources**

- Anuario Demográfico de la Argentina: 1912, 1913, 1914, 1915
- United Nations Yearbook: 1936-2010
- World Health Organization. Health statistics and information systems: 1966-2016 https://www.who.int/healthinfo/statistics/mortality_rawdata/en/

### 9.2.2 Population censuses

- 1869 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79, 80+)
- 1895, 1914, 1947, 1960, 1970, 1980, 1991, 2001, 2010 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+)

**Population censuses sources**

- Census from Argentinean Statistical Office: 1869, 1895, 1914, 1947
- United Nations Yearbook: 1948-2012

### 9.2.3 Adult pivotal life tables (above age 5)

**Adjustments to generate adult pivotal life tables**

1. Completeness

   Pivotal years for life table estimated using generalized OGIVE method: 1869, 1882, 1895, 1904, 1914, 1930

   Pivotal year for life table estimated using adjustments based on Brass: 1914

   Pivotal years for life tables estimated using adjustments based on Bennett-Horiuchi: 1947, 1953, 1965, 1975, 1985, 1996, 2005

2. Age misreporting

   OGIVE: No

   Brass: Yes

   Bennett-Horiuchi: Yes

3. Conciliation Yes

### 9.2.4 Estimates of child mortality: 0, 1-4 and 0-5

**Sources of raw data**

- For pivotal yearly estimates from OGIVE method: Model Life Tables (West and South) 1869, 1882, 1895, 1895, 1904, 1914, 1930, 1947
- Adjusted Estimates from Vital Statistics: 1904, 1914, 1947, 1950-2012
- Census: 1970, 1980, 1991, 2001 (Indirect estimates)
- Surveys: None
- Third party estimates:

  UN/CELADE: 1955, 1960, 1965, 1970, 1975, 1980, 1985, 1990, 1995, 2000, 2005, 2010, 2015

  Others: 1882, 1904, 1914, 1947

**Origins of final child mortality estimate**

Table 9.1 summarizes the origin of the final estimates of child mortality for each pivotal year

Table 9.1: Class of final child mortality estimates by time periods

| Age group | Before 1900 | 1900-1950 | 1950-2012 |
|---|---|---|---|
| 0-1 | Gompertz | Gompertz | Lowess & Splines |
| 1-4 | Gompertz | Gompertz | Lowess & Splines |

### 9.2.5 Computation of complete pivotal life tables

- Single-year of age life tables: 1869, 1882, 1895, 1904, 1914, 1930, 1947, 1953, 1965, 1975, 1985, 1996, 2005
- Abridged life tables: 1869, 1882, 1895, 1904, 1914, 1930, 1947, 1953, 1965, 1975, 1985, 1996, 2005
- Treatment of open age group: based on adjusted (for completeness and age misreporting) $L_{85+} = l_{85}/M_{85+}$

### 9.2.6 Construction of life tables by single calendar year

Methods used to produce single calendar year life tables from pivotal

- Intercensal life tables from OGIVE pivotal: 1870-1946, yearly life tables are computed using linear interpolation between two successive pivotal
- Intercensal life tables from Vital Statistics: 1944-2010, yearly life tables are computed directly from adjusted vital statistics

# 9.3 Bolivia

## 9.3.1 Vital statistics

### Deaths

None

### Births

None

### Vital statistics sources

None

## 9.3.2 Population censuses

- 1900 (ages: 0, 7, 14, 18, 26, 31, 41, 100+)
- 1950, 1976, 1992, 2001, 2012 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+)

### Population censuses sources

- Census from Bolivian Statistical Office: 1900, 1950, 1976, 1992, 2001, 2012
- United Nations Yearbook: 1948-2012

## 9.3.3 Adult pivotal life tables (above age 5)

### Adjustments to generate adult pivotal life tables

1. Completeness

   Pivotal years for life table estimated using generalized OGIVE method: 1900, 1925, 1950, 1963, 1976, 1984, 1992, 1996, 2001, 2006, 2012

   Pivotal year for life table estimated using adjustments based on Brass: None

   Pivotal years for life tables estimated using adjustments based on Bennett-Horiuchi: None

2. Age misreporting

   OGIVE: No

   Brass: NA

   Bennett-Horiuchi: No

3. Conciliation Yes

### 9.3.4 Estimates of child mortality: 0, 1-4 and 0-5

**Sources of raw data**

- For pivotal yearly estimates from OGIVE method:
- Model Life Tables (West and South) 1900, 1925, 1950
- Adjusted Estimates from Vital Statistics: NA
- Census: 1976, 1992, 2001, 2012 (Indirect estimates)
- Surveys:

  1975 (Indirect estimates from National Demographic Survey I)

  1989 (Indirect estimates from National Demographic Survey II)

  1988 (Indirect estimates from Population and Household National Survey)

  2000 (Indirect estimates from Multiple Indicator Cluster Surveys)

  1989, 1994, 1998, 2003, 2008 (Direct and indirect estimates from Demographic and Health Surveys)
- Third party estimates:

  UN/CELADE: 1955, 1960, 1965, 1970, 1975, 1980, 1985, 1990, 1995, 2000, 2005, 2010, 2015

  Others: 1900, 1950

**Origins of final child mortality estimate**

Table 9.2 summarizes the origin of the final estimates of child mortality for each pivotal year

Table 9.2: Class of final child mortality estimates by time periods

| **Age group** | **Before 1900** | **1900-1950** | **1950-2012** |
|---|---|---|---|
| 0-1 None | Gompertz | Lowess & Splines | |
| 1-4 | None | Gompertz | Lowess & Splines |

### 9.3.5 Computation of complete pivotal life tables

- Single-year of age life tables: 1900, 1925, 1950, 1963, 1976, 1984, 1992, 1996, 2001, 2006, 2012
- Abridged life tables: 1900, 1925, 1950, 1963, 1976, 1984, 1992, 1996, 2001, 2006, 2012
- Treatment of open age group: based on adjusted (for completeness and age misreporting) $L_{85+} = l_{85}/M_{85+}$

### 9.3.6 Construction of life tables by single calendar year

Methods used to produce single calendar year life tables from pivotal

- Intercensal life tables from OGIVE pivotal: 1901-2011, yearly life tables are computed using linear interpolation between two successive pivotal
- Intercensal life tables from Vital Statistics: None

## 9.4 Brazil

### 9.4.1 Vital statistics

#### Deaths

- 1980- 1995 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+)
- 1996-2016 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85-89, 90-94, 95+)

#### Causes of Death

- ICD9: 1979-1995
- ICD10: 1996-2016

#### Births

- 1913-2012

#### Vital statistics sources

- U.S. Bureau of Census. Brazil: Summary of Biostatistics
- United Nations Yearbook: 1936-2010
- World Health Organization. Health statistics and information systems: 1979-2016 https://www.who.int/healthinfo/statistics/mortality_rawdata/en/

### 9.4.2 Population censuses

- 1872, 1890, 1900, 1920 (ages: 0, 1-4, 5-9, 10-14, 15-29, 30-39, 40-49, 50-59, 60-69, 70-79, 80+)
- 1940, 1950, 1960, 1970, 1982, 1991, 2000, 2010 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+)

#### Population censuses sources

- Census from Brazilian Statistical Office: 1872, 1890, 1900, 1920, 1940
- United Nations Yearbook: 1948-2012

### 9.4.3 Adult pivotal life tables (above age 5)

**Adjustments to generate adult pivotal life tables**

1. Completeness

   Pivotal years for life table estimated using generalized OGIVE method: 1869, 1882, 1895, 1904, 1914, 1930

   Pivotal year for life table estimated using adjustments based on Brass: None

   Pivotal years for life tables estimated using adjustments based on Bennett-Horiuchi: 1985, 1995, 2005

2. Age misreporting

   OGIVE: No

   Brass: NA

   Bennett-Horiuchi: Yes

3. Conciliation Yes

### 9.4.4 Estimates of child mortality: 0, 1-4 and 0-5

**Sources of raw data**

- For pivotal yearly estimates from OGIVE method: Model Life Tables (West and South) 1872, 1881, 1890, 1900, 1910, 1920, 1930, 1940, 1945
- Adjusted Estimates from Vital Statistics: None
- Census: 1970, 1980, 1991, 2001 (Indirect estimates)
- Surveys:

  1972, 1973, 1976, 1977, 1978, 1984, 2006 (Indirect estimates from National Health Survey)

  1986, 2005, 2007, 2008, 2009 (Indirect estimates from National Household Sample Survey)

  1986, 1996 (Direct and indirect estimates from Demographic and Health Survey)
- Third party estimates:

  UN/CELADE: 1955, 1960, 1965, 1970, 1975, 1980, 1985, 1990, 1995, 2000, 2005, 2010, 2015

  Others: 1940

**Origins of final child mortality estimate**

Table 9.3 summarizes the origin of the final estimates of child mortality for each pivotal year

Table 9.3: Class of final child mortality estimates by time periods

| Age group | Before 1900 | 1900-1950 | 1950-2012 |
|---|---|---|---|
| 0-1 | Gompertz | Gompertz | Lowess & Splines |
| 1-4 | Gompertz | Gompertz | Lowess & Splines |

### 9.4.5 Computation of complete pivotal life tables

- Single-year of age life tables: 1872, 1881, 1895, 1910, 1930, 1945, 1955, 1965, 1975, 1985, 1995, 2005
- Abridged life tables: 1872, 1881, 1895, 1910, 1930, 1945, 1955, 1965, 1975, 1985, 1995, 2005
- Treatment of open age group: based on adjusted (for completeness and age misreporting) $L_{85+} = l_{85}/M_{85+}$

### 9.4.6 Construction of life tables by single calendar year

Methods used to produce single calendar year life tables from pivotal

- Intercensal life tables from OGIVE pivotal: 1873-1984, yearly life tables are computed using linear interpolation between two successive pivotal
- Intercensal life tables from Vital Statistics: 1974-2010, yearly life tables are computed directly from adjusted vital statistics

# 9.5 Chile

## 9.5.1 Vital statistics

### Deaths

- 1936-1995 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+)
- 1996 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85-89, 90-94, 95+)
- 1997-2016 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85-89, 90-94, 95-99, 100+)

### Causes of Death

- ICD7: 1955-1967
- ICD8: 1968-1979
- ICD9: 1980-1996
- ICD10: 1997-2016

### Births

- 1850-2015

### Vital statistics sources

- U.S. Bureau of Census. Chile: Summary of Biostatistics
- United Nations Yearbook: 1936-2010
- World Health Organization. Health statistics and information systems: 1955-2016 https://www.who.int/healthinfo/statistics/mortality_rawdata/en/

## 9.5.2 Population censuses

- 1854, 1865, 1875 (ages: 0-6, 7-14, 15-29, 30-49, 50-79, 80-84, 85+)
- 1885, 1895 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+)
- 1907 (ages: 0-5, 6-9, 10-14, 15-20, 21-24, 25-29, 30-39, 40-49, 50-59, 60-69, 70-79, 80+)
- 1920, 1930, 1940, 1952, 1960, 1970, 1982, 1992, 2002 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+)

**Population censuses sources**

- Census from Chilean Statistical Office: 1854, 1865, 1875, 1885, 1895, 1907, 1920, 1930, 1940

- United Nations Yearbook: 1948-2012

### 9.5.3 Adult pivotal life tables (above age 5)

**Adjustments to generate adult pivotal life tables**

1. Completeness

   Pivotal years for life table estimated using generalized OGIVE method: 1854, 1859, 1865, 1870, 1875, 1880, 1885, 1890, 1895, 1901, 1907, 1913, 1920, 1925, 1930, 1940

   Pivotal year for life table estimated using adjustments based on Brass: None

   Pivotal years for life tables estimated using adjustments based on Bennett-Horiuchi: 1935, 1946, 1956, 1965, 1976, 1987, 1997, 2006

2. Age misreporting

   OGIVE: No

   Brass: NA

   Bennett-Horiuchi: Yes

3. Conciliation Yes

### 9.5.4 Estimates of child mortality: 0, 1-4 and 0-5

**Sources of raw data**

- For pivotal yearly estimates from OGIVE method: Model Life Tables (West and South) 1854, 1859, 1865, 1870, 1875, 1880, 1885, 1890, 1895, 1901, 1907, 1913, 1920, 1925, 1930, 1935, 1940, 1946

- Adjusted Estimates from Vital Statistics: 1920, 1925, 1930, 1935, 1940, 1946, 1955-2012

- Census: 1970, 1982, 1992, 2002 (Indirect estimates)

- Surveys: None

- Third party estimates:

  UN/CELADE: 1955, 1960, 1965, 1970, 1975, 1980, 1985, 1990, 1995, 2000, 2005, 2010, 2015

  Others: 1920, 1930, 1935, 1940, 1946

Table 9.4: Class of final child mortality estimates by time periods

| Age group | Before 1900 | 1900-1950 | 1950-2012 |
|---|---|---|---|
| 0-1 | Gompertz | Gompertz | Lowess & Splines |
| 1-4 | Gompertz | Gompertz | Lowess & Splines |

**Origins of final child mortality estimate**

Table 9.4 summarizes the origin of the final estimates of child mortality for each pivotal year

### 9.5.5 Computation of complete pivotal life tables

- Single-year of age life tables: 1854, 1859, 1865, 1870, 1875, 1880, 1885, 1890, 1895, 1901, 1907, 1913, 1920, 1925, 1930, 1935, 1940, 1946, 1956, 1965, 1976, 1987, 1997, 2006
- Abridged life tables: 1854, 1859, 1865, 1870, 1875, 1880, 1885, 1890, 1895, 1901, 1907, 1913, 1920, 1925, 1930, 1935, 1940, 1946, 1956, 1965, 1976, 1987, 1997, 2006
- Treatment of open age group: based on adjusted (for completeness and age misreporting) $L_{85+} = l_{85}/M_{85+}$

### 9.5.6 Construction of life tables by single calendar year

Methods used to produce single calendar year life tables from pivotal

- Intercensal life tables from OGIVE pivotal: 1855-1945, yearly life tables are computed using linear interpolation between two successive pivotal
- Intercensal life tables from Vital Statistics: 1920-2010, yearly life tables are computed directly from adjusted vital statistics

## 9.6 Colombia

### 9.6.1 Vital statistics

#### Deaths

- 1938-1946 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-49, 50-59, 60-69, 70+)
- 1947-1949 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-49, 50-59, 60-69, 70-79, 80+)
- 1980- 1981 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+)
- 1950-1996 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+)
- 1997-2016 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85-89, 90-94, 95+)

#### Causes of Death

- ICD7: 1953-1967
- ICD8: 1968, 1969, 1978
- ICD9: 1979-1997
- ICD10: 1998-2016

#### Births

- 1915-2012

#### Vital statistics sources

- U.S. Bureau of Census. Colombia: Summary of Biostatistics
- United Nations Yearbook: 1936-2010
- World Health Organization. Health statistics and information systems: 1953-2016 https://www.who.int/healthinfo/statistics/mortality_rawdata/en/

### 9.6.2 Population censuses

- 1912, 1918, 1928, 1938, 1951, 1964, 1973, 1985, 1993, 2005, 2012 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+)

**Population censuses sources**

- Census from Colombia Statistical Office: 1912, 1918, 1928, 1938
- United Nations Yearbook: 1948-2010

### 9.6.3 Adult pivotal life tables (above age 5)

**Adjustments to generate adult pivotal life tables**

1. Completeness

   Pivotal years for life table estimated using generalized OGIVE method: 1905, 1908, 1912, 1915, 1918, 1923, 1928, 1933, 1938

   Pivotal year for life table estimated using adjustments based on Brass: None

   Pivotal years for life tables estimated using adjustments based on Bennett-Horiuchi: 1944, 1957, 1968, 1979, 1989, 1999, 2008

2. Age misreporting

   OGIVE: No

   Brass: NA

   Bennett-Horiuchi: Yes

3. Conciliation Yes

### 9.6.4 Estimates of child mortality: 0, 1-4 and 0-5

**Sources of raw data**

- For pivotal yearly estimates from OGIVE method: Model Life Tables (West and South) 1905, 1908, 1912, 1915, 1918, 1923, 1928, 1933, 1938, 1944
- Adjusted Estimates from Vital Statistics: 1938-1944
- Census: 1973, 1985 (Indirect estimates)
- Surveys:

  1973 (Direct and indirect estimates from World Fertility Survey)

  1978 (Indirect estimates from Contraceptive Prevalence Survey)

  1978, 1980 (Indirect estimates from Household Survey)

  1986, 2005, 2007, 2008, 2009 (Indirect estimates from National Household Sample Survey)

  1986, 1990, 1995, 2000, 2005, 2010 (Direct and indirect estimates from Demographic and Health Survey)

- Third party estimates:

  UN/CELADE: 1955, 1960, 1965, 1970, 1975, 1980, 1985, 1990, 1995, 2000, 2005, 2010, 2015

  Others: 1938

**Origins of final child mortality estimate**

Table 9.5 summarizes the origin of the final estimates of child mortality for each pivotal year

Table 9.5: Class of final child mortality estimates by time periods

| **Age group** | **Before 1900** | **1900-1950** | textbf1950-2012 |
|---|---|---|---|
| 0-1 | None | Gompertz | Lowess & Splines |
| 1-4 | None | Gompertz | Lowess & Splines |

### 9.6.5 Computation of complete pivotal life tables

- Single-year of age life tables: 1905, 1908, 1912, 1915, 1918, 1923, 1928, 1933, 1938, 1944, 1957, 1968, 1979, 1989, 1999, 2008
- Abridged life tables: 1905, 1908, 1912, 1915, 1918, 1923, 1928, 1933, 1938, 1944, 1957, 1968, 1979, 1989, 1999, 2008
- Treatment of open age group: based on adjusted (for completeness and age misreporting) $L_{85+} = l_{85}/M_{85+}$

### 9.6.6 Construction of life tables by single calendar year

Methods used to produce single calendar year life tables from pivotal

- Intercensal life tables from OGIVE pivotal: 1906-1943, yearly life tables are computed using linear interpolation between two successive pivotal
- Intercensal life tables from Vital Statistics: 1936-2010, yearly life tables are computed directly from adjusted vital statistics

# 9.7 Costa Rica

## 9.7.1 Vital statistics

### Deaths

- 1937-1949 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-49, 50-59, 60-69, 70-79, 80+)
- 1950- 1969 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+)
- 1970-2014 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85-89, 90-94, 95-99, 100+)

### Causes of Death

- ICD7: 1961-1967
- ICD8: 1968-1979
- ICD9: 1980-1996
- ICD10: 1997-2014

### Births

- 1900-2013

### Vital statistics sources

- Anuario Estadístico Costa Rica, 1937-1949
- United Nations Yearbook: 1950-2015
- World Health Organization. Health statistics and information systems: 1961-2014 https://www.who.int/healthinfo/statistics/mortality_rawdata/en/

## 9.7.2 Population censuses

- 1864, 1883, 1892, 1927, 1950, 1963, 1973, 1984, 2000, 2011 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+)

### Population censuses sources

- Census from Costa Rican Statistical Office: 1864, 1883, 1892, 1927
- United Nations Yearbook: 1948-2015

### 9.7.3 Adult pivotal life tables (above age 5)

**Adjustments to generate adult pivotal life tables**

1. Completeness

   Pivotal years for life table estimated using generalized OGIVE method: 1864, 1873, 1883, 1887, 1892, 1909, 1927

   Pivotal year for life table estimated using adjustments based on Brass: 1927

   Pivotal years for life tables estimated using adjustments based on Bennett-Horiuchi: 1938, 1956, 1968, 1978, 1992, 2005

2. Age misreporting

   OGIVE: No

   Brass: Yes

   Bennett-Horiuchi: Yes

3. Conciliation Yes

### 9.7.4 Estimates of child mortality: 0, 1-4 and 0-5

**Sources of raw data**

- For pivotal yearly estimates from OGIVE method: Model Life Tables (West and South) 1864, 1873, 1883, 1887, 1892, 1909, 1927, 1938

- Adjusted Estimates from Vital Statistics: 1900, 1910, 1920, 1927, 1930, 1938, 1956-2010

- Census: 1973, 1984, 2000, 2011 (Indirect estimates)

- Surveys:

  1976 (Direct and indirect estimates from World Fertility Survey)

  1978 (Indirect estimates from Contraceptive Prevalence Survey)

  1981 (Indirect estimates from Contraceptive Prevalence Survey)

  1986 (Indirect estimates from Fertility and Health National Survey)

- Third party estimates:

  UN/CELADE: 1955, 1960, 1965, 1970, 1975, 1980, 1985, 1990, 1995, 2000, 2005, 2010, 2015

  Others: 1900, 1910, 1927, 1938

Table 9.6: Class of final child mortality estimates by time periods

| Age group | Before 1900 | 1900-1950 | 1950-2012 |
|---|---|---|---|
| 0-1 | Gompertz | Gompertz | Lowess & Splines |
| 1-4 | Gompertz | Gompertz | Lowess & Splines |

**Origins of final child mortality estimate**

Table 9.6 summarizes the origin of the final estimates of child mortality for each pivotal year

### 9.7.5 Computation of complete pivotal life tables

- Single-year of age life tables: 1864, 1873, 1883, 1887, 1892, 1909, 1927, 1938, 1956, 1968, 1978, 1992, 2005
- Abridged life tables: 1864, 1873, 1883, 1887, 1892, 1909, 1927, 1938, 1956, 1968, 1978, 1992, 2005
- Treatment of open age group: based on adjusted (for completeness and age misreporting) $L_{85+} = l_{85}/M_{85+}$

### 9.7.6 Construction of life tables by single calendar year

Methods used to produce single calendar year life tables from pivotal

- Intercensal life tables from OGIVE pivotal: 1865-1937, yearly life tables are computed using linear interpolation between two successive pivotal
- Intercensal life tables from Vital Statistics: 1927-2010, yearly life tables are computed directly from adjusted vital statistics

# 9.8 Cuba

## 9.8.1 Vital statistics

### Deaths

- 1927-1936 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-39, 40-59, 60+)
- 1937- 2000 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+)
- 2000-2016 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85-89, 90-94, 95+)

### Causes of Death

- ICD7: 1964
- ICD8: 1968-1978
- ICD9: 1979-2000
- ICD10: 2001-2016

### Births

- 1890-2013

### Vital statistics sources

- U.S. Bureau of Census. Cuba: Summary of Biostatistics: 1927-1936
- United Nations Yearbook: 1936-2015
- World Health Organization. Health statistics and information systems: 1964-2016 https://www.who.int/healthinfo/statistics/mortality_rawdata/en/

## 9.8.2 Population censuses

- 1841 (ages: 0-15, 16-60, 61+)
- 1861 (ages: 0, 1-7, 8-15, 16-20, 21-25, 26-30, 31-40, 41-50, 51-60, 61-70, 71-80, 81-85, 86+)
- 1877, 1887, 1899, 1907, 1919, 1931, 1943, 1953, 1970, 1981, 2002, 2012 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+)

**Population censuses sources**

- Census from Cuban Statistical Office: 1841, 1861, 1877, 1887, 1899, 1907, 1919, 1931, 1943
- United Nations Yearbook: 1948-2015

### 9.8.3 Adult pivotal life tables (above age 5)

**Adjustments to generate adult pivotal life tables**

1. Completeness

   Pivotal years for life table estimated using generalized OGIVE method: 1841, 1851, 1861, 1869, 1877, 1882, 1887, 1893, 1899, 1903, 1907, 1913, 1919

   Pivotal year for life table estimated using adjustments based on Brass: 1925

   Pivotal years for life tables estimated using adjustments based on Bennett-Horiuchi: 1937, 1948, 1961, 1975, 1991, 2006

2. Age misreporting

   OGIVE: No

   Brass: Yes

   Bennett-Horiuchi: Yes

3. Conciliation Yes

### 9.8.4 Estimates of child mortality: 0, 1-4 and 0-5

**Sources of raw data**

- For pivotal yearly estimates from OGIVE method: Model Life Tables (West and South) 1841, 1861, 1869, 1877, 1882, 1887, 1893, 1899, 1903, 1907, 1913, 1919, 1925, 1931, 1937, 1943, 1948
- Adjusted Estimates from Vital Statistics: 1919, 1931, 1943, 1953, 1955-2012
- Census: 1981 (Indirect estimates)
- Surveys:

  1974 (Indirect estimates from National Population Survey on Income and Expenditures)

  1979 (Indirect estimates from National Demographic Survey)

  1987 (Indirect estimates from National Fertility Survey)

- Third party estimates:

  UN/CELADE: 1955, 1960, 1965, 1970, 1975, 1980, 1985, 1990, 1995, 2000, 2005, 2010, 2015

  Others: 1925, 1937, 1948

**Origins of final child mortality estimate**

Table 9.7 summarizes the origin of the final estimates of child mortality for each pivotal year

Table 9.7: Class of final child mortality estimates by time periods

| Age group | Before 1900 | 1900-1950 | 1950-2012 |
|---|---|---|---|
| 0-1 | Gompertz | Gompertz | Lowess & Splines |
| 1-4 | Gompertz | Gompertz | Lowess & Splines |

### 9.8.5 Computation of complete pivotal life tables

- Single-year of age life tables: 1841, 1851, 1861, 1869, 1877, 1882, 1887, 1893, 1899, 1903, 1907, 1913, 1919, 1925, 1937, 1948, 1961, 1975, 1991, 2006
- Abridged life tables: 1841, 1851, 1861, 1869, 1877, 1882, 1887, 1893, 1899, 1903, 1907, 1913, 1919, 1925, 1937, 1948, 1961, 1975, 1991, 2006
- Treatment of open age group: based on adjusted (for completeness and age misreporting) $L_{85+} = l_{85}/M_{85+}$

### 9.8.6 Construction of life tables by single calendar year

Methods used to produce single calendar year life tables from pivotal

- Intercensal life tables from OGIVE pivotal: 1842-1947, yearly life tables are computed using linear interpolation between two successive pivotal
- Intercensal life tables from Vital Statistics: 1920-2010, yearly life tables are computed directly from adjusted vital statistics

# 9.9 Dominican Republic

## 9.9.1 Vital statistics

### Deaths

- 1935-1939, 1948-1949 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75+)
- 1940-1947, 1950-1995 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+)
- 1996-2013 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85-89, 90-94, 95+)

### Causes of Death

- ICD7: 1965-1967
- ICD8: 1968-1979
- ICD9: 1980-1992, 1994, 1995
- ICD10: 1996-2013

### Births

- 1906-2012

### Vital statistics sources

- U.S. Bureau of Census. Dominican Republic: Summary of Biostatistics
- United Nations Yearbook: 1936-2010
- World Health Organization. Health statistics and information systems: 1965-2013 https://www.who.int/healthinfo/statistics/mortality_rawdata/en/

## 9.9.2 Population censuses

- 1920 (ages: 0-1, 2-6, 7-14, 15-20, 21-60, 61+)
- 1935, 1950, 1960, 1970, 1981, 1993, 2002, 2010 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+)

### Population censuses sources

- Census from Dominican Republic Statistical Office: 1920, 1935
- United Nations Yearbook: 1948-2010

### 9.9.3 Adult pivotal life tables (above age 5)

**Adjustments to generate adult pivotal life tables**

1. Completeness

   Pivotal years for life table estimated using generalized OGIVE method: 1920, 1927, 1935, 1942

   Pivotal year for life table estimated using adjustments based on Brass: None

   Pivotal years for life tables estimated using adjustments based on Bennett-Horiuchi: 1955, 1965, 1975, 1987, 1997, 2006

2. Age misreporting

   OGIVE: No

   Brass: NA

   Bennett-Horiuchi: Yes

3. Conciliation Yes

### 9.9.4 Estimates of child mortality: 0, 1-4 and 0-5

**Sources of raw data**

- For pivotal yearly estimates from OGIVE method: Model Life Tables (West and South) 1920, 1927, 1935, 1942

- Adjusted Estimates from Vital Statistics: 1935-1942

- Census: 1970, 1981, 2002, 2010 (Indirect estimates)

- Surveys:

  1975 (Direct and indirect estimates from World Fertility Survey)

  1983 (Indirect estimates from Contraceptive Prevalence Survey)

  1978, 1980 (Indirect estimates from Household Survey)

  1986, 1991, 1996, 1999, 2002, 2007, 2013 (Direct and indirect estimates from Demographic and Health Survey)

  2006, 2014 (Direct estimates from Multiple Indicator Cluster Surveys)

- Third party estimates:

  UN/CELADE: 1955, 1960, 1965, 1970, 1975, 1980, 1985, 1990, 1995, 2000, 2005, 2010, 2015

  Others: None

Table 9.8: Class of final child mortality estimates by time periods

| Age group | Before 1900 | 1900-1950 | 1950-2012 |
|---|---|---|---|
| 0-1 | None | Gompertz | Lowess & Splines |
| 1-4 | None | Gompertz | Lowess & Splines |

**Origins of final child mortality estimate**

Table 9.8 summarizes the origin of the final estimates of child mortality for each pivotal year

### 9.9.5 Computation of complete pivotal life tables

- Single-year of age life tables: 1920, 1927, 1935, 1942, 1955, 1965, 1975, 1987, 1997, 2006
- Abridged life tables: 1920, 1927, 1935, 1942, 1955, 1965, 1975, 1987, 1997, 2006
- Treatment of open age group: based on adjusted (for completeness and age misreporting) $L_{85+} = l_{85}/M_{85+}$

### 9.9.6 Construction of life tables by single calendar year

Methods used to produce single calendar year life tables from pivotal

- Intercensal life tables from OGIVE pivotal: 1921-1941, yearly life tables are computed using linear interpolation between two successive pivotal
- Intercensal life tables from Vital Statistics: 1935-2010, yearly life tables are computed directly from adjusted vital statistics

## 9.10 Ecuador

### 9.10.1 Vital statistics

#### Deaths

- 1950-1966 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-34, 35-44, 45-54, 55-64, 65+)
- 1967-1996 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+)
- 1997-2016 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85-89, 90-94, 95+)

#### Causes of Death

- ICD7: 1961, 1963-1967
- ICD8: 1968-1978
- ICD9: 1979-1996
- ICD10: 1997-2016

#### Births

- 1911-2012

#### Vital statistics sources

- U.S. Bureau of Census. Ecuador: Summary of Biostatistics
- United Nations Yearbook: 1936-2010
- World Health Organization. Health statistics and information systems: 1961-2016 https://www.who.int/healthinfo/statistics/mortality_rawdata/en/

### 9.10.2 Population censuses

- 1950, 1962, 1974, 1982, 1990, 2001, 2010 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+)

#### Population censuses sources

- United Nations Yearbook: 1948-2010

### 9.10.3 Adult pivotal life tables (above age 5)

#### Adjustments to generate adult pivotal life tables

1. Completeness

   Pivotal years for life table estimated using generalized OGIVE method: None

   Pivotal year for life table estimated using adjustments based on Brass: None

   Pivotal years for life tables estimated using adjustments based on Bennett-Horiuchi: 1956, 1968, 1978, 1986, 1995, 2005

2. Age misreporting

   OGIVE: No

   Brass: NA

   Bennett-Horiuchi: Yes

3. Conciliation Yes

### 9.10.4 Estimates of child mortality: 0, 1-4 and 0-5

#### Sources of raw data

- For pivotal yearly estimates from OGIVE method: Model Life Tables (West and South) None

- Adjusted Estimates from Vital Statistics: None

- Census: 1974, 1982, 1990, 2001, 2010 (Indirect estimates)

- Surveys:

  1979 (Direct and indirect estimates from World Fertility Survey)

  1987 (Direct and indirect estimates from Demographic and Health Survey)

  1989, 1994, 1999 (Direct estimates from Demographic, Maternal and Child Health)

- Third party estimates:

  UN/CELADE: 1955, 1960, 1965, 1970, 1975, 1980, 1985, 1990, 1995, 2000, 2005, 2010, 2015

  Others: None

#### Origins of final child mortality estimate

Table 9.9 summarizes the origin of the final estimates of child mortality for each pivotal year

Table 9.9: Class of final child mortality estimates by time periods

| Age group | Before 1900 | 1900-1950 | 1950-2012 |
|---|---|---|---|
| 0-1 | None | None | Lowess & Splines |
| 1-4 | None | None | Lowess & Splines |

### 9.10.5 Computation of complete pivotal life tables

- Single-year of age life tables: 1956, 1968, 1978, 1986, 1995, 2005
- Abridged life tables: 1956, 1968, 1978, 1986, 1995, 2005
- Treatment of open age group: based on adjusted (for completeness and age misreporting) $L_{85+} = l_{85}/M_{85+}$

### 9.10.6 Construction of life tables by single calendar year

Methods used to produce single calendar year life tables from pivotal

- Intercensal life tables from OGIVE pivotal: None
- Intercensal life tables from Vital Statistics: 1950-2009, yearly life tables are computed directly from adjusted vital statistics

## 9.11 El Salvador

### 9.11.1 Vital statistics

**Deaths**

- 1933-1935 (ages: 0, 1-4, 5-9, 10-14, 15-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75-84, 85+)
- 1936-1996 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+)
- 1997-2014 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85-89, 90-94, 95+)

**Causes of Death**

- ICD7: 1950-1967
- ICD8: 1968-1974
- ICD9: 1981-1984, 1990-1993, 1995-1996
- ICD10: 1997-2014

**Births**

- 1900-2012

**Vital statistics sources**

- U.S. Bureau of Census. El Salvador: Summary of Biostatistics
- United Nations Yearbook: 1933-2010
- World Health Organization. Health statistics and information systems: 1950-2014 https://www.who.int/healthinfo/statistics/mortality_rawdata/en/

### 9.11.2 Population censuses

- 1930 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80+)
- 1950, 1961, 1971, 1992, 2007 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+)

**Population censuses sources**

- Census from El Salvador Statistical Office: 1930
- United Nations Yearbook: 1948-2010
- World Health Organization. Health statistics and information systems: 1966-2016 https://www.who.int/healthinfo/statistics/mortality_rawdata/en/

### 9.11.3 Adult pivotal life tables (above age 5)

#### Adjustments to generate adult pivotal life tables

1. Completeness

   Pivotal years for life table estimated using generalized OGIVE method: 1930, 1940

   Pivotal year for life table estimated using adjustments based on Brass: None

   Pivotal years for life tables estimated using adjustments based on Bennett-Horiuchi: 1955, 1966, 1981, 1999, 2008

2. Age misreporting

   OGIVE: No

   Brass: NA

   Bennett-Horiuchi: Yes

3. Conciliation Yes

### 9.11.4 Estimates of child mortality: 0, 1-4 and 0-5

#### Sources of raw data

- For pivotal yearly estimates from OGIVE method: Model Life Tables (West and South) 1930, 1940

- Adjusted Estimates from Vital Statistics: 1930-1940

- Census: 1971, 1992, 2007 (Indirect estimates)

- Surveys:

  1973 (Direct estimates from National Fertility Survey)

  1985 (Direct and indirect estimates from Demographic and Health Survey)

  1992 (Indirect estimates from Household and Health Survey)

  1992 (Indirect estimates from Multiple Indicator Cluster Survey)

  1998 (Direct and indirect estimates from National Health Survey)

  2002, 2008 (Direct estimates from National Family Health Survey)

- Third party estimates:

  UN/CELADE: 1955, 1960, 1965, 1970, 1975, 1980, 1985, 1990, 1995, 2000, 2005, 2010, 2015

  Others: 1930

Table 9.10: Class of final child mortality estimates by time periods

| Age group | Before 1900 | 1900-1950 | 1950-2012 |
|---|---|---|---|
| 0-1 | None | Gompertz | Lowess & Splines |
| 1-4 | None | Gompertz | Lowess & Splines |

**Origins of final child mortality estimate**

Table 9.10 summarizes the origin of the final estimates of child mortality for each pivotal year

### 9.11.5 Computation of complete pivotal life tables

- Single-year of age life tables: 1930, 1940, 1955, 1966, 1981, 1999, 2008
- Abridged life tables: 1930, 1940, 1955, 1966, 1981, 1999, 2008
- Treatment of open age group: based on adjusted (for completeness and age misreporting) $L_{85+} = l_{85}/M_{85+}$

### 9.11.6 Construction of life tables by single calendar year

Methods used to produce single calendar year life tables from pivotal

- Intercensal life tables from OGIVE pivotal: 1931-1939, yearly life tables are computed using linear interpolation between two successive pivotal
- Intercensal life tables from Vital Statistics: 1929-2009, yearly life tables are computed directly from adjusted vital statistics

## 9.12 Guatemala

### 9.12.1 Vital statistics

**Deaths**

- 1935-1939, 1948-1949 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75+)
- 1940-1947, 1950-1995 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+)
- 1996-2016 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85-89, 90-94, 95+)

**Causes of Death**

- ICD7: 1963-1968
- ICD8: 1969-1971, 1974-1978
- ICD9: 1979-1981, 1984, 1986-2004
- ICD10: 2005-2016

**Births**

- 1900-2012

**Vital statistics sources**

- U.S. Bureau of Census. Guatemala: Summary of Biostatistics
- United Nations Yearbook: 1935-2010
- World Health Organization. Health statistics and information systems: 1963-2016 https://www.who.int/healthinfo/statistics/mortality_rawdata/en/

### 9.12.2 Population censuses

- 1921 (ages: 0, 1-4, 5-6, 7-13, 14-17, 18-20, 21-29, 30-39, 40-49, 50-59, 60-69, 70-79, 80+)
- 1940, 1950, 1964, 1973, 1981, 1994, 2002, 2011 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+)

**Population censuses sources**

- Census from Guatemalan Statistical Office: 1921, 1940
- United Nations Yearbook: 1948-2010

### 9.12.3 Adult pivotal life tables (above age 5)

**Adjustments to generate adult pivotal life tables**

1. Completeness

   Pivotal years for life table estimated using generalized OGIVE method: 1880, 1886, 1893, 1907, 1921, 1930, 1940, 1945

   Pivotal year for life table estimated using adjustments based on Brass: None

   Pivotal years for life tables estimated using adjustments based on Bennett-Horiuchi: 1957, 1968, 1977, 1987, 1998, 2005

2. Age misreporting

   OGIVE: No

   Brass: NA

   Bennett-Horiuchi: Yes

3. Conciliation Yes

### 9.12.4 Estimates of child mortality: 0, 1-4 and 0-5

**Sources of raw data**

- For pivotal yearly estimates from OGIVE method: Model Life Tables (West and South) 1880, 1886, 1893, 1907, 1921, 1930, 1940, 1945

- Adjusted Estimates from Vital Statistics: 1940-1949

- Census: 1973, 1981, 2002 (Indirect estimates)

- Surveys:

  1978 (Indirect estimates from National Fertility Survey)

  1987, 1989 (Indirect estimates from National Sociodemographic Survey)

  1987, 1995, 1999 (Direct and indirect estimates from Demographic and Health Survey)

  2002 (Direct and indirect estimates from Reproductive Health Survey)

  2008 (Direct estimates from National Maternal and Child Health Survey)

- Third party estimates:

  UN/CELADE: 1955, 1960, 1965, 1970, 1975, 1980, 1985, 1990, 1995, 2000, 2005, 2010, 2015

  Others: 1940, 1945

Table 9.11: Class of final child mortality estimates by time periods

| Age group | Before 1900 | 1900-1950 | 1950-2012 |
|---|---|---|---|
| 0-1 | Gompertz | Gompertz | Lowess & Splines |
| 1-4 | Gompertz | Gompertz | Lowess & Splines |

**Origins of final child mortality estimate**

Table 9.11 summarizes the origin of the final estimates of child mortality for each pivotal year

### 9.12.5 Computation of complete pivotal life tables

- Single-year of age life tables: 1880, 1886, 1893, 1907, 1921, 1930, 1940, 1945, 1957, 1968, 1977, 1987, 1998, 2005
- Abridged life tables: 1880, 1886, 1893, 1907, 1921, 1930, 1940, 1945, 1957, 1968, 1977, 1987, 1998, 2005
- Treatment of open age group: based on adjusted (for completeness and age misreporting) $L_{85+} = l_{85}/M_{85+}$

### 9.12.6 Construction of life tables by single calendar year

Methods used to produce single calendar year life tables from pivotal

- Intercensal life tables from OGIVE pivotal: 1881-1944, yearly life tables are computed using linear interpolation between two successive pivotal
- Intercensal life tables from Vital Statistics: 1939-2009, yearly life tables are computed directly from adjusted vital statistics

# 9.13 Honduras

## 9.13.1 Vital statistics

### Deaths

- 1933-1947 (ages: 0, 1-5, 6-9, 10-14, 15-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79, 80+)

- 1948-1950 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70-74, 75-79, 80+)

- 1951-1963 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+)

- 1964-1965 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70+)

- 1966-1990 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+)

### Births

- 1910-2012

### Vital statistics sources

- U.S. Bureau of Census. Honduras: Summary of Biostatistics

- United Nations Yearbook: 1936-2010

## 9.13.2 Population censuses

- 1930, 1935, 1940, 1945, 1950, 1961, 1974, 1988, 2001, 2013 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+)

### Population censuses sources

- Census from Honduras Statistical Office: 1930, 1935, 1940, 1945

- United Nations Yearbook: 1948-2015

### 9.13.3 Adult pivotal life tables (above age 5)

**Adjustments to generate adult pivotal life tables**

1. Completeness

   Pivotal years for life table estimated using generalized OGIVE method: 1930, 1932, 1935, 1937, 1940, 1942, 1945, 1947, 1955, 1961, 1967, 1974, 1981, 1988, 1994, 2001, 2007, 2013

   Pivotal year for life table estimated using adjustments based on Brass: None

   Pivotal years for life tables estimated using adjustments based on Bennett-Horiuchi: None

2. Age misreporting

   OGIVE: No

   Brass: NA

   Bennett-Horiuchi: No

3. Conciliation Yes

### 9.13.4 Estimates of child mortality: 0, 1-4 and 0-5

**Sources of raw data**

- For pivotal yearly estimates from OGIVE method: Model Life Tables (West and South) 1930, 1932, 1935, 1937, 1940, 1942, 1945, 1947

- Adjusted Estimates from Vital Statistics: 1940-1945

- Census: 1974, 1988, 2001 (Indirect estimates)

- Surveys:

  1970, 1972, 1983 (Direct and indirect estimates from National Demographic Survey)

  1984 (Indirect estimates from National Survey of Maternal and Child Health)

  1987, 1991, 1996, 2001 (Direct and indirect estimates from National Survey of Epidemiology and Family Health)

  2005, 2011 (Direct and indirect estimates from Demographic and Health Survey)

- Third party estimates:

  UN/CELADE: 1955, 1960, 1965, 1970, 1975, 1980, 1985, 1990, 1995, 2000, 2005, 2010, 2015

  Others: 1940

**Origins of final child mortality estimate**

Table 9.12 summarizes the origin of the final estimates of child mortality for each pivotal year

Table 9.12: Class of final child mortality estimates by time periods

| Age group | Before 1900 | 1900-1950 | 1950-2012 |
|---|---|---|---|
| 0-1 | None | Gompertz | Lowess & Splines |
| 1-4 | None | Gompertz | Lowess & Splines |

### 9.13.5 Computation of complete pivotal life tables

- Single-year of age life tables: 1930, 1932, 1935, 1937, 1940, 1942, 1945, 1947, 1955, 1961, 1967, 1974, 1981, 1988, 1994, 2001, 2007, 2013
- Abridged life tables: 1930, 1932, 1935, 1937, 1940, 1942, 1945, 1947, 1955, 1961, 1967, 1974, 1981, 1988, 1994, 2001, 2007, 2013
- Treatment of open age group: based on adjusted (for completeness and age misreporting) $L_{85+} = l_{85}/M_{85+}$

### 9.13.6 Construction of life tables by single calendar year

Methods used to produce single calendar year life tables from pivotal

- Intercensal life tables from OGIVE pivotal: 1931-2012, yearly life tables are computed using linear interpolation between two successive pivotal
- Intercensal life tables from Vital Statistics: 1939-1990, yearly life tables are computed directly from adjusted vital statistics

## 9.14 Mexico

### 9.14.1 Vital statistics

#### Deaths

- 1936-1995 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+)
- 1996 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85-89, 90-94, 95+)
- 1997-2016 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85-89, 90-94, 95-99, 100+)

#### Causes of Death

- ICD7: 1955-1967
- ICD8: 1968-1978
- ICD9: 1979-1997
- ICD10: 1998-2016

#### Births

- 1921-2012

#### Vital statistics sources

- United Nations Yearbook: 1936-2015
- World Health Organization. Health statistics and information systems: 1955-2016 https://www.who.int/healthinfo/statistics/mortality_rawdata/en/

### 9.14.2 Population censuses

- 1854, 1865, 1875 (ages: 0-6, 7-14, 15-29, 30-49, 50-79, 80-84, 85+)
- 1885, 1895 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80+)
- 1907 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80+)
- 1920, 1930, 1940, 1952, 1960, 1970, 1982, 1992, 2002 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+)

**Population censuses sources**

- Census from Mexican Statistical Office: 1854, 1865, 1875, 1885, 1895, 1907, 1920, 1930, 1940
- United Nations Yearbook: 1948-2010

### 9.14.3 Adult pivotal life tables (above age 5)

**Adjustments to generate adult pivotal life tables**

1. Completeness

   Pivotal years for life table estimated using generalized OGIVE method: 1895, 1897, 1900, 1905, 1910, 1915, 1921, 1925, 1930, 1935, 1940

   Pivotal year for life table estimated using adjustments based on Brass: None

   Pivotal years for life tables estimated using adjustments based on Bennett-Horiuchi: 1945, 1955, 1965, 1975, 1985, 1995, 2005

2. Age misreporting

   OGIVE: No

   Brass: NA

   Bennett-Horiuchi: Yes

3. Conciliation Yes

### 9.14.4 Estimates of child mortality: 0, 1-4 and 0-5

**Sources of raw data**

- For pivotal yearly estimates from OGIVE method: Model Life Tables (West and South) 1895, 1897, 1900, 1905, 1910, 1915, 1921, 1925, 1930, 1935, 1940, 1945
- Adjusted Estimates from Vital Statistics: 1900, 1910, 1921, 1925, 1935, 1940, 1945
- Census: 1980, 1990, 2000, 2005, 2010 (Indirect estimates)
- Surveys:

  1976 (Direct and indirect estimates from World Fertility Survey)

  1979 (Indirect estimates from Contraceptive Prevalence Survey)

  1987 (Direct and indirect estimates Demographic and Health Survey)

  2006 (Indirect estimates from National Survey of Demographic Dynamics)

  2009 (Indirect estimates from National Survey of Demographic Dynamics)

- Third party estimates:

  UN/CELADE: 1955, 1960, 1965, 1970, 1975, 1980, 1985, 1990, 1995, 2000, 2005, 2010, 2015

  Others: 1940, 1945

**Origins of final child mortality estimate**

Table 9.13 summarizes the origin of the final estimates of child mortality for each pivotal year

Table 9.13: Class of final child mortality estimates by time periods

| Age group | Before 1900 | 1900-1950 | 1950-2012 |
|---|---|---|---|
| 0-1 | Gompertz | Gompertz | Lowess & Splines |
| 1-4 | Gompertz | Gompertz | Lowess & Splines |

## 9.14.5 Computation of complete pivotal life tables

- Single-year of age life tables: 1895, 1897, 1900, 1905, 1910, 1915, 1921, 1925, 1930, 1935, 1940, 1945, 1955, 1965, 1975, 1985, 1995, 2005
- Abridged life tables: 1895, 1897, 1900, 1905, 1910, 1915, 1921, 1925, 1930, 1935, 1940, 1945, 1955, 1965, 1975, 1985, 1995, 2005
- Treatment of open age group: based on adjusted (for completeness and age misreporting) $L_{85+} = l_{85}/M_{85+}$

## 9.14.6 Construction of life tables by single calendar year

Methods used to produce single calendar year life tables from pivotal

- Intercensal life tables from OGIVE pivotal: 1896-1944, yearly life tables are computed using linear interpolation between two successive pivotal
- Intercensal life tables from Vital Statistics: 1921-2010, yearly life tables are computed directly from adjusted vital statistics

# 9.15 Nicaragua

## 9.15.1 Vital statistics

### Deaths

- 1933-1935 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-39, 40-49, 50-59, 60-69, 70-79, 80+)
- 1936 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+)
- 1937-1938 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80+)
- 1939-1996, 1950-1995 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+)
- 1997-2017 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85-89, 90-94, 95+)

### Causes of Death

- ICD7: 1961-1965
- ICD8: 1968, 1969, 1973-1978
- ICD9: 1988-1994, 1996
- ICD10: 1997-2017

### Births

- 1933-2012

### Vital statistics sources

- U.S. Bureau of Census. Nicaragua: Summary of Biostatistics
- United Nations Yearbook: 1935-2010
- World Health Organization. Health statistics and information systems: 1961-2017 https://www.who.int/healthinfo/statistics/mortality_rawdata/en/

## 9.15.2 Population censuses

- 1940, 1950, 1963, 1971, 1995, 2005 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+)

**Population censuses sources**

- Census from Nicaraguan Statistical Office: 1940
- United Nations Yearbook: 1948-2015

### 9.15.3 Adult pivotal life tables (above age 5)

**Adjustments to generate adult pivotal life tables**

1. Completeness

   Pivotal years for life table estimated using generalized OGIVE method: 1940, 1945

   Pivotal year for life table estimated using adjustments based on Brass: none

   Pivotal years for life tables estimated using adjustments based on Bennett-Horiuchi: 1956, 1967, 1983, 2000, 2007

2. Age misreporting

   OGIVE: No

   Brass: NA

   Bennett-Horiuchi: Yes

3. Conciliation Yes

### 9.15.4 Estimates of child mortality: 0, 1-4 and 0-5

**Sources of raw data**

- For pivotal yearly estimates from OGIVE method: Model Life Tables (West and South) 1940, 1945
- Adjusted Estimates from Vital Statistics: 1940-1945
- Census: 1971, 1995, 2005 (Indirect estimates)
- Surveys:

  1978 (Indirect estimates from National Fertility Survey)

  1985 (Indirect estimates from National Sociodemographic Survey)

  1992 (Direct estimates from Reproductive Health Survey)

  1992 (Direct and indirect estimates from National Fertility Survey)

  1998, 2006 (Direct and indirect estimates from Demographic and Health Survey)

- Third party estimates:

  UN/CELADE: 1955, 1960, 1965, 1970, 1975, 1980, 1985, 1990, 1995, 2000, 2005, 2010, 2015

  Others: None

**Origins of final child mortality estimate**

Table 9.14 summarizes the origin of the final estimates of child mortality for each pivotal year

Table 9.14: Class of final child mortality estimates by time periods

| Age group | Before 1900 | 1900-1950 | 1950-2012 |
|---|---|---|---|
| 0-1 | None | Gompertz | Lowess & Splines |
| 1-4 | None | Gompertz | Lowess & Splines |

### 9.15.5 Computation of complete pivotal life tables

- Single-year of age life tables: 1940, 1945, 1956, 1967, 1983, 2000, 2007
- Abridged life tables: 1940, 1945, 1956, 1967, 1983, 2000, 2007
- Treatment of open age group: based on adjusted (for completeness and age misreporting) $L_{85+} = l_{85}/M_{85+}$

### 9.15.6 Construction of life tables by single calendar year

Methods used to produce single calendar year life tables from pivotal

- Intercensal life tables from OGIVE pivotal: 1941-1944, yearly life tables are computed using linear interpolation between two successive pivotal
- Intercensal life tables from Vital Statistics: 1936-2010, yearly life tables are computed directly from adjusted vital statistics

## 9.16 Panama

### 9.16.1 Vital statistics

#### Deaths

- 1941-1947 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-49, 50-59, 60-69, 70-79, 80+)
- 1948-2001 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+)
- 2002-2016 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85-89, 90+)

#### Causes of Death

- ICD7: 1955-1967
- ICD8: 1968-1978
- ICD9: 1979-1989, 1996, 1997
- ICD10: 1998-2016

#### Births

- 1908-2012

#### Vital statistics sources

- U.S. Bureau of Census. Panama: Summary of Biostatistics
- United Nations Yearbook: 1941-2010
- World Health Organization. Health statistics and information systems: 1955-2016 https://www.who.int/healthinfo/statistics/mortality_rawdata/en/

### 9.16.2 Population censuses

- 1911, 1920, 1930 (ages: 0, 1-6, 7-15, 16-20, 21-30, 31-40, 41-50, 51-60, 61-70, 71-80, 81+)
- 1940 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80+)
- 1950, 1960, 1970, 1980, 1990, 2000, 2010 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+)

**Population censuses sources**

- Census from Panama Statistical Office: 1911, 1920, 1930. 1940
- United Nations Yearbook: 1948-2015

### 9.16.3 Adult pivotal life tables (above age 5)

**Adjustments to generate adult pivotal life tables**

1. Completeness

   Pivotal years for life table estimated using generalized OGIVE method: 1911, 1915, 1920, 1925, 1930, 1935, 1940, 1945

   Pivotal year for life table estimated using adjustments based on Brass: None

   Pivotal years for life tables estimated using adjustments based on Bennett-Horiuchi: 1955, 1965, 1975, 1985, 1995, 2005

2. Age misreporting

   OGIVE: No

   Brass: NO

   Bennett-Horiuchi: Yes

3. Conciliation Yes

### 9.16.4 Estimates of child mortality: 0, 1-4 and 0-5

**Sources of raw data**

- For pivotal yearly estimates from OGIVE method: Model Life Tables (West and South) 1911, 1915, 1920, 1925, 1930, 1935, 1940, 1945
- Adjusted Estimates from Vital Statistics: 1940-1945, 1955-2012
- Census: 1980, 1990, 2000, 2010 (Indirect estimates)
- Surveys:

  1976 (Direct and indirect estimates from World Fertility Survey)
- Third party estimates:

  UN/CELADE: 1955, 1960, 1965, 1970, 1975, 1980, 1985, 1990, 1995, 2000, 2005, 2010, 2015

  Others: None

Table 9.15: Class of final child mortality estimates by time periods

| Age group | Before 1900 | 1900-1950 | 1950-2012 |
|---|---|---|---|
| 0-1 | None | Gompertz | Lowess & Splines |
| 1-4 | None | Gompertz | Lowess & Splines |

#### Origins of final child mortality estimate

Table 9.15 summarizes the origin of the final estimates of child mortality for each pivotal year

### 9.16.5 Computation of complete pivotal life tables

- Single-year of age life tables: 1911, 1915, 1920, 1925, 1930, 1935, 1940, 1945, 1955, 1965, 1975, 1985, 1995, 2005
- Abridged life tables: 1911, 1915, 1920, 1925, 1930, 1935, 1940, 1945, 1955, 1965, 1975, 1985, 1995, 2005
- Treatment of open age group: based on adjusted (for completeness and age misreporting) $L_{85+} = l_{85}/M_{85+}$

### 9.16.6 Construction of life tables by single calendar year

Methods used to produce single calendar year life tables from pivotal

- Intercensal life tables from OGIVE pivotal: 1912-1944, yearly life tables are computed using linear interpolation between two successive pivotal
- Intercensal life tables from Vital Statistics: 1940-2010, yearly life tables are computed directly from adjusted vital statistics

# 9.17 Paraguay

## 9.17.1 Vital statistics

### Deaths

- 1936-1948 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-79, 80+)
- 1950 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-34, 35-44, 45-54, 55-64, 65+)
- 1951 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75+)
- 1952-1954 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+)
- 1955-1958 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-34, 35-44, 45-54, 55-64, 65+)
- 1959 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-29, 30-39, 40-49, 50-59, 60+)
- 1960-1961 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+)
- 1962-1966 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-34, 35-44, 45-54, 55-64, 65+)
- 1967-1987 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75+)
- 1988-1995 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+)
- 1996-2016 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85-89, 90-94, 95+)

### Causes of Death

- ICD7: 1961-1963, 1965-1967
- ICD8: 1968-1978
- ICD9: 1980–1991, 1994, 1995
- ICD10: 1997-2016

### Births

- 1905-2012

**Vital statistics sources**

- U.S. Bureau of Census. Paraguay: Summary of Biostatistics
- United Nations Yearbook: 1948-2012
- World Health Organization. Health statistics and information systems: 1961-2016 https://www.who.int/healthinfo/statistics/mortality_rawdata/en/

### 9.17.2 Population censuses

- 1950, 1962, 1972, 1982, 1992, 2002, 2012 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+)

**Population censuses sources**

- United Nations Yearbook: 1948-2015

### 9.17.3 Adult pivotal life tables (above age 5)

**Adjustments to generate adult pivotal life tables**

1. Completeness

   Pivotal years for life table estimated using generalized OGIVE method: None

   Pivotal year for life table estimated using adjustments based on Brass: None

   Pivotal years for life tables estimated using adjustments based on Bennett-Horiuchi: 1956, 1967, 1977, 1987, 1997, 2006

2. Age misreporting

   OGIVE: No

   Brass: NA

   Bennett-Horiuchi: Yes

3. Conciliation Yes

### 9.17.4 Estimates of child mortality: 0, 1-4 and 0-5

**Sources of raw data**

- For pivotal yearly estimates from OGIVE method: Model Life Tables (West and South) None
- Adjusted Estimates from Vital Statistics: None
- Census: 1972, 1982, 1992, 2002 (Indirect estimates)

- Surveys:

  1977 (Direct and indirect estimates from World Fertility Survey)

  1990 (Direct and indirect estimates from Demographic and Health Survey)

  1995, 2004, 2008 (Direct estimates from National Survey of Demography and Reproductive Health)

- Third party estimates:

  UN/CELADE: 1955, 1960, 1965, 1970, 1975, 1980, 1985, 1990, 1995, 2000, 2005, 2010, 2015

  Others: None

**Origins of final child mortality estimate**

Table 9.16 summarizes the origin of the final estimates of child mortality for each pivotal year

Table 9.16: Class of final child mortality estimates by time periods

| Age group | Before 1900 | 1900-1950 | 1950-2012 |
|---|---|---|---|
| 0-1 | None | None | Lowess & Splines |
| 1-4 | None | None | Lowess & Splines |

### 9.17.5 Computation of complete pivotal life tables

- Single-year of age life tables: 1956, 1967, 1977, 1987, 1997, 2006
- Abridged life tables: 1956, 1967, 1977, 1987, 1997, 2006
- Treatment of open age group: based on adjusted (for completeness and age misreporting) $L_{85+} = l_{85}/M_{85+}$

### 9.17.6 Construction of life tables by single calendar year

Methods used to produce single calendar year life tables from pivotal

- Intercensal life tables from OGIVE pivotal: None
- Intercensal life tables from Vital Statistics: 1950-2010, yearly life tables are computed directly from adjusted vital statistics

## 9.18 Peru

### 9.18.1 Vital statistics

#### Deaths

- 1939-1943 (ages: 0, 1-4, 5-14, 15-24, 25-34, 35-44, 45-54, 55-64, 65+)
- 1944-1951 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+)
- 1952-1954 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75+)
- 1955-1956 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+)
- 1957-1961 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60+)
- 1962-1964 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75+)
- 1965-1998 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+)
- 1999-2015 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85-89, 90-94, 95+)

#### Causes of Death

- ICD7: 1966-1967
- ICD8: 1968-1973, 1977, 1978
- ICD9: 1980-1983, 1986-1992, 1994-1998
- ICD10: 1999-2015

#### Births

- 1923-2012

#### Vital statistics sources

- U.S. Bureau of Census. Peru: Summary of Biostatistics
- United Nations Yearbook: 1935-2010
- World Health Organization. Health statistics and information systems: 1966-2015 https://www.who.int/healthinfo/statistics/mortality_rawdata/en/

### 9.18.2 Population censuses

- 1876, 1940, 1961, 1972, 1981, 1993, 2007 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+)

**Population censuses sources**

- Census from Peru Statistical Office: 1876, 1940
- United Nations Yearbook: 1948-2010

### 9.18.3 Adult pivotal life tables (above age 5)

**Adjustments to generate adult pivotal life tables**

1. Completeness

   Pivotal years for life table estimated using generalized OGIVE method: 1876, 1908, 1940

   Pivotal year for life table estimated using adjustments based on Brass: None

   Pivotal years for life tables estimated using adjustments based on Bennett-Horiuchi: 1950, 1966, 1976, 1987, 2000, 2008

2. Age misreporting

   OGIVE: No

   Brass: NA

   Bennett-Horiuchi: Yes

3. Conciliation Yes

### 9.18.4 Estimates of child mortality: 0, 1-4 and 0-5

**Sources of raw data**

- For pivotal yearly estimates from OGIVE method: Model Life Tables (West and South) 1876, 1908, 1940
- Adjusted Estimates from Vital Statistics: 1940-1945
- Census: 1972, 1981, 1993, 2007 (Indirect estimates)
- Surveys:

  1974, 1976 (Direct and indirect estimates from Demographic Survey)

  1978 (Direct and indirect estimates from World Fertility Survey)

  1981 (Indirect estimates from Contraceptive Prevalence Survey)

  1986, 1992, 1996, 2000, 2004, 2005, 2008, 2009, 2011, 2012, 2013, 2014 (Direct and indirect estimates from Demographic and Health Survey)

- Third party estimates:

  UN/CELADE: 1955, 1960, 1965, 1970, 1975, 1980, 1985, 1990, 1995, 2000, 2005, 2010, 2015

  Others: 1940

**Origins of final child mortality estimate**

Table 9.17 summarizes the origin of the final estimates of child mortality for each pivotal year

Table 9.17: Class of final child mortality estimates by time periods

| Age group | Before 1900 | 1900-1950 | 1950-2012 |
|---|---|---|---|
| 0-1 | Gompertz | Gompertz | Lowess & Splines |
| 1-4 | Gompertz | Gompertz | Lowess & Splines |

### 9.18.5 Computation of complete pivotal life tables

- Single-year of age life tables: 1876, 1908, 1940, 1950, 1966, 1976, 1987, 2000, 2008
- Abridged life tables: 1876, 1908, 1940, 1950, 1966, 1976, 1987, 2000, 2008
- Treatment of open age group: based on adjusted (for completeness and age misreporting) $L_{85+} = l_{85}/M_{85+}$

### 9.18.6 Construction of life tables by single calendar year

Methods used to produce single calendar year life tables from pivotal

- Intercensal life tables from OGIVE pivotal: 1877-1939, yearly life tables are computed using linear interpolation between two successive pivotal
- Intercensal life tables from Vital Statistics: 1940-2009, yearly life tables are computed directly from adjusted vital statistics

# 9.19 Uruguay

## 9.19.1 Vital statistics

### Deaths

- 1905-1907, 1909-1921, 1923, 1929-1996 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+)
- 1997-2016 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85-89, 90-94, 95+)

### Causes of Death

- ICD7: 1955-1960, 1963-1967
- ICD8: 1968-1978
- ICD9: 1980-1990, 1993-1996
- ICD10: 1997-2010, 2012-2016

### Births

- 1899-2012

### Vital statistics sources

- Fortalecimiento Institucional del Sector Salud/ Ministerio de Salud Publica: 1905-1947
- United Nations Yearbook: 1948-2010
- World Health Organization. Health statistics and information systems: 1955-2016 https://www.who.int/healthinfo/statistics/mortality_rawdata/en/

## 9.19.2 Population censuses

- 1900, 1908, 1963, 1975, 1985, 1996, 2004, 2011 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+)

### Population censuses sources

- Census from Uruguay Statistical Office: 1900, 1908
- United Nations Yearbook: 1948-2015

### 9.19.3 Adult pivotal life tables (above age 5)

#### Adjustments to generate adult pivotal life tables

1. Completeness

   Pivotal years for life table estimated using generalized OGIVE method: 1900, 1904, 1908, 1935

   Pivotal year for life table estimated using adjustments based on Brass: 1908

   Pivotal years for life tables estimated using adjustments based on Bennett-Horiuchi: 1969, 1980, 1990, 2000, 2007

2. Age misreporting

   OGIVE: No

   Brass: Yes

   Bennett-Horiuchi: Yes

3. Conciliation Yes

### 9.19.4 Estimates of child mortality: 0, 1-4 and 0-5

#### Sources of raw data

- For pivotal yearly estimates from OGIVE method: Model Life Tables (West and South) 1900, 1904, 1908, 1935

- Adjusted Estimates from Vital Statistics: 1908-1935, 1955-2012

- Census: 1975, 1985, 1996, 2011 (Indirect estimates)

- Surveys: None

- Third party estimates:

  UN/CELADE: 1955, 1960, 1965, 1970, 1975, 1980, 1985, 1990, 1995, 2000, 2005, 2010, 2015

  Others: None

#### Origins of final child mortality estimate

Table 9.18 summarizes the origin of the final estimates of child mortality for each pivotal year

Table 9.18: Class of final child mortality estimates by time periods

| Age group | Before 1900 | 1900-1950 | 1950-2012 |
|---|---|---|---|
| 0-1 | None | Gompertz | Lowess & Splines |
| 1-4 | None | Gompertz | Lowess & Splines |

### 9.19.5 Computation of complete pivotal life tables

- Single-year of age life tables: 1900, 1904, 1908, 1935, 1969, 1980, 1990, 2000, 2007
- Abridged life tables: 1900, 1904, 1908, 1935, 1969, 1980, 1990, 2000, 2007
- Treatment of open age group: based on adjusted (for completeness and age misreporting) $L_{85+} = l_{85}/M_{85+}$

### 9.19.6 Construction of life tables by single calendar year

Methods used to produce single calendar year life tables from pivotal

- Intercensal life tables from OGIVE pivotal: 1901-1934, yearly life tables are computed using linear interpolation between two successive pivotal
- Intercensal life tables from Vital Statistics: 1905-2010, yearly life tables are computed directly from adjusted vital statistics

## 9.20 Venezuela

### 9.20.1 Vital statistics

**Deaths**

- 1933-1942 (ages: 0, 1-4, 5-9, 10-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70+)
- 1943-1946 (ages: 0, 1-4, 5-9, 10-19, 20-29, 30-39, 40-49, 50-59, 60+)
- 1947-1954 (ages: 0, 1-4, 5-9, 10-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70+)
- 1955-1995 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+)
- 1996-2013 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85-89, 90-94, 95+)

**Causes of Death**

- ICD7: 1955-1967
- ICD8: 1968-1978
- ICD9: 1979-1983, 1985-1990, 1992-1994
- ICD10: 1996-2013

**Births**

- 1904-2012

**Vital statistics sources**

- U.S. Bureau of Census. Venezuela: Summary of Biostatistics
- United Nations Yearbook: 1948-2010
- World Health Organization. Health statistics and information systems: 1955-2013 https://www.who.int/healthinfo/statistics/mortality_rawdata/en/

### 9.20.2 Population censuses

- 1926, 1936, 1941, 1950, 1961, 1971, 1981, 1990, 2001, 2011 (ages: 0, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+)

**Population censuses sources**

- Census from Venezuela Statistical Office: 1926, 1936, 1941
- United Nations Yearbook: 1948-2011

### 9.20.3 Adult pivotal life tables (above age 5)

#### Adjustments to generate adult pivotal life tables

1. Completeness

   Pivotal years for life table estimated using generalized OGIVE method: 1926, 1931, 1936, 1938, 1941, 1945

   Pivotal year for life table estimated using adjustments based on Brass: None

   Pivotal years for life tables estimated using adjustments based on Bennett-Horiuchi: 1955, 1966, 1976, 1985, 1995, 2006

2. Age misreporting

   OGIVE: No

   Brass: NA

   Bennett-Horiuchi: Yes

3. Conciliation Yes

### 9.20.4 Estimates of child mortality: 0, 1-4 and 0-5

#### Sources of raw data

- For pivotal yearly estimates from OGIVE method: Model Life Tables (West and South) 1926, 1931, 1936, 1938, 1941, 1945

- Adjusted Estimates from Vital Statistics: 1938-1945, 1955-2009

- Census: 1981, 1990, 2001, 2011 (Indirect estimates)

- Surveys:

  1977 (Direct and indirect estimates from World Fertility Survey)

- Third party estimates:

  UN/CELADE: 1955, 1960, 1965, 1970, 1975, 1980, 1985, 1990, 1995, 2000, 2005, 2010, 2015

  Others: 1938, 1941, 1945

#### Origins of final child mortality estimate

Table 9.19 summarizes the origin of the final estimates of child mortality for each pivotal year

Table 9.19: Class of final child mortality estimates by time periods

| Age group | Before 1900 | 1900-1950 | 1950-2012 |
|---|---|---|---|
| 0-1 | None | Gompertz | Lowess & Splines |
| 1-4 | None | Gompertz | Lowess & Splines |

### 9.20.5 Computation of complete pivotal life tables

- Single-year of age life tables: 1926, 1931, 1936, 1938, 1941, 1945, 1955, 1966, 1976, 1985, 1995, 2006
- Abridged life tables: 1926, 1931, 1936, 1938, 1941, 1945, 1955, 1966, 1976, 1985, 1995, 2006
- Treatment of open age group: based on adjusted (for completeness and age misreporting) $L_{85+} = l_{85}/M_{85+}$

### 9.20.6 Construction of life tables by single calendar year

Methods used to produce single calendar year life tables from pivotal

- Intercensal life tables from OGIVE pivotal: 1927-1944, yearly life tables are computed using linear interpolation between two successive pivotal
- Intercensal life tables from Vital Statistics: 1935-2010, yearly life tables are computed directly from adjusted vital statistics

# Chapter 10

# Supporting materials

This final chapter of the documentation contains supplemental information used but not described in detail in preceding chapters.

## 10.1 Appendix: Life tables for the period 2010-2020

### 10.1.1 Introduction

The estimation of life tables requires data on deaths and populations by age. As described in this chapter data for Latin American countries, even during the 21st century, require adjustments for completeness and age reporting before they can be used as input for life tables.

### 10.1.2 Data sources: 2010-2020

Population and death counts for the period 2010-2020[1] were retrieved from different sources: primarily, UN Demographic Yearbooks and WHO on-line mortality data. We also resorted to national vital statistics, published by countries' National Statistics Offices and/or Secretaries of Health. Data were available by sex and 5-year age groups, from age 5 to 9, 10 to 14, up to 85+ (open age group). We also collected information on birth counts from the aforementioned sources. This information was used to aid the estimation of infant and child mortality. The list of countries and year with information on vital statistics is in Table 10.1.

Population counts originate mainly from final reports of population censuses carried out between 2010 and 2020. When not available, we use population estimates or projections elaborated by the United Nations Population Division or CELADE. This information is displayed in Table 10.2.

### 10.1.3 Definition of pseudo-pivotal years

We will refer to the period post-2010 as a pseudo intercensal period, $t, t+k$. This is because $t$ can be either the year of the last or next to last population census and $t+k$ can be

[1]Although we refer throughout to the interval 2010-2020, the new information on which life tables for this period are computed is from sources whose year of reference is 2018 at the latest. The precise year up to which information for each country is available will be noted in each case.

Table 10.1: Information for birth and death counts for the period 2010-2020

| Country | Year |
|---|---|
| Argentina | 2011-2017 |
| Brazil | 2011-2017 |
| Chile | 2011-2017 |
| Colombia | 2011-2018 |
| Costa Rica | 2011-2018 |
| Cuba | 2011-2017 |
| Dom Rep | 2011-2018 |
| Ecuador | 2011-2017 |
| El Salvador | 2011-2014 |
| Guatemala | 2011-2017 |
| Mexico | 2011-2017 |
| Nicaragua | 2011-2017 |
| Panama | 2011-2017 |
| Paraguay | 2011-2018 |
| Peru | 2011-2017 |
| Uruguay | 2011-2017 |
| Venezuela | 2011-2017 |

Table 10.2: Country-years with available population censuses or estimates or projections

| Census | Projections |
|---|---|
| Chile 2017 | Argentina 2017 |
| Colombia 2018 | Brazil 2017 |
| Guatemala 2018 | Costa Rica 2018 |
| Peru 2017 | Cuba 2017 |
| NA | Dominican Rep 2018 |
| NA | Ecuador 2017 |
| NA | El Salvador 2014 |
| NA | Mexico 2017 |
| NA | Nicaragua 2017 |
| NA | Panama 2017 |
| NA | Paraguay 2018 |
| NA | Peru 2017 |
| NA | Uruguay 2017 |
| NA | Venezuela 2017 |

Table 10.3: Pseudo-pivotal years by country

| Country | Pivotal Year |
|---|---|
| Argentina | 2014 |
| Brazil | 2013 |
| Chile | 2009 |
| Colombia | 2011 |
| Costa Rica | 2014 |
| Cuba | 2014 |
| Dominican Rep | 2014 |
| Ecuador | 2013 |
| El Salvador | 2010 |
| Guatemala | 2010 |
| Mexico | 2013 |
| Nicaragua | 2011 |
| Panama | 2013 |
| Paraguay | 2010 |
| Peru | 2012 |
| Uruguay | 2014 |
| Venezuela | 2014 |

either the year of the last population census or, alternatively, the year of the last population projection/estimation available to us. For example, for Chile $t = 2000$ is the year of the next to last census whereas $t + k = 2017$ is the year of the last population census. In this case, the middle of the period (the year 2009) is indeed a pivotal year according to the definition we used for the period before 2010. Instead, in the case of Argentina $t = 2010$ is the year of the last census but $t = 2017$ is the year of the last population projection. In this case, the entire period is not, strictly speaking, an intercensal period (centered in the year 2014). To distinguish them from what were strictly intercensal pivotal years for the period before 2010, we will refer to the middle of these pseudo-intercensal periods as pseudo-pivotal years, $PY$. Table 10.3 displays the list of pseudo-pivotal years for each country.

### 10.1.4 Adjustment of adult mortality for completeness and age reporting

To estimate life tables for the year $PV$ we first adjusted for completeness observed mortality rates for ages 5 and above by age and sex using Brass I and Bennett-Horiuchi variants. We also corrected for net age overreporting. The Bennett-Horiuchi variants and Brass correction factors were computed using observed accumulated deaths for the entire pseudo-intercensal period and the two sets of population estimates (from census and or estimates and projections), with $PY$ as a centering year. We also corrected for old age net overreporting applying estimates of $\theta_1$ and $theta_3$ that were applicable to the last pivotal year during the period 2000-2010.[2] The adjustment factors for completeness and age misreporting where then applied to observed mortality rates for the three year period centered in $PY$ and a pseudo-pivotal life table was computed for $PY$. The same adjustment factors were applied to years before and after $PY$ to construct yearly adjusted adult life tables.Table 10.4 lists

[2]See body of chapter for a description of these two parameters

Table 10.4: Estimates of relative completeness for pseudo-pivotal years during 2010-2020

| Country | Pivotal year | Sex | Brass | BH |
|---|---|---|---|---|
| Argentina | 2013 | Male | 1.033 | 0.995 |
| | 2013 | Female | 1.022 | 0.99 |
| Brazil | 2013 | Male | 1.003 | 0.996 |
| | 2013 | Female | 0.999 | 0.959 |
| Chile | 2009 | Male | 0.996 | 0.98 |
| | 2009 | Female | 0.999 | 0.978 |
| Colombia | 2011 | Male | 0.994 | 0.8 |
| | 2011 | Female | 0.992 | 0.835 |
| Costa Rica | 2014 | Male | 0.974 | 0.98 |
| | 2014 | Female | 0.976 | 0.97 |
| Cuba | 2014 | Male | 0.979 | 0.989 |
| | 2014 | Female | 0.97 | 0.995 |
| Dom Rep | 2014 | Male | 1.038 | 0.604 |
| | 2014 | Female | 1.02 | 0.6 |
| Ecuador | 2013 | Male | 1.015 | 0.8 |
| | 2013 | Female | 1.025 | 0.795 |
| El Salvador | 2010 | Male | 0.958 | 0.72 |
| | 2010 | Female | 0.961 | 0.76 |
| Guatemala | 2010 | Male | 1.003 | 0.94 |
| | 2010 | Female | 1.009 | 0.9 |
| Mexico | 2013 | Male | 1.021 | 0.97 |
| | 2013 | Female | 0.97 | 0.96 |
| Nicaragua | 2011 | Male | 0.935 | 0.561 |
| | 2011 | Female | 0.94 | 0.52 |
| Panama | 2013 | Male | 1.001 | 0.92 |
| | 2013 | Female | 1.002 | 0.9 |
| Paraguay | 2010 | Male | 0.988 | 0.691 |
| | 2010 | Female | 0.97 | 0.651 |
| Peru | 2012 | Male | 0.99 | 0.593 |
| | 2012 | Female | 0.977 | 0.603 |
| Uruguay | 2014 | Male | 0.996 | 0.996 |
| | 2014 | Female | 0.985 | 0.968 |
| Venezuela | 2014 | Male | 1.035 | 0.895 |
| | 2014 | Female | 1.005 | 0.842 |

estimates of relative completeness centered at $PY$, the mid-point of the period $t, t+k$ for each country

### 10.1.5 Direct estimates of child mortality

#### Estimation for pseudo-pivotal years

To complete abridged life tables, we estimate infant and child mortality for all years during the period 2010-2020. To accomplish this we rely on data from three different sources: vital statistics (births, deaths), population censuses, and our own estimates for the period 2000-2010. Infant and child mortality rates for the last pseudo-pivotal years were estimated using a simple procedure described below.[3] The procedure for estimating infant and child mortality uses as inputs observed infant and child deaths, births and census populations

[3] Recall that a pseudo-pivotal year for the period 2010-2020 does not always refer to the midpoint between two censuses but may identify the midpoint between the last recorded census and the last population projection/estimation available.

counts. What follows is a description of computations

- Estimates of infant mortality were obtained using the following expressions:

$$Q_0(PY) = D_0(PY)/B(PY) \tag{10.1.1}$$

where $B(PY)$ and $D_0$ are the observed number of births and infant deaths in the the three years centered in the pseudo-pivotal year $PY$. We correct these figures using adjustment factors derived from adjusted infant mortality rates for the years before 2010. The adjustment factors are computed as follows:

$$R_0(PY^*) = Q_0^{vital}(PY^*) - 1)/Q_0^{splines}(PY^*) \tag{10.1.2}$$

where $Q_0^{splines}(PY^*)$ is the adjusted infant mortality rates for the year $PY^*$ or *the last pivotal year before 2010*[4] and $Q_0^{vital}(PY^*)$ is infant mortality rate computed from the observed data (births and infant deaths) for the same year. The adjusted mortality rate for the period centered in the pseudo-pivotal year is

$$Q_0^{adj}(PY) = Q_0^{vital}(PY^*)/R_0(PY^*) \tag{10.1.3}$$

for any $PY > 2010$.

- To compute early child mortality rates we consider all deaths in the age group $1-4$ observed during the three year period centerd in $PY$ We then estimate the average population in the age group $1-4$ in the same three years centered in $PY$ and use it as. That is,

$$Q_{1-4}(PY) = \widehat{D_{1-4}^{obs}(PY)}/\widehat{P_{1-4}}(PY) \tag{10.1.4}$$

where $\widehat{D_{1-4}^{obs}(PY)}$ is the average number of deaths in the age group $1-4$ during the three year period centered in $PY$ and $\widehat{P_{1-4}(PY)}$ is the average exposure in the same age interval and for the three year period. The corresponding adjustment factor is calculated as

$$R_{1-4}(PY) = Q^{vital}(PY)/Q^{splines}(PY) \tag{10.1.5}$$

and the adjusted rates for all years $t$ in the interval 2010-2020

$$Q_{1-4}^{adj}(t) = Q_{1-4}^{vital}(t/R_{1-4}(PY) \tag{10.1.6}$$

The probabilities thus obtained were transformed into mortality rates for single ages 0 to 4 using the Coale-Demeny West model sex-specific separation factors. To complete the pseudo-pivotal and all off- pseudo pivotal complete life tables we joined the estimated adult life tables with estimates for infant and early child mortality rates.

[4]See body of chapter for description of method of estimation based on splines.

### 10.1.6 Summary

The complete set of abridged life tables for the period 2010-2020 includes both a pseudo-pivotal and multiple yearly life tables. The pivotal life tables are those constructed for the middle of the reference period, a year bounded by a population single census count on or before 2010 and either a population census or a population projection or estimates for the most recent year with available data. The yearly life tables are computed using adjusted yearly death rates during the period. The raw rates are computed using observed deaths counts in vital statistics and interpolated population counts. The adjustment factors for completeness and age misreporting are those that apply to the pseudo-pivotal year, e.g. those used to compute the pseudo-pivotal life table.

Finally, single year of age life tables were constructed applying Sprague polynomial interpolation on the observed populations counts in 5-year-age groups. The converted counts were then used to compute death rates in single years of age and these, in turn, were adjusted with the same adjustment factors derived before.

These new pseudo-pivotal and yearly life tables for the period 2010-2020, abridged or in single years of age, are consistent with those estimated for years before 2010 and, together, they summarize the mortality experience of the LAC region for nearly one hundred and seventy years.

## 10.2 Definition of demographic profiles for the simulation

Five different master populations profiles were created, one stable and four non-stable populations. In each case we start with a stable population in 1900 and we compute yearly populations until the year 2000. The age distribution is in single years of age but for totals (not by gender).

The four non-stable populations were generated following approximately the mortality and fertility schedules for Costa Rica, Mexico, Guatemala, Argentina, and Uruguay for the period 1900-2000.

### 10.2.1 Stable population

The stable population is generated using constant values for $GRR = 3.03$ and $E(0) = 45$ for the period 1900 and 2000 with a natural rate of increase $r = 0.025$.

### 10.2.2 Non-stable populations (a)(b)(c)

Table 10.5 displays values of fertility and mortality time-dependent parameters that fully characterize four different non-stable population profiles.

Table 10.5: Non-stable population profiles.$_{a,b,c}$

| | I | | | II | | | III | | | IV | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Year$E(0)$ | $GRR$ | $r$ | $E(0)$ | $GRR$ | $r$ | $E(0)$ | $GRR$ | $r$ | $E(0)$ | $GRR$ | $r$ |
| 190034.70 | 3.60 | 0.05 | 26.30 | 6.20 | 0.04 | 22.10 | 5.80 | 0.03 | 45.40 | 1.80 | 0.02 |
| 191035.10 | 3.40 | 0.05 | 29.60 | 5.70 | 0.04 | 25.40 | 5.70 | 0.03 | 48.90 | 1.70 | 0.02 |
| 192035.10 | 3.20 | 0.05 | 32.90 | 5.20 | 0.04 | 28.70 | 5.20 | 0.03 | 51.30 | 1.60 | 0.02 |
| 193042.20 | 2.60 | 0.05 | 36.20 | 4.70 | 0.04 | 32.00 | 4.70 | 0.03 | 54.40 | 1.50 | 0.02 |
| 194046.90 | 2.50 | 0.05 | 41.80 | 4.20 | 0.04 | 37.40 | 3.80 | 0.03 | 59.60 | 1.40 | 0.02 |
| 195055.60 | 2.40 | 0.05 | 50.70 | 3.40 | 0.04 | 40.20 | 3.50 | 0.03 | 66.30 | 1.30 | 0.02 |
| 196062.60 | 2.30 | 0.05 | 58.50 | 3.30 | 0.04 | 47.00 | 3.30 | 0.03 | 68.40 | 1.40 | 0.02 |
| 197065.40 | 2.10 | 0.05 | 62.60 | 3.20 | 0.04 | 53.90 | 3.10 | 0.03 | 68.80 | 1.50 | 0.02 |
| 198072.60 | 1.70 | 0.05 | 67.70 | 2.10 | 0.04 | 58.20 | 3.00 | 0.03 | 71.00 | 1.30 | 0.02 |
| 199075.70 | 1.50 | 0.05 | 71.50 | 1.50 | 0.04 | 62.60 | 2.60 | 0.03 | 72.80 | 1.20 | 0.02 |
| 200077.30 | 1.30 | 0.05 | 73.40 | 1.20 | 0.04 | 65.90 | 2.20 | 0.03 | 75.20 | 1.10 | 0.02 |

(a) Non Stable population I, II, III and IV follow the patterns of mortality and fertility between 1900 and 2000 assessed with current (Adjusted data) for Costa Rica, Mexico, Guatemala and Argentina/Uruguay respectively.
(b) Population parameters were directly estimated for each decade and then interpolated linearly within each decade to obtain yearly values.

(c) The initial population age distribution for I, II and III correspond to the stable population associated with parameter values in 1900. In case IV the initial population corresponded to the average of census populations closest to 1900.

## 10.3 Proof of lack of identification of parameters of net age overstatement

Using the same notation as in the text we have

$$\Pi^T = (1/\phi^{no})[\hat{\Theta}^S]^{-1}\Pi^O$$

and

$$\Delta^T = (1/\lambda^{no})[\hat{\Theta}^S]^{-1}\Delta^O.$$

In a closed population the relation between the vectors for populations in two successive censuses and the vector of intercensal deaths is:

$$\Pi^T_{t+k} = \Pi^T_t + \Delta^T_{[t,t+k]}. \tag{10.3.1}$$

Using the first two expressions in (10.3.1) yields:

$$(1/\phi^{no})[\hat{\Theta}^S]^{-1}\Pi^O_{t+k} = (1/\phi^{no})[\hat{\Theta}^S]^{-1}\Pi^O_t - (1/\lambda^{no})[\hat{\Theta}^S]^{-1}\Delta^O_{[t,t+k]}. \tag{10.3.2}$$

From (10.3.2) we see that only $(\phi^{no}/\lambda^{no})$ is identifiable with the available information.

## 10.4 Behavior of age misreporting index $cmR^o_{x,[t_1,t_2]}$

The expression of the age misreporting index is

$$cmR^o_{x,[t_1,t_2]} = \frac{cmP^o_{x+k,t_2}/cmP^o_{x,t_1}}{1-(cmD^o_{x,[t_1,t_2]}/cmP^o_{x,t_1})}$$

a ratio of two different estimators of the same quantity, namely the cumulative probability of survival of the population aged $x$ and over at time $t_1$ to age $(x+k)$ and over at time $t_2$. Use of cumulative quantities in the index is an important prerequisite since it minimizes the impact of age misreporting within the bounds of the cumulative quantities. Thus, erroneous transfers over age $x$ do not affect population counts at ages $x$ and over. These quantities are influenced only by transfers from ages younger than $x$ into ages $x$ and above or by transfers from ages $x$ and above to ages younger than $x$. Admittedly, however, use of cumulative quantities complicates the algebra and muddles interpretation. To circumvent this difficulty and preserving the same set up and assumptions defined in the text, we redefine the expression for single years of age to obtain:

$$R^o_{x,[t_1,t_2]} = \frac{P^o_{x+k,t_2}/P^o_{x,t_1}}{1-(D_{x,[t_1,t_2]}/P^o_{x,t_1})}$$

or the ratio of a conventional survival ratio computed from two successive population counts to the survival ratio computed from the complement of a measure of the conditional probability of dying between the two censuses. If the population is stationary, the numerator is simply the ratio $L_{x+k}/L_x$ in a life table and the denominator is the complement of the probability of dying in the intercensal period, namely, $1-(1-L_{x+k}/L_x)$. From this it follows that,

$$\ln\left(R^o_{x,[t_1,t_2]}\right) \sim -I^N_{x,x+k} - \ln\left(1-\left[1-\exp\left(-I^D_{x,x+k}\right)\right]\right) \tag{10.4.1}$$

where $I^D_{x,x+k}$ and $I^N_{x,x+k}$ are estimators of the integrated hazards between $x$ and $x+k$ consistent with the survival ratios in the denominator and numerator respectively. When the population is closed to migration, there is perfect coverage and no net age overstatement, expression (10.4.1) equals 0 as both estimators of the integrated hazards are identical. When there is age overstatement expression (10.4.1) becomes

$$\ln\left(R^o_{x,[t_1,t_2]}\right) \sim \ln\left(\frac{h(x+k)}{h(x)}\right) - I^N_{x,x+k} - \ln\left(1-\frac{g(x)}{h(x)}\left[1-\exp(-I^D_{x,x+k})\right]\right) \tag{10.4.2}$$

where $h(.)$ and $g(.)$ are defined in the text and refer to increasing functions of age that reflect age overstatement of ages of population and deaths respectively. When these functions are equal to 1, there is neither population nor death age overstatement or, if there is, their effects cancel each other out. Expression (10.4.2) can be simplified if we expand the inner log expression in a Taylor series around a value of $f(x) = g(x)/h(x) = 1$:

$$\ln\left(R^o_{x,[t_1,t_2]}\right) \sim \ln\left(\frac{h(x+k)}{h(x)}\right) - I^N_{x,x+k} + \left(\frac{g(x)}{h(x)}-1\right)(1+I^D_{x,x+k}) + I^D_{x,x+k} \tag{10.4.3}$$

an expression that reduces to 0 when $h(x+k)/h(x)$ =1 and $f(x) = 1$.

Expression (10.4.3) is the analytic support for inferences regarding the effects of age overstatement on the index of age misstatement $cmR_{x,[t_1,t_2]}$ (see text). Deviations from the assumption of population stationarity introduce only minor changes in the algebra but leave the implications of expression (10.4.3) intact. However, when, as required by the original index, we restore the cumulative functions, the algebra becomes intractable even in the case of a stationary population. The way out of this conundrum is to think of the cumulative ratios as functions not of the exact integrated hazards, as in expressions (10.4.1)-(10.4.3) but rather as expressions of mean values of corresponding integrated hazards. Thus, in a stationary population, the survival ratio of the cumulative populations at ages $x$ and $x+k$ is the ratio $T(x+k)/T(x)$ which can be written as $\int_{x+k}^{\infty}[\exp(-\int_0^y \mu(s)ds)]dx/\int_x^{\infty}[\exp(-\int_0^y \mu(s)ds)]dx$. Using the mean value theorem in numerator and denominator leads to the approximation $\exp(-\int_{x+i}^{x+k+i'} \mu(s)ds)$ or, more generally, $\exp(-\int_{x^*}^{x^{**}} \mu(s)ds)$ where $x^* > x$ and $x^{**} > x+k$. Upon taking logs in this expression we retrieve an integrated hazard that expresses integration of the force of mortality over two ages that are not fixed *ex ante* (such as $x$ and $x+k$) but, rather, between limits (ages) that are a function of the underlying force of mortality. For this reason, in the text, we use the symbols $I^N_{x,x+k}$ and $I^D_{x,x+k}$ associated with cumulative quantities as "integrated hazard analogues".

## 10.5 Comparison of relative census completeness

Table 10.6 displays alternative estimates of relative completeness of two consecutive censuses during the post 1950 period. The table displays LAMBdA estimates of $(C_1/C_2)$ from Brass-Hill method and alternative ratios computed by CELADE (these are based on computations using indirect techniques, population projections, and post-census enumeration surveys).

## 10.6 Brief description of methods to estimate completeness of death registration

**Bennett-Horiuchi No 1, (BH_1)**

Bennett-Horiuchi (1981) completeness factor can be estimated in two different ways. First (cumulated from bottom to top), as the ratio of the estimated number of persons-years in the age group "a" to "a+5" (${}_{10}\hat{N}_{a-5}$) to the observed average number of persons-years in the age group "a" to "a+5" over the ten-year period (${}_{10}N_{a-5}$). This was labeled as $C5$ because the computation of cumulated numbers started at age 5. Second (cumulated from top to bottom), a more robust measurement was created by cumulating ${}_{10}\hat{N}_{a-5}$ and ${}_{10}N_{a-5}$ from top to bottom; since the accumulation started at age 75, the ratio was tagged as $C75$.

- Bennett-Horiuchi No 1 is the average of these two estimates.

**Bennett-Horiuchi No 2, (BH_2)**

Bennett-Horiuchi (1984) is a variation of the previous estimates that introduces a slight correction in the calculation of ${}_{10}\hat{N}_{a-5}$ for ages above 60. This new indicator can also be computed in two different ways: first, as the ratio of ${}_{10}\hat{N}_{a-5}$ to ${}_{10}N_{a-5}$ ($C5$), and as the ratio of the accumulated values of ${}_{10}\hat{N}_{a-5}$ to ${}_{10}N_{a-5}$ ($C75$).

- Bennett-Horiuchi No 2 is the average of these two estimates.

Table 10.6: Alternative estimates of consecutive census relative completeness

| Country, | Males | | Females | | Country, | Males | | Females | |
|---|---|---|---|---|---|---|---|---|---|
| Year | LAMBdA | CELADE* | LAMBdA | CELADE* | Year | LAMBdA | CELADE* | LAMBdA | CELADE* |
| Argentina | | | | | Guatemala | | | | |
| 1947-1960 | 0.960 | 0.967 | 0.940 | 0.944 | 1950-1964 | 0.912 | 0.965 | 0.924 | 0.952 |
| 1960-1970 | 0.981 | 0.989 | 0.995 | 0.999 | 1964-1973 | 1.047 | 1.073 | 1.105 | 1.058 |
| 1970-1980 | 0.984 | 0.974 | 0.987 | 0.988 | 1973-1981 | 1.062 | 1.057 | 1.073 | 1.054 |
| 1980-1991 | 0.997 | 1.012 | 0.986 | 1.010 | 1981-1994 | 0.974 | 0.999 | 0.984 | 1.013 |
| 1991-2001 | 1.033 | 1.017 | 1.022 | 1.012 | 1994-2002 | 0.966 | 0.898 | 1.011 | 0.904 |
| Brazil | | | | | Honduras | | | | |
| 1980-1991 | 0.998 | 1.015 | 1.000 | 1.010 | 1950-1961 | 0.944 | 1.016 | 0.951 | 1.003 |
| 1991-2000 | 0.975 | 0.990 | 0.950 | 0.991 | 1961-1974 | 0.992 | 1.038 | 0.973 | 1.031 |
| 2000-2010 | 1.003 | 1.011 | 0.999 | 1.001 | 1974-1988 | 0.964 | 0.958 | 0.977 | 0.964 |
| Chile | | | | | Mexico | | | | |
| 1952-1960 | 0.950 | 0.967 | 0.937 | 0.964 | 1950-1960 | 0.976 | 0.995 | 0.926 | 1.017 |
| 1960-1970 | 1.035 | 1.034 | 1.050 | 1.031 | 1960-1970 | 0.963 | 0.987 | 0.936 | 0.986 |
| 1970-1982 | 0.938 | 0.949 | 0.949 | 0.952 | 1970-1980 | 0.981 | 0.969 | 0.956 | 0.954 |
| 1982-1992 | 1.002 | 0.997 | 0.994 | 0.998 | 1980-1990 | 0.904 | 1.017 | 0.916 | 1.009 |
| 1992-2002 | 1.011 | 1.011 | 1.002 | 1.016 | 1990-2000 | 0.952 | 0.990 | 0.968 | 0.981 |
| Colombia | | | | | 2000-2010 | 0.954 | 1.006 | 0.958 | 1.001 |
| 1951-1964 | 0.947 | 0.951 | 0.930 | 0.931 | Nicaragua | | | | |
| 1964-1973 | 1.060 | 1.085 | 1.029 | 1.070 | 1950-1963 | 0.971 | 1.035 | 0.945 | 1.030 |
| 1973-1985 | 0.915 | 0.962 | 0.955 | 0.983 | 1963-1971 | 0.941 | 1.048 | 0.955 | 1.033 |
| 1985-1993 | 0.929 | 0.984 | 0.920 | 0.979 | 1971-1995 | 0.883 | 0.793 | 0.879 | 0.815 |
| 1993-2005 | 1.005 | 0.954 | 1.008 | 0.955 | 1995-2005 | 0.935 | 0.985 | 0.940 | 0.992 |
| Costa Rica | | | | | Panama | | | | |
| 1950-1963 | 0.936 | 0.919 | 0.919 | 0.931 | 1950-1960 | 1.031 | 0.981 | 1.019 | 0.988 |
| 1963-1973 | 0.979 | 0.959 | 0.959 | 0.967 | 1960-1970 | 0.975 | 0.952 | 0.949 | 0.951 |
| 1973-1984 | 0.995 | 1.020 | 1.028 | 1.031 | 1970-1980 | 0.958 | 1.014 | 0.958 | 1.024 |
| 1984-2000 | 0.961 | 0.935 | 0.935 | 0.944 | 1980-1990 | 1.023 | 0.979 | 1.016 | 0.978 |
| 2000-2011 | 0.974 | 1.057 | 0.976 | 1.021 | 1990-2000 | 1.000 | 1.005 | 0.988 | 1.006 |
| Cuba | | | | | 2000-2010 | 1.001 | 0.997 | 1.002 | 0.996 |
| 1953-1970 | 0.941 | 0.927 | 0.970 | 0.955 | Paraguay | | | | |
| 1970-1981 | 0.965 | 1.008 | 0.992 | 0.992 | 1950-1962 | 0.967 | 0.991 | 0.980 | 0.997 |
| 1981-2002 | 0.932 | 0.987 | 0.956 | 0.983 | 1962-1972 | 0.990 | 1.004 | 1.025 | 0.994 |
| 2002-2012 | 0.995 | 1.016 | 0.990 | 1.016 | 1972-1982 | 0.994 | 1.006 | 0.968 | 1.012 |
| Dominican Republic | | | | | 1982-1992 | 1.017 | 0.961 | 1.008 | 0.959 |
| 1950-1960 | 0.986 | 0.971 | 0.914 | 0.981 | 1992-2002 | 0.988 | 0.997 | 0.981 | 1.002 |
| 1960-1970 | 0.905 | 1.047 | 0.979 | 1.031 | Peru | | | | |
| 1970-1981 | 0.930 | 0.943 | 0.962 | 0.966 | 1961-1972 | 0.995 | 0.986 | 0.967 | 1.004 |
| 1981-1993 | 0.980 | 1.004 | 0.985 | 0.944 | 1972-1981 | 1.035 | 1.007 | 1.025 | 1.005 |
| 1993-2002 | 0.948 | 0.963 | 0.988 | 1.015 | 1981-1993 | 1.006 | 1.008 | 1.020 | 1.002 |
| 2002-2010 | 1.038 | 1.005 | 1.047 | 1.026 | 1993-2007 | 0.990 | 0.982 | 0.977 | 0.982 |
| Ecuador | | | | | Uruguay | | | | |
| 1950-1962 | 0.993 | 0.996 | 0.998 | 0.989 | 1963-1975 | 0.960 | 1.003 | 0.939 | 0.996 |
| 1962-1974 | 0.951 | 0.973 | 0.988 | 0.965 | 1975-1985 | 0.961 | 0.997 | 0.953 | 1.011 |
| 1974-1982 | 0.912 | 1.001 | 0.990 | 0.997 | 1985-1996 | 0.977 | 1.005 | 0.981 | 1.008 |
| 1982-1990 | 0.900 | 1.016 | 0.985 | 1.024 | 1996-2004 | 0.998 | 0.999 | 0.991 | 0.999 |
| 1990-2001 | 0.970 | 1.000 | 0.975 | 0.997 | 2004-2011 | 0.996 | 1.019 | 0.998 | 1.011 |
| 2001-2010 | 0.937 | 0.975 | 0.950 | 0.978 | Venezuela | | | | |
| El Salvador | | | | | 1950-1961 | 1.041 | 1.030 | 1.004 | 1.033 |
| 1950-1961 | 0.944 | 0.959 | 0.951 | 0.949 | 1961-1971 | 0.958 | 0.998 | 0.985 | 0.970 |
| 1961-1971 | 0.992 | 0.939 | 0.973 | 0.937 | 1971-1981 | 1.016 | 0.973 | 1.005 | 0.986 |
| 1971-1992 | 1.038 | 1.002 | 1.030 | 0.999 | 1981-1990 | 0.943 | 1.009 | 0.953 | 1.007 |
| 1992-2007 | 0.958 | 1.003 | 0.977 | 0.999 | 1990-2001 | 0.928 | 0.982 | 0.943 | 0.976 |
| | | | | | 2001-2011 | 1.035 | 1.013 | 1.005 | 1.031 |

* CELADE's estimates of completeness provided by Guiomar Bay

**Bennett-Horiuchi No 3 (BH_3) & Bennett-Horiuchi No 4 (BH_4)**
These estimates are computed as **Bennett-Horiuchi No 1** & **Bennett-Horiuchi No 2** but now we use adjusted age-specific rate of population growth. The adjustment factors are estimated using Brass classic method.

**Preston-Lahiri 1-2**
Preston-Lahiri method can also estimate two completeness measurements depending on the age at which the functions are calculated (birth rates, death rates, mean age, etc.) We computed two variants: using age 5+ (labeled $C5$) and 10+ (labeled $C10$).

- These are respectively called Preston-Lahiri No 1 (PL_1) and Preston-Lahiri No 2 (PL_2).

## 10.7 Shortcut to estimate mortality adjustment (2SBH_4) in the presence of differential census completeness

### 10.7.1 Introduction

To adjust the family of Bennett-Horiuchi estimates (any of the four variants we use in the construction of LAMBdA; see Table **??**) when the completeness of the first census, $C_1$ is not equal to the completeness of the second census, $C_2$ or $Cc = C_1/C_2 \neq 1$, is not quite as straightforward as one would like and requires a few additional steps. While the rate of intercensal growth $r$ can be corrected using an estimate of $Cc$ and this entered as input in the Bennett-Horiuchi algorithm to compute a completeness estimate, the expression for number of persons years lived in the intercensal period used in the Bennett-Horiuchi algorithm contains in the numerator the *true* difference between populations in census 2 and census 1. This difference cannot be obtained unless one has separate estimates of the corresponding completeness factors which, as a rule, we do not have. Therefore, it is not sufficient to compute an estimate of relative completeness of death registration, $Cm$, with an adjusted intercensal rate of growth instead of the observed one.

It turns out that there is near perfect correspondence between the 'true' estimate of $Cm$ and two ancillary estimates. The first is one computed *after adjusting the rate of intercensal increase* using an estimate of $Cc$ (for example, from Brass-Hill method). The second is computed with no adjustments at all for $Cc$. As shown below, the correspondence is closely approximated by the 'predicted' value of $Cm$ computed from regressions estimated separately in simulated data with $Cc > 1$ and $Cc < 1$. When Cc=1 there is no need for correction and hence these observation are uninformative. Below we describe the relations of interest.

### 10.7.2 Empirical relations

We use eligible simulated populations and estimate regression equations of the following form:

$$Cm = a_1 + b_1 * BH\_1 + c_1 * BH\_1_r + d_1 * Cc$$
$$Cm = a_2 + b_2 * BH\_2 + c_2 * BH\_2_r + d_2 * Cc$$
$$Cm = a_3 + b_3 * BH\_3 + c_3 * BH\_3_r + d_3 * Cc$$
$$Cm = a_4 + b_4 * BH\_4 + c_4 * BH\_4_r + d_4 * Cc$$

where $Cm$ is the correct relative completeness of death registration in a simulated population, labels $BH\_x$ refer to Bennett-Horiuchi estimates, all computed with no correction for differential completeness of census registration and, finally, labels $BH\_x_r$ refer to the same estimates but computed with an intercensal rate of growth adjusted for differential population census completeness, $Cc$.

All regressions are estimated constraining the constants, $a_*$, to 0 and using two sets of simulated subpopulations, with $0.8 < C1 < 1$ and $1 < C1 > 1.20$ respectively. Thus, for each regression equation above, there are 2 sets of estimates that yield predicted values of $Cm$. Table 10.7 below contains estimates and summaries of prediction proportionate errors.

The results of these regressions are in Table 10.7.

### 10.7.3 Summary of shortcut

To arrive at a final estimate associated with the estimator labeled 2SBH_4 in Table **??** proceed as follows:

1. Compute values for $BH\_1_r, BH\_2_r, BH\_3_r$, and $BH\_4_r$ using an external estimate of $Cc$ (for example, Brass-Hill method);

2. Compute predicted values from the regression corresponding to the observed (estimated) value of Cc.

The new estimates are highly accurate 2-stage variants of the original ones used in the regressions and are adjusted for differential census completeness.

Table 10.7: Regression models to predict adjusted BH_4 (2SBH_4) an estimate adjusted for differences in completeness of two censuses.

| Variables | Coef. | Std. Err. | t | P > t | [95% Conf. | Interval] | adjR-square |
|---|---|---|---|---|---|---|---|
| **Model 1: BH_1 & CC <1** | | | | | | | |
| Cc | 0.46081 | 0.01549 | 29.75 | 0.000 | 0.43040 | 0.49123 | 0.996 |
| BH_1 | 0.28842 | 0.00707 | 40.82 | 0.000 | 0.27455 | 0.30230 | |
| BH_1_r | 0.25025 | 0.01388 | 18.03 | 0.000 | 0.22300 | 0.27750 | |
| | | | Prediction Error | | | | |
| Median | | | 0.052 | | | | |
| Mean error | | | 0.061 | | | | |
| Standard deviation | | | 0.047 | | | | |
| **Model 2: BH_1 & Cc > 1** | | | | | | | |
| Cc | 0.45016 | 0.01100 | 40.94 | 0.000 | 0.42857 | 0.47175 | 0.994 |
| BH_1 | 0.64328 | 0.01724 | 37.31 | 0.000 | 0.60943 | 0.67714 | |
| BH_1_r | 0.00253 | 0.00105 | 2.41 | 0.016 | 0.00047 | 0.00459 | |
| | | | Prediction Error | | | | |
| Median | | | 0.059 | | | | |
| Mean | | | 0.068 | | | | |
| Standard deviation | | | 0.048 | | | | |
| **Model 3: BH_2 & Cc < 1** | | | | | | | |
| Cc | 0.47712 | 0.01543 | 30.91 | 0.000 | 0.44681 | 0.50742 | 0.994 |
| BH_2 | 0.28285 | 0.00713 | 39.7 | 0.000 | 0.26886 | 0.29684 | |
| BH_2_r | 0.23944 | 0.01403 | 17.07 | 0.000 | 0.21190 | 0.26698 | |
| | | | Prediction Error | | | | |
| Median | | | 0.053 | | | | |
| Mean | | | 0.062 | | | | |
| Standard Deviation | | | 0.048 | | | | |
| **Model 4: BH_2 & Cc > 1** | | | | | | | |
| Cc | 0.45211 | 0.01100 | 41.12 | 0.000 | 0.43052 | 0.47370 | 0.994 |
| BH_2 | 0.64181 | 0.01728 | 37.14 | 0.000 | 0.60788 | 0.67573 | |
| BH2_r | 0.00216 | 0.00099 | 2.18 | 0.029 | 0.00022 | 0.00411 | |
| | | | Prediction Error | | | | |
| Median | | | 0.059 | | | | |
| Mean | | | 0.069 | | | | |
| Standard Deviation | | | 0.048 | | | | |
| **Model 5: BH_3 & Cc < 1** | | | | | | | |
| Cc | 0.45958 | 0.01549 | 29.67 | 0.000 | 0.42916 | 0.48999 | 0.995 |
| BH_3 | 0.29005 | 0.00709 | 40.9 | 0.000 | 0.27612 | 0.30397 | |
| BH_3_r | 0.24986 | 0.01387 | 18.01 | 0.000 | 0.22263 | 0.27709 | |
| | | | Prediction Error | | | | |
| Median | | | 0.051 | | | | |
| Mean | | | 0.060 | | | | |
| Standard Deviation | | | 0.043 | | | | |
| **Model 6: BH_3 & Cc > 1** | | | | | | | |
| Cc | 0.44971 | 0.01101 | 40.85 | 0.000 | 0.42810 | 0.47133 | 0.993 |
| BH_3 | 0.64412 | 0.01726 | 37.32 | 0.000 | 0.61023 | 0.67800 | |
| BH_3_r | 0.00260 | 0.00107 | 2.42 | 0.016 | 0.00049 | 0.00470 | |
| | | | Prediction Error | | | | |
| Median | | | 0.060 | | | | |
| Mean | | | 0.068 | | | | |
| Standard Deviation | | | 0.051 | | | | |
| **Model 7: BH_4 & Cc < 1** | | | | | | | |
| Cc | 0.47593 | 0.01543 | 30.84 | 0.000 | 0.44563 | 0.50623 | 0.995 |
| BH_4 | 0.28444 | 0.00715 | 39.77 | 0.000 | 0.27040 | 0.29848 | |
| BH_4 | 0.23905 | 0.01402 | 17.05 | 0.000 | 0.21153 | 0.26658 | |
| | | | Prediction Error | | | | |
| Median | | | 0.053 | | | | |
| Mean | | | 0.062 | | | | |
| Standard Deviation | | | 0.046 | | | | |
| **Model 8: BH_4 & Cc > 1** | | | | | | | |
| Cc | 0.45171 | 0.01101 | 41.04 | 0.000 | 0.43010 | 0.47332 | 0.993 |
| BH_4 | 0.64262 | 0.01730 | 37.15 | 0.000 | 0.60866 | 0.67658 | |
| BH_4 | 0.00220 | 0.00101 | 2.18 | 0.029 | 0.00022 | 0.00418 | |
| | | | Prediction Error | | | | |
| Median | | | 0.058 | | | | |
| Mean | | | 0.068 | | | | |
| Standard Deviation | | | 0.048 | | | | |

Note: Prediction error is absolute value of proportionate difference between 2SBH_4 and simulated values of relative completeness ($C_m$).

# Bibliography


Anuarios estadísticos, Republica del Uruguay (1976), "Anuarios Estadísticos," Tech. rep., Instituto Nacional de Estadística.

Arriaga, E. E. (1968), *New Life Tables for Latin American Populations in the Nineteenth and Twentieth Centuries*, Berkeley, CA: Institute of International Studies, University of California, population monograph series no. 3 ed.

Bennett, N. G. and Horiuchi, S. (1981), "Estimating the Completeness of Death Registration in a Closed Population," *Population Index*, 47, 207–221.

Bhat, M. P. (1987), "Mortality in India: Levels, Trends and Patterns," Ph.d., University of Pennsylvania.

Bhat, P. N. M. (1990), "Estimating Transition Probabilities of Age Misstatement," *Demography*, 27, 149–163.

Borges, G. and Sacco, N. (2016), "Has census coverage in Latin America improved with time?" in *International Seminar on Registering and Counting Population: The Production and Exploration of Census Information from Early Modern Times to the End of the 20th Century UNICAMP, NEPO, IUSSP. Campinas, Sao Paolo.*

Brass, W. (1975), *Methods for Estimating Fertility and Mortality from Limited and Defective Data: based on seminars held 16-24 September 1971 at the Centro Latinoamericano de Demografía (CELADE) San José, Costa Rica*, Chapel Hill, NC: International Program of Laboratories for Population Statistics, Dept. of Biostatistics, School of Public Health, Carolina Population Center, Univ. of North Carolina at Chapel Hill.

— (1979), "Evaluation of Birth and Death Registration Using Age Distribution and Child Survivorship Data," *Asian and Pacific Census Forum*, 5, 9–20.

Brass, W. and Bamgboye, E. A. (1981), "The Time Location of Reports of Survivorship: Estimates for maternal and Paternal Orphanhood and the Ever-Widowed," Report, CPS Working Paper No. 81-1, Center for Population Studies, London School of Hygiene and Tropical Medicine, University of London, London, England.

Brass, W. and Coale, A. (1968), "Methods of Analysis and Estimation," in *The Demography of Tropical Africa*, Princeton University Press, book section 3, pp. 88–150.

Chidambaram, V. and Sathar, Z. (1984), "Age and date reporting," in *Comparative Studies: Cross-National Summaries*, Voorburg, Netherlands: International Statistical Institute, vol. 5, pp. 1–37.

Coale, A. and Demeny, P. (1967), *Methods of estimating basic demographic measures from incomplete data*, New York: United Nations.

Coale, A. J., Demeny, P., and Vaughan, B. (1983), "Model Life Tables and Stable Populations," in *Regional Model Life Tables and Stable Populations (Second Edition)*, ed. Vaughan, A. J. C. D., Academic Press, pp. 41–154.

Coale, A. J. and Kisker, E. E. (1986), "Mortality Crossovers: Reality or Bad Data?" *Population Studies*, 40, 389–401.

Collver, O. A. (1965), *Birth Rates in Latin America: New Estimates of Historical Trends and Fluctuations*, Berkeley, CA: Institute of International Studies Research, University of California Press, monograph no. 7 ed.

Condran, G. A., Himes, C., and Preston, S. H. (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*, 30, 23–60.

Danel, I., Graham, W., Stupp, P., and Castillo, P. (1996), "Applying the sisterhood method for estimating maternal mortality to a health facility-based sample: a comparison with results from a household-based sample," *Int J Epidemiol*, 25, 1017–22.

Dechter, A. R. and Preston, S. H. (1991), "Age misreporting and its effects on adult mortality estimates in Latin America," *Popul Bull UN*, 31-32, 1–16.

Del Popolo, F. (2000), "Los Problemas en la Declaración de la Edad de la Población Adulta Mayor en los Censos," Report, CELADE, CEPAL, ECLAC.

Elo, I. T. and Preston, S. H. (1994), "Estimating African-American Mortality from Inaccurate Data," *Demography*, 31, 427–458.

Ewbank, D. C. (1981), *Age Misreporting and age-selection and Remuneration: Sources, Patterns, and Consequences for Demographics Analysis*, Committee on Population and Demography, Washington DC: National Reserach Coucil.

Ferreira-Levy, M. (1974), "O Papel da migracao international na evolucao da populacao brasileira :1872-1972," *Revista de Saude Publica*, 8 (suppl), 49–90.

Gakidou, E., Hogan, M., and Lopez, A. D. (2004), "Adult mortality: time for a reappraisal," *Int J Epidemiol*, 33, 710–7.

Gakidou, E. and King, G. (2006), "Death by survey: Estimating adult mortality without selection bias from sibling survival data," *Demography*, 43, 569–585.

Grushka, C. O. (1996), “Adult and old age mortality in Latin America: Evaluation, adjustments and a debate over a distinct pattern,” Thesis, University of Pennsylvania.

Hill, K. (2000), *Methods for Measuring Adult Mortality in Developing Countries: A comparative review*, Johns Hopkins University.

Hill, K. and Choi, Y. (2004), “Performance of GGB and SEG given various simulated data errors,” .

Hill, K., Choi, Y., and Timaeus, I. (2005), “Unconventional approaches to mortality estimation,” *Demographic Research*, S4, 281–300.

Hill, K. and Trussell, J. (1977), “Further Developments in Indirect Mortality Estimation,” *Population Studies*, 31, 313–334.

Hill, K., You, D., and Choi, Y. (2009), “Death distribution methods for estimating adult mortality: Sensitivity analysis with simulated data errors,” *Demographic Research*, 21, 235–254.

Himes, C., Preston, S., and Condran, G. A. (1994), “A Relational Model of Mortality at Older Ages in Low Mortality Countries,” *Population Studies*, 48, 240–291.

Horiuchi, S. and Coale, A. J. (1985), “Age Patterns of Mortality for Older Women,” .

Kamps E., J. (1976), *La declaración de la edad en los Censos de Población de la América Latina; exactitud y preferencia de dígitos en los Censos de 1950, 1960 y 1970*, vol. Serie C of *Series Históricas*, San José, Costa Rica: ONU, CEPAL, CELADE.

Kannisto, V. (1994), *Development of oldest-old mortality, 1950-1990: evidence from 28 developed countries*, Odense, Denmark: Odense University Press.

Kannisto, V., Lauritsen, L., Thatcher, A., and Vaupel, J. (1994), “Reductions in Mortality at Advanced Ages: Several Decades of Evidence from 27 Countries,” *Population and Development Review*, 20, 793–810.

LAMBdA team (2020), *Two Hundred Years of Mortality Changes in Latin America: From Hunger to Longevity*, To be announced.

Martin, L. (1980), “A Modificationfor Use in Destabilized Populations of Brass’s Technqiue for Estimating Completeness of Death Registration,” *Population Studies*, 34, 381–95.

Mazess, R. B. and Forman, S. H. (1979), “Longevity and age exaggeration in Vilcabamba, Ecuador,” *J Gerontol*, 34, 94–8.

McNeil, D. R., Trussell, T. J., and Turner, J. C. (1977), “Spline interpolation of demographic data,” *Demography*, 14, 245–252.

Mortara, G. (1954), "The development and structure of Brazil's population," *Population Studies*, 8, 121–139.

Nuñez, L. (1984), "Una Aproximacion al Efecto de la Mala Declaracion de la Edad en la Informacion Demografica Recabada en Mexico," Report, Direccion General del Registro Nacional de Poblacion e Identificacion Personal.

Ortega, A. and Garcia, V. (1985), "Estudio Sobre la Mortalidad y Algunas Caracteristicas Socioeconomicas de las Personas de la Tercera Edad: Informe de la Investigacion Efectuada en los Cantones de Puriscal y Coronado del 3 al 20 de Junio de 1985," Report, CELADE.

Palloni, A. and Heligman, L. (1985), "Re-estimation of Structural Parameters to Obtain Estimates of Mortality in Developing Countries," *Population Bulletin of the United Nations*, 18, 10–33.

Palloni, A., Massagli, M., and Marcotte, J. (1984), "Estimating adult mortality with maternal orphanhood data: Analysis of sensitivity of the techniques," *Population Studies*, 38, 255–279.

Palloni, A. and Pinto, G. (2004), "One hundred years of mortality in Latin America and the Caribbean: the fragile path from hunger to longevity," Presented at the annual meeting of the Population Association of America, Boston, Massachusetts, April 1-3, 2004, pp. 1–40.

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

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

Preston, S. H. and Coale, A. J. (1982), "Age Structure, Growth, Attrition and Accession: A New Synthesis," *Population Index*, 48, 217–259.

Preston, S. H., Elo, I. T., Rosenwaike, I., and Hill, M. (1996), "African-American mortality at older ages: results of a matching study," *Demography*, 33, 193–209.

Preston, S. H., Elo, T. I., Hill, M. E., and Rosenwaike, I. (2003), *The Demography of African Americans, 1930-1990*, Boston: Kluwer Academic Publishers.

Preston, S. H., Heuveline, P., and Guillot, M. (2001), *Demography: measuring and modeling population processes*, Malden: Blackwell Publishing.

Preston, S. H. and Lahiri, S. (1991), "A short-cut method for estimating death registration completeness in destabilized populations," *Math Popul Stud*, 3, 39–51.

Raymer, J. and Rogers, A. (2006), "Applying model migration schedules to represent age-specific migration flows," Tech. rep., Research Program on Population Processes, Institute of Behavioral Sciences, University of Colorado at Boulder.

Rosenwaike, I. (1987), "Mortality Differentials among Persons Born in Cuba, Mexico and Puerto Rico Residing in the United States, 1979-81," *American Journal of Public Health*, 77, 603–606.

Rosenwaike, I. and Preston, S. H. (1984), "Age Overstatement and Puerto Rican Longevity," *Human Biology*, 56, 503–525.

Sanchez-Albornoz, N. (1976), *The Population of Latin America*, Duke University Press.

Secretaria de Hacienda, Republica de Cuba (1976), "Reporte de la Secretaria de Hacienda: 1902-1930," Tech. rep., Secretaria de Hacienda.

Shryock, H. S., Siegel, J. S., and Associates (1976), *The methods and materials of Demography*, Academic Press, New York.

Somoza, J. L. (1971), *La Mortalidad en la Argentina Entre 1869 y 1960*, Buenos Aires: Instituto Torcuato di Tella, Editorial del Instituto.

Spencer, G. (1984), "Mortality among the Elderly Spanish Surnamed Population in the Medicare Files: 1968 to 1979," .

Sprague, T. B. (1880), "Explanation of a new formula for interpolation," *Journal of the Institute of Actuaries and Assurance Magazine*, 22, 270–285.

Timaeus, I., Chackiel, J., and Ruzicka, L. (1996), *Adult Mortality in Latin America*, Clarendon Press.

Timaeus, I. M. (1991), "Measurement of adult mortality in less developed countries: a comparative review," *Population Index*, 57, 552–68.

Trussell, T. J. (1975), "A Re-estimation of the Multiplying Factors for the Brass Technique for Determining Childhood Survivorship Rates," *Population Studies*, 29, 97–107.

UN IGME (2017), "Levels & Trends in Child Mortality: Report 2017, Estimates Developed by the UN Inter-agency Group for Child Mortality Estimation," Tech. rep., United Nations Children's Fund: United Nations-Inter-Agency-Group for Child Mortality Estimation.

United Nations (1982), *Model Life Tables for Developing Countries*, New York: Department of International Economic and Social Affairs United Nations.

United-Nations (1983), *Indirect techniques for demographic estimation*, New York: United Nations.

— (1988), "MORTPAK-LITE: The United Nations software package for mortality measurement," Report, United Nations.