This paper is presented at the 2018 annual meeting of the Population Association of America (session Modeling mortality and population health), 26-28 April 2018, Denver, CO, USA.

Establishing a Developing Countries Mortality Database (DCMD) on the Empirical Basis of Child, Adult, and Old-age Mortality

Nan Li[1], Hong Mi[2], Patrick Gerland[1], Cheng Li[3], and Lingxue Sun[3]

1. Population Division, Department of Economic and Social Affairs, United Nations, 2 UN Plaza, 233 East 44th Street, Room DC2-1938, 1934, New York, NY 10017, USA (Email: li32@un.org, gerland@un.org). The views expressed in this paper are those of the author and do not necessarily reflect those of the United Nations.

2. Corresponding author， School of Public Affairs, Zhejiang University, Room 265#, Mengmin'wei Building, Zijin'gang Campus, Hangzhou, Zhejiang, P. R. China. (E-mail: spsswork@163.com). The work on this paper was supported by the Nature Science Foundation of China (NSFC) project (NO.71490732), NSFC project (NO.71490733) , Zhejiang Social Science Planning Project Key Program(NO.17NDJC029Z), and Nature Science Foundation of Zhejiang project(NO.LZ13G030001)

3. School of Public Affairs, Zhejiang University, Mengmin'wei Building, Zijin'gang Campus, Hangzhou, Zhejiang, P. R. China. (E-mail: bigbosscool1990@aliyun.com, sunlx_2011@163.com).

Keywords: Old-age mortality, Life table, Database, Developing countries

Abstract Life tables of developing countries are often estimated using child (or child and adult mortality), leaving old-age mortality unchecked and unreliable. However, the number of developing countries' old-age deaths is the largest among various age and development groups. This irony requires establishing a developing-countries mortality database (DCMD) that utilizes observed data on old-age mortality. We introduce the DCMD method that includes estimating old-age mortality, and uses child, adult, and old-age mortality to produce life tables. A validation of using the DCMD method to the data of the Human Mortality Database shows that the errors of estimating old-age mortality are reduced for more than 70% countries compared to using only child and adult mortality. Preliminary estimates of the DCMD old-age mortality, as are shown in this paper, differ remarkably from the predictions of existing model life tables using only child and adult mortality. More results are available at **www.lifetables.org**.

1 Introduction

Empirical data used in estimating life tables are collected from death registration, census, and sample survey (United Nations Statistics Division (UNSD), 2004). For developed countries, death registration and census provide necessary data for calculating life tables. For most developing countries, however, death registrations are either unavailable or unreliable, estimating life tables rely mainly on surveys, and occasionally also on census when it collects information on deaths in the household in the previous 12 months (UNSD, 2008).

Typical sample surveys often collect information only from a small portion of the population. For example, since the 1990s Demographic and Health Surveys sample on average 15,000 households[1]; and Living Standard

[1] Measure DHS (2017). What do we do: Survey search. http://dhsprogram.com/What-We-Do/survey-search.cfm?pgType=main&SrvyTp=type, Accessed 8 Aug. 2017.

Measurement Surveys on average sample 4,630 households[2]. Subsequently, they cannot produce reliable life tables based on household deaths due to large sampling errors for a 12 or 24 recall period. This is because death rates at some ages could be very low, and hence require a large population to be estimated reliably (e.g., Bangladesh Maternal Mortality Survey sampled 104,323 households in 2001 (NIPORT et al. 2003) and 175,000 in 2010 (NIPORT et al. 2012); the Ghana 2007 Maternal Health Survey surveyed 227,715 households (GSS et al. 2009)). Nonetheless, sample surveys could provide reliable indicators of mortality for certain age groups when deaths are not rare or when the age group is wide enough.

The most commonly sampled mortality indicator is child mortality, which is the probability of dying between birth and age 5, and is often denoted as ${}_5q_0$. The United Nations Children's Fund (UNICEF) as part of the United Nations Inter-agency Group for Child Mortality Estimation (IGME) has been regularly collecting, analyzing, and publishing child mortality for most of the countries back to the 1970s or earlier (see United Nations Children's Fund, 2017; http://www.childmortality.org). Based on the values of ${}_5q_0$ and in the absence of mortality shocks or epidemics like HIV/AIDS mortality affecting disproportionately young adults (before the availability of Antiretroviral Therapy), life tables can be estimated using one-input model life tables (MLT) such as the Coale-Demeny MLT (Coale, Demeny and Vaughan, 1983), or the United Nations MLT (United Nations Population Division, 1982).

Starting from the 1990s, surveys such as the Demographic and Health Surveys (DHS, http://www.measuredhs.com/) have been collecting sibling histories (Timæus, 2013), leading to estimations of adult mortality, which is the probability of dying between age 15 and 60 years, and is denoted as ${}_{45}q_{15}$. Combining data of surveys and other sources (including additional time covariates for countries and periods with missing or deficient data), Wang and colleagues (2012) at the Institute for Heath Metrics and Evaluation (IHME) estimated adult mortality for 187 countries from 1970 to 2010. Utilizing the values of ${}_5q_0$ and ${}_{45}q_{15}$, life tables can be estimated using the two-input-parameter MLT such as the modified logit life table system (Murray and others, 2003) or the log-quadratic model life table with two input parameters (Wilmoth and others, 2012).

In principle, the two-input-parameter MLT describes relationships between old-age and young-age mortality. These relationships are found mainly from the data of developed countries. For developing countries, whether or not applying these relationships to estimate life tables is reliable, in fact, was unknown. Nonetheless, these relationships have been widely applied to estimate life tables for developing countries, because they have been almost the only means.

Using population data of two successive censuses, old-age mortality, or the probability of dying between age 60 and 75 years (${}_{15}q_{60}$), could be estimated for the period between the two censuses, because at old ages migration is negligible compared to deaths (Li and Gerland, 2013) and potential age-reporting errors can be adjusted. For most of the countries, census data on population by age and sex can be found from the United Nations Demographic Yearbook (e.g., UNSD, 2013a). Moreover, a three-input-parameter MLT is proposed by Li (2014), which can utilize ${}_5q_0$, ${}_{45}q_{15}$ and ${}_{15}q_{60}$ to compute life tables.

Three-input-parameter MLT include adjustments to the relationships between old-age and young-age mortality described by the two-input-parameter MLT; and these adjustments are made according to the ${}_{15}q_{60}$ estimated from observed data. Furthermore, in terms of applying the relationships obtained from the data of developed countries to estimate life tables for developing countries, the adjustments are made according to data observed from developing countries.

Life-table databases have been established for developed countries (e.g., Human Mortality Database (HMD), 2016) and effectively used for various purposes. For developing countries of which the deaths counted 78% that of the world in 2010-2015 (United Nations Population Division (UNPD), 2017), however, reliable life tables can hardly

[2] World Bank (2017). LSMS Data: LSMS Information Table. http://go.worldbank.org/XP4GRPITV0, Accessed 8 Aug. 2017.

be found. Indirect estimates of life tables have been produced by the UNPD (2017) and IHME (Wang and colleagues, 2012) for developing countries, using empirical data on child mortality (${}_5q_0$) and adult mortality (${}_{45}q_{15}$). But more than half of all deaths already occurred at age 60 and higher in developing countries in 2010-2015, and this extra information on old-age mortality is not taken into account when only one or two input parameters are used. Thus, estimating old-age mortality (${}_{15}q_{60}$), and using it together with estimates of ${}_5q_0$ and ${}_{45}q_{15}$ such as those provided by UNICEF and IHME or other similar processes, to calculate life tables as the Developing-countries Mortality Database (DCMD), is an important task. The DCMD method is proposed to fulfill this task, which includes (1) estimating old-age mortality ${}_{15}q_{60}$, and (2) the three-input-parameter MLT that utilizes ${}_5q_0$, ${}_{45}q_{15}$, and ${}_{15}q_{60}$ to calculate life tables.

A basic validation exercise involves simply applying the DCMD method to the data based on old-age population from the HMD after 1950. In this case the errors of fitting old-age mortality using model life tables are reduced for more than 70% of all the countries in HMD, compared to using only child and adult mortality (Li, Mi, and Gerland, 2017). These findings indicate that, even for developed countries, the existing relationships in model life tables between old-age and young-age mortality within this group of countries can be improved significantly by taking into account observed data on old ages. Accordingly, applying these relationships to estimate life tables for developing countries are unlikely to be reliable, and adjustments seem to be necessary to take into account old-age mortality data upon availability.

Applying the DCMD method to census data of old-age population of the 151 developing countries that have about 100 thousand or more populations in 2015 (UNPD, 2017), we obtained life tables for 122 countries as the preliminary results of the application of the DCMD. The old-age mortality of these life tables differ remarkably from that of using only child and adult mortality. These remarkable differences indicate that, applying the model life tables found from developed countries using only child and adult mortality to developing countries to estimate old-age mortality is unreliable, and adjustments are indeed necessary. More results are available at **www.lifetables.org.**

2 The DCMD method

The DCMD method involves two components, first estimating old-age mortality and secondly the three-input-parameter MLT including not only child and adult mortality, but also old-age mortality summary estimates.

2.1 Estimating old-age mortality

In this paper, estimating old-age mortality is carried out by the Census Method that includes two models, and the two-input-parameter MLT. In the Census Method, the first model uses the variable-r model (Bennett and Horiuchi,1981; Li and Gerland, 2013), and the second is the survival model at older ages. Although the two-input-parameter MLT seems unreliable, it provides a richer experience of historical and contemporary human mortality patterns than typically included in older model life table systems (e.g., Coale, Demeny and Vaughan, 1983). Thus, we use the two-parameters MLT to estimate old-age mortality as references with weights to be determined according to specific situations. In the future, estimating old-age mortality should utilize data on deaths by age and sex from vital registrations, surveys and censuses.

2.1.1 The variable-r model

The variable-r model (Bennett and Horiuchi, 1981) assumes zero migration and evenly distributed enumeration errors over age. Let $p(x, t)$ be the observed number of population in age group $[x,x+5)$ enumerated from a census conducted at time t, where x=60, 65, 70. The growth rates at age x are computed as:

$$r(x) = \log[\frac{p(x,t_2)}{p(x,t_1)}]/(t_2 - t_1), \quad x = 60,65,70, \qquad (1)$$

Where t_1 and t_2 represent the date of the first and second census, respectively. And the accumulated growth rates are:

$$
\begin{aligned}
s(60) &= 2.5r(60),\\
s(65) &= 5r(60) + 2.5r(65),\\
s(70) &= 5[r(60) + r(65)] + 2.5r(70).
\end{aligned} \tag{2}
$$

Further, the middle-point population in age group $[x,x+5)$, $N(x)$, are estimated as:

$$N(x) = \sqrt{p(x,t_1)p(x,t_2)}, \quad x = 60,65,70. \tag{3}$$

Furthermore, the person-years lived in 5-year age group $[x,x+5)$, L_x, in the underlying stationary population, are obtained as (Bennett And Horiuchi, 1981):

$$L_x = N(x)\exp[s(x)], x = 60,65,70. \tag{4}$$

In developing countries, however, the errors in enumerating population often occur unevenly across age. A typical example is age heaping. When such errors are severe, the L_x resulted from (4), would show implausible patterns of increasing with age, which cannot occur in a stationary population. When such implausible situations occur, adjusting L_x is necessary. Li and Gerland (2013) proposed such an adjustment as is shown in the appendix, which provides the adjusted $\hat{L}_x$.

After adjusting the age-reporting errors, the number of survivors at age x, l_x, can be estimated using nonlinear optimization and a Gompertz model (Li and Gerland, 2013), or it can be estimated locally linearly as below:

$$
\begin{aligned}
l_{65} &= \frac{\hat{L}_{60} + \hat{L}_{65}}{2.5} \frac{\hat{L}_{65}}{(\hat{L}_{60} + 2\hat{L}_{65} + \hat{L}_{70})},\\
l_{70} &= \frac{\hat{L}_{65} + \hat{L}_{70}}{2.5} \frac{\hat{L}_{65}}{(\hat{L}_{60} + 2\hat{L}_{65} + \hat{L}_{70})},\\
l_{60} &= \frac{\hat{L}_{60}}{2.5} - l_{65},\\
l_{75} &= \frac{\hat{L}_{70}}{2.5} - l_{70}.
\end{aligned} \tag{5}
$$

In (5), the $\frac{\hat{L}_{60} + \hat{L}_{65}}{2.5}$ and $\frac{\hat{L}_{65} + \hat{L}_{70}}{2.5}$ are the first-step estimates of l_{65} and l_{70}, which are linear interpolations between $\hat{L}_{60}$, $\hat{L}_{65}$ and $\hat{L}_{70}$. The $\frac{\hat{L}_{65}}{(\hat{L}_{60} + 2\hat{L}_{65} + \hat{L}_{70})}$ is an adjustment that makes $2.5 \cdot (l_{60} + l_{65}) = \hat{L}_{65}$. The last two lines in (5) are linear formulas of calculating $\hat{L}_{60}$ and $\hat{L}_{70}$.

Finally, after estimating l_x, ${}_{15}q_{60}$ is obtained as:

$${}_{15}q_{60}^{(v)} = 1 - \frac{l_{75}}{l_{60}}, \tag{6}$$

where superscript "*v*" stands for variable-r method.

2.1.2 The survival model at older-ages

When the period between the two successive censuses is closer to 5 or 10 years, the populations between the period of exactly 5 or 10 years can be estimated assuming over-time constant growth rates and using (1). Consequently, the 10-year survival ratio of the stationary population is estimated as:

$$S = \frac{L_{70}}{L_{60}} = \begin{cases} \frac{p(65-69,t_2)}{p(60-64,t_1)} \cdot \frac{p(70-74,t_2)}{p(65-69,t_1)}, & t_2 - t_1 = 5, \\ \frac{p(70-74,t_2)}{p(60-64,t_1)}, & t_2 - t_1 = 10. \end{cases} \quad (7)$$

Assuming that the over-age survival ratio is constant, the 1-year and 15-year survival ratios are therefore $S^{\frac{1}{10}}$ and $S^{\frac{15}{10}}$, respectively. Subsequently, the 15-year probability of death between age 60 and 75 can be estimated as:

$$q = 1 - S^{\frac{15}{10}}. \quad (8)$$

The assumption of constant over-age survival ratio can be adjusted using the United Nations general model life table (UNPD, 1982), which leads to a more accurate estimate of old-age mortality as

$$_{15}q_{60}^{(s)} = \begin{cases} q \cdot (1.021 - 0.0002 \cdot q + 0.0002 \cdot q^2), & R^2 = 0.999, \quad \textit{female}, \\ q \cdot (1.0153 - 0.0003 \cdot q + 0.0002 \cdot q^2), & R^2 = 0.999, \quad \textit{male}, \end{cases} \quad (9)$$

where superscript "*s*" represents the survival model.

2.1.3 The two-input-parameter MLT

The two-input-parameter MLT, or the flexible two-dimensional log-quadratic model life table (Wilmoth et al, 2012), is expressed as

$$\log(m_x) = a_x + b_x \cdot \log({}_5q_0) + c_x \cdot [\log({}_5q_0)]^2 + v_x \cdot k, \quad (10)$$

where m_x stands for the five-year age-specific death rates with *x=0,1,5,10,…*; coefficient vectors a_x, b_x, c_x, and v_x are obtained from fitting mortality data of the Human Mortality Database; and parameter *k* is flexible, which can be solved to fit an additional ${}_{45}q_{15}$. Obviously, given ${}_5q_0$ and ${}_{45}q_{15}$ the two-input-parameter MLT produces a life table, which includes an old-age mortality. This summary old-age mortality estimate is denoted as ${}_{15}q_{60}^{(t)}$, where superscript "t" stands for the two-input-parameter MLT. Although ${}_{15}q_{60}^{(t)}$ is not estimated from observed data, it is based on the relationships between old-age and child and adult mortality found from developed countries, and is therefore worth using.

Based on the three estimates of old-age mortality, the DCMD method estimates old-age mortality as:

$$_{15}\hat{q}_{60} = w_1 \cdot {}_{15}q_{60}^{(v)} + w_2 \cdot {}_{15}q_{60}^{(s)} + w_3 \cdot {}_{15}q_{60}^{(t)}, \quad (11)$$

where weights w_1, w_2, and w_3 are non-negative and obey:

$$w_1 + w_2 + w_3 = 1. \tag{12}$$

The default values of each weight are 1/3, and can be changed for special situations.

The errors in the Census Method estimates (${}_{15}q_{60}^{(v)}$ and ${}_{15}q_{60}^{(s)}$) are caused by the errors in census population, and would be small if the errors in census population are small. The errors in the two-input-parameter MLT estimate (${}_{15}q_{60}^{(t)}$), however, is caused not only by the errors in young-age mortality but also by applying developed countries' relationships to developing countries. Thus, the errors in ${}_{15}q_{60}^{(t)}$ may not be small even if the errors in young-age mortality are small. Consequently, the errors in the DCMD estimate (${}_{15}\hat{q}_{60}$) would also be small if the errors in census population were small. Besides, because ${}_{15}q_{60}^{(v)}$, ${}_{15}q_{60}^{(s)}$, and ${}_{15}q_{60}^{(t)}$ use different assumption and data, the errors in them should be independent each other, and therefore should cancel each other. Thus, the errors in the DCMD estimate (${}_{15}\hat{q}_{60}$) should be smaller than the errors in the two-input-parameter MLT estimate (${}_{15}q_{60}^{(t)}$), not only because it would be small when the errors in census population are small, but also because of the cancellations of the errors in the three individual estimates.

In the future, the DCMD should include estimates of old-age mortality from sources other than census population at old ages. Consequently, the estimates from the DCMD will be more reliable because the cancellations of the errors in individual estimates.

2.2 The three-input-parameter model life table

How to utilize the estimated old-age mortality (${}_{15}\hat{q}_{60}$)? A simple answer (Li, 2014) can be found by following the logic of the Logit transformation: $\log[{}_x\hat{q}_0/(1-{}_x\hat{q}_0)] = \alpha + \beta\log[{}_xq_0/(1-{}_xq_0)]$, in which the standard ${}_xq_0$ is naturally that of the two-input-parameter MLT, and level α and pattern β can be chosen to fit some function of observed probability of death (${}_x\hat{q}_0$). When there is only ${}_{15}\hat{q}_{60}$ to fit, it is customary to set $\beta = 1$ and solve α to fit ${}_{15}\hat{q}_{60}$ (see Preston, Heuveline and Guillot, 2001; p.200). The rationale for using the Logit transformation is that $\log[{}_xq_0/(1-{}_xq_0)]$ would be close to linear at all the ages. It is worth noting that, at old ages, $\log(\hat{m}_x)$ would be close to linear according to the Gompertz law. Thus, at old ages, the linear relationship of the Logit transformation can be simplified as:

$$\log(\hat{m}_x) = \alpha + \log(m_x). \tag{13}$$

Because

$${}_{15}q_{60} \approx 1 - \exp[-5\cdot(m_{60} + m_{65} + m_{70})], \tag{14}$$

α is solved by inserting (13) to (14):

$$\alpha \approx \log[\frac{\log(1-{}_{15}\hat{q}_{60})}{\log(1-{}_{15}q_{60})}] \tag{15}$$

where ${}_{15}q_{60}$ is the old-age mortality of the two-input-parameter MLT. Subsequently, (10) is augmented to the three-input-parameter MLT:

$$\log(m_x) = \hat{a}_x + b_x \cdot \log({}_5q_0) + c_x \cdot [\log({}_5q_0)]^2 + v_x \cdot k, \tag{16}$$

$$\hat{a}_x = \begin{cases} a_x, & x < 60, \\ a_x + \log[\dfrac{\log(1-{}_{15}\hat{q}_{60})}{\log(1-{}_{15}q_{60})}], & x \geq 60, \end{cases} \tag{17}$$

When ${}_{15}\hat{q}_{60}$ differs remarkably from ${}_{15}q_{60}$, (16) will show discontinuous change in m_x between age group 55-59 and 60-64, which can be smoothed as following. Let

$$d = m_{60} - \sqrt{m_{55} \cdot m_{65}}\ . \tag{18}$$

Then, the smoothed death rates by age, or the DCMD death rates by age are:

$$\hat{m}_x = \begin{cases} m_x, & x < 60, \\ m_x - d, & x = 60, \\ m_x, & x = 65, \\ m_x + d, & x \geq 70, \end{cases} \tag{19}$$

where m_x are given in (18). Using the three-input-parameter MLT described by (16)-(19), the child, adult, and old-age mortality will equal exactly their input values.

The DCMD method is tested using the data of the Human Mortality Database (HMD). Applying the DCMD method to the old-age population data of HMD after 1950 lead to a reduction of the errors (root-mean-squared error, RMSE) of fitting old-age mortality using model life tables for more than 70% of all the countries (Li, Mi, and Gerland, 2017), compared to using only child and adult mortality. In figure 1, the location on the horizontal axis of each diamond describes the RMSE of using the two-input-parameter MLT to predict the old-age mortality of a country's male or female population over the six 10-year periods from 1950 through 2010; and the corresponding vertical-axis location indicates the RMSE of using DCMD. In figure 1, more than 70% diamonds are below the equal line, indicating that the DCMD method improved estimation of old-age mortality for more than 70% of the populations of HMD. Moreover, using DCMD reduced the RMSE of the two-input-parameter MLT by 19%.

Figure 1. Root-mean-squared errors in predicting old-age mortality for the 74 male and female populations by sex in HMD

These findings indicate that, even for developed countries, the relationships between old-age and young-age mortality is not well captured using a two-input parameter model life table system with only child and adult mortality as inputs. Prediction of old-mortality can be improved significantly by using observed data on old ages. Accordingly, applying these relationships to estimate life tables for developing countries are unlikely to be reliable, and adjustments seem to be necessary.

3 The preliminary results from the DCMD method

In ideal situation, for one country, if N censuses were conducted after 1970 (before which the IHME estimates of adult mortality are unavailable), (N-1) life tables for each gender could be estimated by the DCMD method for each sex. In reality, however, some countries may have two successive censuses that are separated by too many years (e.g, more than 15 years), some may have two inconsistent successive censuses (of which the survivors in the later one is bigger than the population in the earlier one), and some do not have the open age group higher than 75+ for each census. As a result, the number of estimated life tables will be smaller than that of the ideal situation. We applied the DCMD method to the data on old-age population of all the 534 censuses conducted after 1970 in all the 151 developing countries that have about 100 thousand or more populations in 2015 (UNPD, 2017). We obtained life tables by sex for 122 countries as preliminary results from the application of the DCMD method to this set of countries based upon census data availability and the before-mentioned conditions.

We use the root-mean-squared deviation (RMSD) to indicate the difference between the values of old-age mortality from the DCMD method (${}_{15}\hat{q}_{60}$) and the two-input-parameter MLT (${}_{15}q_{60}^{(t)}$):

$$RMSD = \sqrt{\sum_{i=1}^{n} ({}_{15}\hat{q}_{60}(i) - {}_{15}q_{60}^{(t)}(i))^2 / n}, \qquad (22)$$

where i stands for the i th estimate and n represents the total number of estimates. The value of RMSD is about 0.09, which can be interpreted as the average difference between ${}_{15}\hat{q}_{60}$ and ${}_{15}q_{60}^{(t)}$. Compared to the mean of ${}_{15}\hat{q}_{60}$ and ${}_{15}q_{60}^{(t)}$, 0.42, the relative difference between ${}_{15}\hat{q}_{60}$ and ${}_{15}q_{60}^{(t)}$ is about 21%. Recalling that one-third of ${}_{15}\hat{q}_{60}$ is ${}_{15}q_{60}^{(t)}$, the difference made by using the DCMD method is remarkable. These remarkable differences indicate that, applying the relationships between old-age mortality and young-age mortality observed from developed countries to developing countries is unreliable, and adjustments using the DCMD method are necessary

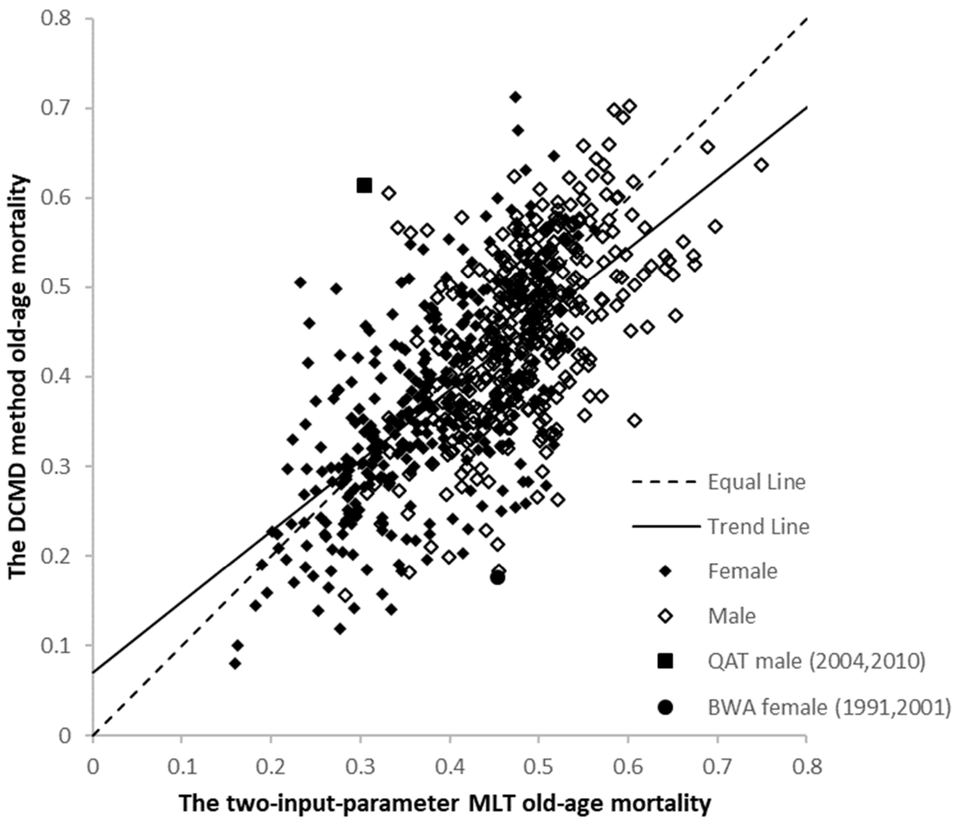

Figure 2. Estimates from the DCMD method and two-input-parameter MLT for old-age mortality

More details about the differences between the old-age mortality estimates from the DCMD method (${}_{15}\hat{q}_{60}$) and two-input-parameter MLT (${}_{15}q_{60}^{(t)}$) are shown in figure 2, in which each diamond describes the old-age mortality of a country's male or female population for a certain period between two successive censuses. The horizontal-axis location of a diamond represents the old-age mortality estimated by the two-input-parameter MLT; and the vertical-axis location indicates the corresponding estimate from the DCMD method. Taking into account the findings from section 2.2 showing that for HMD countries the DCMD method estimates of old age mortality are more accurate than those from the two-input-parameter MLT, Figure 2 indicates that at higher mortality levels for developing countries, the two-input-parameter MLT tends to overestimate old-age mortality (equal line higher than trend line); while at lower mortality levels, the two-input-parameter MLT is more likely to underestimate old-age mortality (equal line lower than trend line). It is also clear that Botswana (BWA) female (1991,2001) and Qatar (QAT) male (2004,2010) female are the two extreme outliers (which differ the farthest from the equal line), of which the life tables may be more likely abnormal. The values of death rates by age of the two countries are displayed in figure 3. Without the smoothing procedure described in (18)-(19), the changes between age group 55-59 (the last on of adult ages) and 60-64 (the first one of old ages) are indeed discontinued, and are unlikely to be observed in large populations. After using the smoothing procedure described in (18)-(19), the pattern of increase by age of the death rate becomes more plausible.

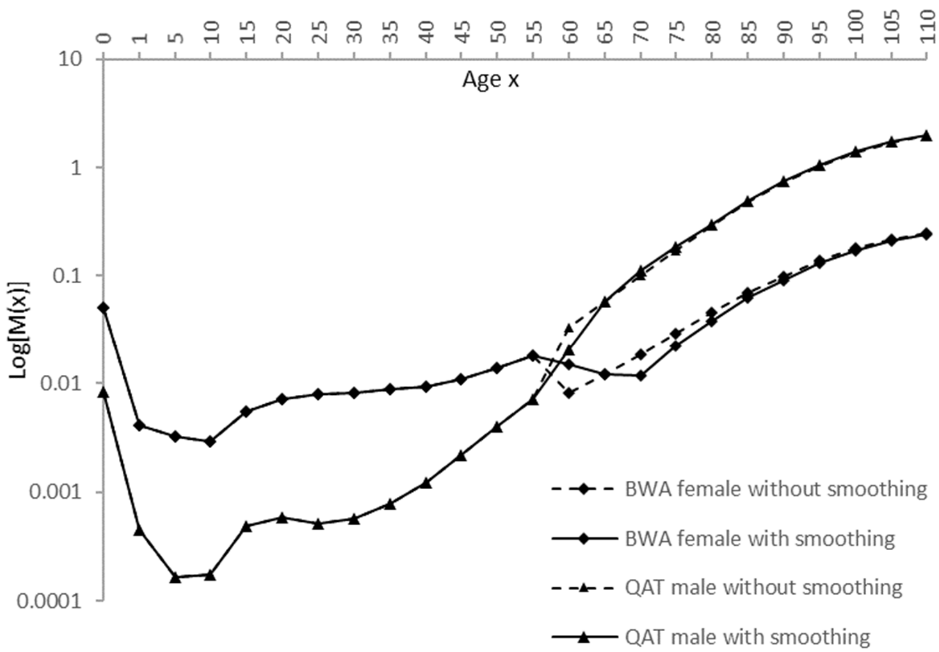

Figure 3. Age patterns of death rate, Botswana (BWA) female (1991,2001) and Qatar (QAT) male (2004,2010), the DCMD method estimates

Since the DCMD method provides not only old-age mortality but also life tables, we can discuss the differences between the life expectancy at birth (e(0)) estimated by the two-input-parameter MLT and the DCMD method. In figure 4, each diamond describes the e(0) of a country's male or female population. The horizontal-axis location of a diamond represents the e(0) estimated by the two-input-parameter MLT; and the vertical-axis location indicates the corresponding estimate from the DCMD method. At first glance the e(0) estimated by the two-input-parameter MLT and DCMD method are similar (figure 4). Indeed, the RMSD, or the average difference between the e(0) estimated by the two-input-parameter MLT and DCMD method, is 1.8 years, which is small compared to the mean of e(0), 64.1 years. Moreover, the trend line is also quite close to the equal line. Nonetheless, the average difference between the e(0) of developing countries in 2010-2015 was only about 0.2 years (UNPD, 2017). Although the 0.2 years difference in e(0) is even smaller, it affects the ranks of Human Development Index (HDI, http://hdr.undp.org/) for many countries, because e(0) has a weight of 1/3 in computing the HDI with the other two index components. If the error of estimating e(0) is about 1.8 years because of not using observed data on old ages, what would be the mistake of ranking countries' developing levels by HDI, behind which the average difference of e(0) is merely 0.2 years? In other words, the difference of e(0) produced using the DCMD method may not look large, its impact on ranking the development levels, and on many other aspects as well, could still be significant.

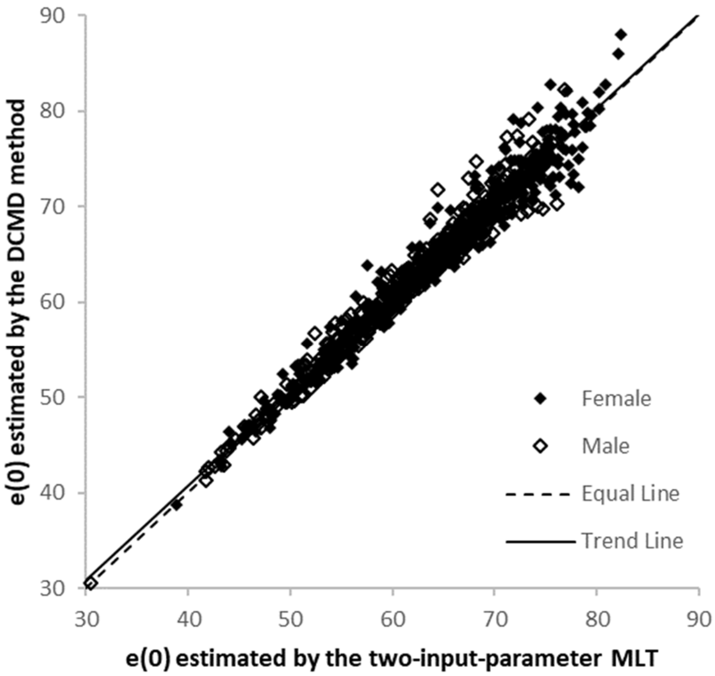

Figure 4. The DCMD method and two-input-parameter MLT estimates of life expectancy at birth, e(0), by sex

In 2013, governments in 102 countries reported that population aging was a “major concern,” (UNPD, 2013). To study the issues of population aging, such as providing pensions and medical services for the elderly, life expectancy at age 60 (e(60)) is a piece essential information. In figure 5, each diamond indicates the e(60) of a country’s male or female population. The horizontal-axis location of a diamond represents the e(0) estimated by the two-input-parameter MLT; and the vertical-axis location indicates the corresponding estimate from the DCMD method.

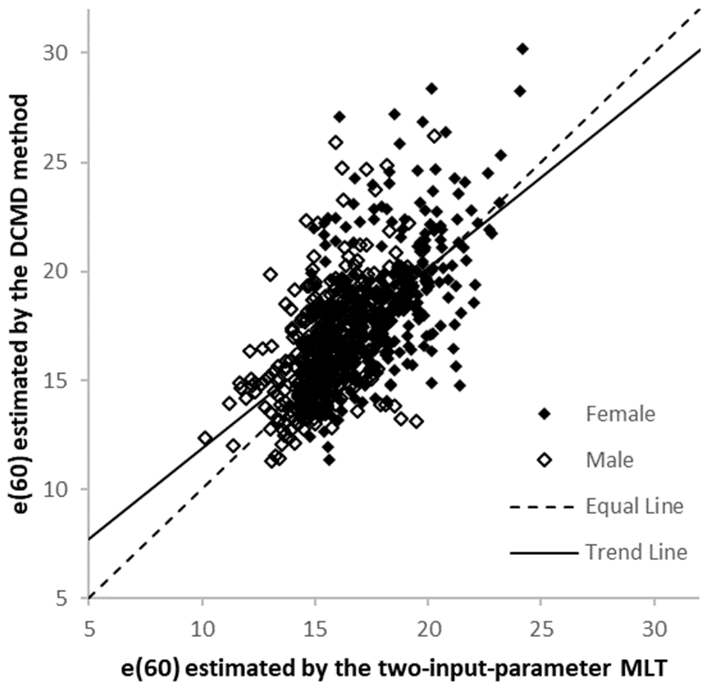

Figure 5. The DCMD method and two-input-parameter MLT estimates of life expectancy at age 60, e(60), by sex

It can be seen, in figure 5, that the two-input-parameter MLT tends to overestimate life expectancy at age 60 (i.e., the model underestimates mortality at these older ages), and vice versa. Since at a higher level of life expectancy at age 60 is equivalent to at a lower level of old-age morality, the pattern in figure 5 is consistent with that in figure 2. It is intuitive that the difference made by using old-age data, or the DCMD method, is remarkable. The RMSD, which indicates the average difference between the e(60) estimated by the two-input-parameter MLT and DCMD method, is about 2.5 years. Compared to the mean value of e(60), 17.0 years, the relative difference is about 15%. If a pension program, for example, made a 15% error on estimating e(60), how could it achieve its target?

4 Summary

In 2010-2015, the deaths at age 60 and older already reached 60% of all deaths worldwide (UNPD, 2017). Compared to the numbers of deaths at child and adult ages, the number of deaths at old ages is the biggest and, ironically, also the least reliable. This is because, for most developing countries, the numbers of old-age deaths are not estimated using observed data. They are based on the relationships between old-age and young-age mortality found mainly from developed countries.

At old ages, migrants are rare comparing to deaths. Thus, census data on population by age and sex could be used to estimate old-age mortality; and such data are available for almost all the countries of the world. Furthermore, in recent years new methodological developments have been made to extend two-input-parameter MLT to utilize old-age mortality, and to use census population to estimate old-age mortality. These methodological developments provided a method for establishing a developing-countries mortality database (DCMD) (see also Li, Mi, and Gerland, 2017). Using the DCMD method to the data on old-age population of HMD after 1950, the errors of fitting old-age mortality are reduced for more than 70% of all the countries in HMD, compared to using only child and adult mortality. These findings indicate that, even for developed countries, the existing relationships between old-age and young-age mortality within this group of countries can be improved significantly by taking into account observed data on old ages. Accordingly, applying these relationships to estimate life tables for developing countries are unlikely to be reliable, and adjustments seem to be necessary to take into account old-age mortality data upon availability. In this paper, we applied the DCMD method to all the censuses after 1970 conducted in the 151 developing countries that have about 100 thousand or more populations in 2010-2015 (UNPD, 2017). We obtained life tables for 122 countries as the preliminary results of DCMD. On average, using the DCMD method for the old-age mortality estimates differ by about 21% from those from the two-input-parameter MLT which only uses child and adult mortality. These remarkable differences indicate that, applying the relationships found from developed countries to developing countries is unreliable, and adjustments using the DCMD method are necessary. More results are available at **www.lifetables.org.**

In comparison to the use of only child and adult mortality to estimate life expectancy at birth (e(0)), the use instead of the DCMD method would make a 1.8-year difference for e(0) on average. Considering that the average difference between the e(0) of developing countries in 2010-2015 was only 0.2 years, and that the 0.2-year difference largely determines the ranks of Human Development Index, the 1.8-year difference on estimating e(0) is important.

Population aging is an issue causing concerns in many countries. To study this issue, life expectancy at age 60 (e(60)) is an essential piece of information. Compared to using only child and adult mortality to estimate e(60), the use of the DCMD method would make a 15% difference on estimating e(60). More specifically, the two-input-parameter MLT tends to overestimate life expectancy at age 60 (i.e., too optimistic estimates of remaining years of life), and vice versa. If a pension program, for example, made a 15% error on estimating e(60), there would be serious consequences.

The preliminary results of the DCMD method include life tables that should be more reliable than those using only child and adult mortality, and therefore could improve the basis of various demographic studies for developing countries. For some developing countries, old-age mortality could also be estimated from death registration (e.g., Algeria), censuses (e.g., China), and surveys (e.g., Indonesia). Furthermore, some international surveys (e.g., CHARLS www.g2aging.org) are also starting to collect data that could be used to estimate old-age mortality. In the future, the DCMD will include these data and, as indicated by the cancellations of errors, will provide more reliable estimates of mortality for developing countries.

References

Bennett, N.G.and S. Horiuchi.1981. Estimating the Completeness of Death Registration in a Closed Population.*Population Index* 47(2):207-221.

Coale, A. J., P. G.Demeny and B. Vaughan (1983). *Regional Model Life Tables and Stable Populations* (2nd edition). New York: Academic Press.

Ghana Statistical Service (GSS), Ghana Health Service (GHS), and Macro International (2009). *Ghana Maternal Health Survey 2007*. Calverton, Maryland, USA: GSS, GHS, and Macro International.

Human Mortality Database, 2016. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de.

Li, N., 2014. Estimating Life Tables for Developing Countries. Technical paper of the United Nations Population Devision 2014/4. Available at http://www.un.org/en/development/desa/population/publications/pdf/technical/TP2014-4.pdf.

Li, N. and P. Gerland. 2013. Using census data to estimate old-age mortality for developing countries. Paper prepared for Session 17-05: Indirect methods of mortality and fertility estimation: new techniques for new realities, in *XXVII IUSSP International Population Conference*. Busan, Korea. http://iussp.org/en/event/17/programme/paper/1959.

Li, N., H. Mi, and P. Gerland. 2017. Using Child, Adult, and Old-age Mortality to Establish a Developing Countries Mortality Database (DCMD). In *Proceedings of the 17th Applied Stochastic Models and Data Analysis International Conference with Demographics Workshop (*ASMDA2017, *p645-655)*. Edited by Christos H Skiadas (available at http://1drv.ms/b/s!ApL_0Av0YGDLijd07zI0KBZJZakF).

Murray, C. J. L. and others (2003). Modified logit life table system: Principles, empirical validation and application. *Population Studies,* Vol. 57, No. 2, pp. 165-182.

National Institute of Population Research and Training (NIPORT), ORC Macro, Johns Hopkins University and ICDDR,B. (2003). *Bangladesh Maternal Health Services and Maternal Mortality Survey 2001*. Dhaka, Bangladesh and Calverton, Maryland (USA): NIPORT, ORC Macro, Johns Hopkins University, and ICDDR,B.

National Institute of Population Research and Training (NIPORT), MEASURE Evaluation, and icddr,b. (2012). *Bangladesh Maternal Mortality and Health Care Survey 2010*. Dhaka, Bangladesh: NIPORT, MEASURE Evaluation, and icddr,b.

Preston, S. H., P. Heuveline and M. Guillot (2001). *Demography, Measuring and Modeling Population Process*. Oxford, United Kingdom: Blackwell Publishers Ltd.

Timæus I.M. 2013. Estimation of adult mortality from sibling histories. In Moultrie T.A., R.E. Dorrington, A.G. Hill, K. Hill, I.M. Timæus and B. Zaba (eds). *Tools for Demographic Estimation*. Paris: International Union for the Scientific Study of Population. http://demographicestimation.iussp.org/content/sibling-histories

United Nations Population Division. 2017. *World Population Prospects: The 2017 Revision*. United Nations, New York.

United Nations Population Division. 2013. World Population Policies Database. New York: United Nations.
United NationsPopulation Division. 1982. *Model life tables for developing countries*. New York: United Nations.

United Nations Statistics Division. (2004). *Handbook on the Collection of Fertility and Mortality Data*. Studies in Methods, Series F. No. 92. New York, NY: United Nations Statistics Division. 2008.*Principles and Recommendations for Population and Housing Censuses, Revision 2,Statistical papers Series M,* Rev. 2 ed., New York, NY: United Nations. http://unstats.un.org/unsd/demographic/sources/census/census3.htm

United Nations Statistics Division. 2013a. *Demographic Yearbook 2012, Sixty-third issue*. New York: United Nations. https://unstats.un.org/unsd/demographic/products/dyb/2010_round_latest.htm

United Nations Statistics Division. 2013b. Availablehttp://unstats.un.org/unsd/demographic/sources/census/2010_PHC/censusclockmore.htm

Wang H., L. Dwyer-Lindgren, K.T. Lofgren, J.K. Rajaratnam, J.R. Marcus, A. Levin-Rector, C. Levitz, A.D. Lopez, C.J.L. Murray. 2012. Age-specific and sex-specific mortality in 187 countries, 1970–2010: a systematic analysis for the Global Burden of Disease Study 2010. *The Lancet*. 2012 Dec 13; 380: 2071–2094.

Wilmoth, J., S. Zureick, V. Canudas-Romo, V. Inoue, and C. Sawyer, 2012. A flexible two-dimensional mortality model for use in indirect estimation. *Population Studies,* (66). 1-28.

Appendix: Adjusting age reporting errors

It is hard to find a proper basis to adjust enumeration errors in a real population, which is affected by historical fertility, mortality and migration. But a stationary population is determined only by mortality. Thus, it is possible to find a proper basis to adjust age errors for stationary populations. According to the United Nations general model life table (United Nations Population Division, 1982), there is a common relationship between the survival ratios $S_{60} = \frac{L_{65}}{L_{60}}$ and $S_{65} = \frac{L_{70}}{L_{65}}$ among model life tables, which is

$$S_{65} = -0.29 + 1.27 \cdot S_{60}, \quad R^2 = 0.998. \qquad \text{(A.1)}$$

This relationship is called the model line. When the observed survival-ratio point, (S_{60}, S_{65}), is above the model line, or when the survival ratio is abnormally rising with age, the difference between the survival-ratio point and the model line is caused mainly by age heaping. Accordingly, assuming that the heaping ratio at age 60 equals to that at age 70, the adjustment is

$$\begin{aligned}
\hat{L}_{60} &= L_{60} - \frac{L_{60}}{L_{70}}\Delta, \\
\hat{L}_{65} &= L_{65} + \Delta, \\
\hat{L}_{70} &= L_{70} - \Delta,
\end{aligned} \tag{A.2}$$

where

$$\begin{aligned}
\Delta &= \frac{-B + \sqrt{B^2 - 4AC}}{2A}, \\
A &= b - a\frac{L_{60}}{L_{70}} - \frac{L_{60}}{L_{70}}, \\
B &= a(L_{60} - \frac{L_{60}}{L_{70}}L_{65}) + 2bL_{65} + L_{60} + \frac{L_{60}}{L_{70}}L_{70}, \\
C &= L_{65}(aL_{60} + bL_{65}) - L_{60}L_{70}, \\
a &= -0.29, \quad b = 1.27.
\end{aligned} \tag{A.3}$$

On the other hand, when the survival-ratio point is below the model line, the difference between the survival-ratio point and the model line is caused by nonspecific errors. Accordingly, the adjustment is to move the survival ratio point into the model line through minimal distance as

$$\begin{aligned}
\hat{S}_{60} &= \frac{-ab + S_{60} + bS_{65}}{1 + b^2}, \\
\hat{S}_{65} &= a + b\hat{S}_{60}.
\end{aligned} \tag{A.4}$$

$$\begin{aligned}
\hat{L}_{60} &= w\frac{L_{60} + \hat{S}_{60}L_{65} + \hat{S}_{60}\hat{S}_{65}L_{70}}{1 + \hat{S}_{60}^2 + \hat{S}_{60}^2\hat{S}_{65}^2} + (1-w)L_{60}, \\
\hat{L}_{65} &= w\hat{S}_{60}\frac{L_{60} + \hat{S}_{60}L_{65} + \hat{S}_{60}\hat{S}_{65}L_{70}}{1 + \hat{S}_{60}^2 + \hat{S}_{60}^2\hat{S}_{65}^2} + (1-w)L_{65}, \\
\hat{L}_{70} &= w\hat{S}_{65}\hat{S}_{60}\frac{L_{60} + \hat{S}_{60}L_{65} + \hat{S}_{60}\hat{S}_{65}L_{70}}{1 + \hat{S}_{60}^2 + \hat{S}_{60}^2\hat{S}_{65}^2} + (1-w)L_{70},
\end{aligned} \tag{A.5}$$

where $0 \le w \le 1$ is the weight, and is used as 0.5.