United Nations
Department of Economic and Social Affairs

Population Division

Technical Paper
No. 2014/4

Estimating Life Tables for Developing Countries

This page is intentionally left blank

Population Division

Technical Paper
No. 2014/4

Estimating Life Tables for Developing Countries

Nan Li

United Nations • New York, 2015

NOTE

PREFACE

The Population Division of the Department of Economic and Social Affairs of the United Nations Secretariat is responsible for providing the international community with up-to-date and impartial information on population and development. The Population Division provides guidance on population and development issues to the United Nations General Assembly, the Economic and Social Council and the Commission on Population and Development, and undertakes regular studies on population estimates and projections, fertility, mortality, migration, reproductive health, population policies and population and development interrelationships.

The purpose of the *Technical Paper* series is to publish substantive and methodological research on population issues carried out by experts within and outside the United Nations system. The series promotes scientific understanding of population issues among Governments, national and international organizations, research institutions and individuals engaged in social and economic planning, research and training.

This paper focuses on methods for estimating life tables for countries without complete mortality data, as a contribution to an internal discussion about how to improve the production of the *World Population Prospects*, the official United Nations publication of population estimates and projections. The paper was written by Nan Li of the Population Division. The author thanks the following colleagues of the Population Division for their comments and help in preparing this paper: Patrick Gerland and Victor Gaigbe-Togbe.

The *Technical Paper* series as well as other population information may be accessed on the Population Division's website at www.unpopulation.org. For further information concerning this publication, please contact the Office of the Director, Population Division, Department of Economic and Social Affairs, United Nations, New York, 10017, USA, by telephone: +1 (212) 963-3209, by fax: +1 (212) 963-2147, or by email: population@un.org.

CONTENTS

Page

FIGURES

TABLE

This page is intentionally left blank

ESTIMATING LIFE TABLES FOR DEVELOPING COUNTRIES

Nan Li

1. INTRODUCTION

The purpose of this paper is to discuss the estimation of life tables for developing countries[1] as part of the preparation of the United Nations *World Population Prospects* (see United Nations Population Division, 2015), which is the official United Nations publication for population estimates and projections. More specifically, the *World Population Prospects* (WPP) requires estimates of life tables by sex in successive five-year intervals from 1950 to 2010 or later, for all the countries that have the requisite size of population to ensure that the estimates are statistically meaningful. The sex differentials, age patterns and time trends of the estimated life tables should be consistent with demographic knowledge and common sense. Moreover, the estimation of life tables should be consistent with the changes in fertility, migration and population. This paper will focus on countries that do not have the empirical data to compute directly all the life tables by sex in successive five-year intervals from 1950 to 2010; and such countries are usually in less developed regions.

Empirical data used in estimating life tables are collected from three types of sources: (1) death registration that counts deaths by sex and age in a certain period, usually a calendar year; (2) census that enumerates the numbers of population by age and sex at a certain time point, and sometimes also death by age and sex during a period before the census time; and (3) sample survey that, in principle, could collect data on both death and population but cover only a small portion of the population in a country. For more specific information, one may consult the country-specific inventory of data sources provided by the United Nations Population Division[2].

Reliable death registration is available in developed countries and has also been established in recent years for a number of developing countries that have good socioeconomic conditions (see Demographic Yearbook, United Nations Statistics Division, 2013). For many developing countries, however, death registration does not exist or is unreliable, because the registration requires conditions that cannot be met at low levels of development. Moreover, death registration alone could only provide the number of deaths by age and sex, which are the numerators of age-specific death rates. The denominators of death rates are the person-years, or simply the mid-year populations, that are estimated using the data of census and international migration. The focus of this paper will not be the calculation of death rates using reliable data on death and population and international migration, but rather the estimation of life tables when the data are unreliable or unavailable for some or all the five-year intervals in 1950-2010.

Censuses are conducted in almost all the countries of the world. For example, among the 233 countries and areas listed in the 2015 revision of WPP, 214 have conducted the 2010-round of censuses between 2005 and 2014.[3] Besides providing mid-year populations used to compute death rates for countries with reliable death registration, some developing countries rely also on censuses to obtain life tables directly. Since census interviewers must visit every household in a country to enumerate the number of residents at a certain time point, they can also ask an additional question about whether there were any deaths in the household in past year, and the associated sex and age at death (United Nations

[1] The mortality of countries experiencing war, famine, natural disaster or major pandemics (e.g., HIV/AIDS) will not be discussed in this paper.
[2] http://www.un.org/en/development/desa/population/publications/mortality/world-mortality-cdrom-2015.shtml.
[3] http://unstats.un.org/unsd/demographic/sources/census/2010_PHC/censusclockmore.htm.

Statistics Division, 2004 and 2008). Furthermore, using population data of two successive censuses, some mortality indicators of the period between the two censuses can be estimated, especially for old ages at which the effect of migration is negligible (Li and Gerland, 2013). For many countries, census data on population by age and sex can be obtained from the UN Demographic Yearbook (e.g., United Nations Statistics Division, 2013). Occasionally, surveys using large sample size could also provide life tables.

Typical sample surveys often collect information only from a small portion of the population. Subsequently, they cannot produce life tables. This is because death rates at some ages, for example 10-20 years, are typically very low and hence require a large population to be estimated reliably. Nonetheless, sample surveys can provide reliable indicators of mortality for certain age groups when death is not a rare event 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 estimates for most countries since the 1970s or earlier (see United Nations Children's Fund, 2015; http://www.childmortality.org). Based on the same principles used to estimate child mortality such as birth histories, surveys such as the Demographic and Health Surveys (DHS, http://www.measuredhs.com/) have been collecting sibling histories since the 1990s to measure adult mortality, allowing to derive the probability of dying between age 15 and 50 or 60 years,[4] namely ${}_{35}q_{15}$ or ${}_{45}q_{15}$, respectively, for an increasing number of developing countries (Timæus, 2013). Combining data of surveys and other sources, Wang and colleagues (2012), at the Institute for Heath Metrics and Evaluation (IHME), estimated adult mortality[5] for 187 countries from 1970 to 2010 (see also GBD 2013 Mortality and Causes of Death Collaborators for the 2015 latest update).

Given the goal of estimating life tables by sex in successive five-year intervals from 1950 to 2010 and the gaps and challenges in empirical data as noted above, models and estimation methods are needed to utilise maximally the available empirical data to reach this goal.

2. METHODOLOGY

The situations of empirical data could be divided into four categories; and the methods and models can then be summarised according to these categories in table 1.

The first category, (A), includes cases of having empirical data only on child mortality. For almost all countries, estimates and data sources of child mortality can be acquired from UNICEF/ IGME from 1950 (or a later year) through 2015 (see http://childmortality.org). These estimates are based on collecting and evaluating multiple sources of data that show different levels and trends, and then modelling the data to produce time-series estimates with standard errors. For these countries, life tables could be estimated using one-dimensional model life tables such as the log-quadratic model based on data from the Human Mortality Database developed by Wilmoth and colleagues (2012), or regional mortality patterns based on Coale-Demeny model life tables (Coale, Demeny and Vaughan, 1983), or the United Nations model life tables (United Nations Population Division, 1982). See, for instance, the procedure MATCH in Mortpak for Windows (United Nations Population Division, 2013a) to compute such model life tables using one parameter input like ${}_5q_0$ through l_5 the number of survivors in the life table at age 5.

---

[4] Adult mortality is taken as the probability of dying between age 15 and 60 in this paper.

[5] Wang and colleagues provided life tables using child and adult mortality and relational model life tables; and this paper uses only their estimates of adult mortality, because UNICEF/IGME estimated child mortality for earlier periods, and because this paper aims to take into account empirical data on old-age mortality.

TABLE 1: A SUMMARY OF DATA AND METHODS

Country type	*Type of data*	*Recommended method*	*Equation number*
(A)	$_5q_0$ only	One-dimensional model life tables (log-quadratic model or other regional model life table system)	Not included in this paper
(B)	$_5q_0$ and adult mortality ($_{35}q_{15}$ or $_{45}q_{15}$)	F2dMLT (Flexible two-dimensional log-quadratic model life table)	(1)
(C)	$_5q_0$, $_{45}q_{15}$, old-age mortality	Augmented F2dMLT	(2) and (7)
(D)	Mortality rates by age and computed from raw data, or estimated from model life tables	CLC (Coherent Lee-Carter model)	(8)

The second category, (B), contains situations that have empirical data on child and adult mortality (at least for one or more years). Also, for almost all countries, time series estimates of adult mortality from 1970 through 2010 or later are produced by the Institute for Health Metrics and Evaluation (IHME) (http://www.healthdata.org/).[6]

The estimation of life tables in WPP is based not only on mortality data, but relies also on the consistency of levels and trends in fertility, migration and population. When the data, evaluation and models used by an analyst of WPP differ from that of the UNICEF/IGME or IHME, the estimates of child or adult mortality of WPP will be different from that of UNICEF/IGME or IHME.

For countries in the second category, life tables can be produced by either 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 this paper, this flexible two-dimensional model life table (F2dMLT) based on data from the Human Mortality Database is augmented to fit old-age mortality, and is therefore recommended for applications. Hereafter, F2dMLT is used to represent the flexible two-dimensional model life table, which is written as

$$\log(m_x) = a_x + b_x \cdot \log({}_5q_0) + c_x \cdot [\log({}_5q_0)]^2 + v_x \cdot k, \qquad (1)$$

where m_x stands for the five-year age-specific death rates with $x=0,1,5,10,\ldots$; coefficient vectors a_x, b_x, c_x, and v_x are given from fitting observed mortality data; and parameter k is flexible, which, for example, can be set as 0 when only child mortality $_5q_0$ is available, or can be solved to fit an additional $_{45}q_{15}$. Obviously, the F2dMLT can be used to produce a life table when only $_5q_0$, or both $_5q_0$ and $_{45}q_{15}$ are available.

The third category, (C), contains situations that have data with more than just child and adult mortality, but fewer than life tables for many time periods. This third category is practically important, because it includes empirical data on old-age mortality. More specifically, for countries without reliable

[6] Due to data deficiencies or limited availability (i.e., only for few years) for adult mortality in many developing countries, IHME $_{45}q_{15}$ time series estimates are based on a multistage statistical modelling relying on time series of socio-economic covariates (e.g., GDP per capita, average years of education by sex for ages 15-59 years, estimated crude death rates from HIV/AIDS by sex for ages 15 – 59 years).

death registration, the basis of empirical data could still be enlarged to include mortality at old ages such as 60 years and over.

Old-age mortality did not get much attention in the early years when surviving to old age was exceptional. In addition, population growth had long been the focus, to which old-age mortality is indeed less relevant. As a result, old-age mortality of many developing countries has been inferred by model life tables using child (and occasionally also adult) mortality, according to some relationships found in other countries. Moreover, model life tables are sometimes used to infer old-age mortality even for countries with reliable death rates at old ages (e.g., Wang and others, 2013). But the situation is entirely different now. In 2010-2015, deaths at age 60 years and older have already reached 60 per cent of all deaths worldwide,[7] according to the 2015 revision of the *World Population Prospects*. Furthermore, without adequate data on old-age mortality, important issues such as population ageing cannot be addressed properly.

Let the observed death rates at old ages be $\hat{m}_x$. The F2dMLT can be augmented to fit $\hat{m}_x$ by modifying coefficient a_x to $\hat{a}_x$:

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

Subtracting (1) from (2), there is $\log(\hat{m}_x) - \log(m_x) = \hat{a}_x - a_x$. Noting that the death rates at ages younger than 60 are the m_x of the F2DMLT, there is

$$\hat{a}_x = \begin{cases} a_x, \quad x < 60, \\ a_x + [\log(\hat{m}_x) - \log(m_x)], \quad x \geq 60. \end{cases} \tag{3}$$

Using (3), the F2DMLT described by (1) is augmented to (2)-(3), which will fit the observed old-age death rates exactly without changing the perfect model values of child and adult mortality.

One difference between the F2dMLT and the Lee-Carter model is that, the coefficients in the Lee-Carter model are country-specific, while in the F2dMLT they are universal for all countries. Comparing (1) and (2)-(3), it can be seen that the augmentation is to modify the universal a_x in (1) into the country specific $\hat{a}_x$ in (2)-(3), according to the country-specific $\hat{m}_x$.

How could the empirical data on old-age death rates be obtained? Although international surveys such as demographic and health surveys (DHS), are not yet collecting data on old-age population and death, national surveys conducted by some countries may be doing so. Furthermore, old-age mortality could be indirectly estimated using data on old-age population. Utilising data on old-age population of two successive censuses, Li and Gerland (2013) proposed a method, namely, the Census method, to estimate the probability of dying between age 60 and 75 years, which is called old-age mortality in this paper, and is denoted as ${}_{15}\hat{q}_{60}$. The reason of choosing age 75 as the upper bound is that many countries, especially those with relatively small population size, often aggregate the population aged 75 years and over into an open age group.

[7] Compared to the numbers of deaths at child and adult ages, the number of deaths at old ages is the biggest and also the least reliable. This irony necessitates improving the estimation of old-age mortality.

The Census method is based on two facts. The first is that migrants are rare at old ages. This fact makes using census population to estimate old-age mortality possible. Besides migration, misreporting of age in census is common in developing countries and raises another major obstacle to estimating old-age mortality. This obstacle is hard to eliminate directly, because the age patterns of real population are affected by historical changes of fertility, mortality and migration. The second fact is that the age pattern of a stationary population depends only on mortality, which could be used to remove the obstacle caused by misreporting of age. Using the variable-r method (Bennett and Horiuchi, 1981) to convert the age pattern of a real population to the age pattern of a stationary population, the Census method could utilise the common age patterns of stationary population found in the Coale-Demeny model life tables (Coale, Demeny and Vaughan, 1983) and the United Nations (United Nations Population Divison, 1982; Li and Gerland, 2011) model life tables to eliminate the errors of age misreporting, as is shown in the appendix. The Census method, however, still need a condition to accurately estimate old-age mortality, which is, that the completeness of the two successive censuses are similar. This condition is shown later in a case study.

When only ${}_{15}\hat{q}_{60}$ is available instead of $\hat{m}_x$, what should be the form of (3)? A simple answer 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 F2dMLT, and level α and pattern β can be chosen to fit some function of ${}_x\hat{q}_0$. When there is only ${}_{15}\hat{q}_{60}$, a customary way (see Preston, Heuveline and Guillot, 2001; p.200) is to set $\beta = 1$ and solve α to fit ${}_{15}\hat{q}_{60}$. 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:

$$\log(\hat{m}_x) = \alpha + \log(m_x). \tag{4}$$

Because

$${}_{15}\hat{q}_{60} \approx 1 - \exp[-5\cdot(\hat{m}_{60} + \hat{m}_{65} + \hat{m}_{70})], \tag{5}$$

α is solved by inserting (4) to (5):

$$\alpha \approx \log[\frac{\log(1-{}_{15}\hat{q}_{60})}{\log(1-{}_{15}q_{60})}] \tag{6}$$

where ${}_{15}q_{60}$ is the old-age mortality of the F2dMLT. Subsequently, (3) becomes

$$\hat{a}_x = \begin{cases} a_x, \quad x < 60, \\ a_x + \log[\frac{\log(1-{}_{15}\hat{q}_{60})}{\log(1-{}_{15}q_{60})}], \quad x \geq 60. \end{cases} \tag{7}$$

Finally, when old-age mortality is estimated at multiple years rather than a single year, α becomes the average over these years. Subsequently, (2) will describe the average of the ${}_{15}\hat{q}_{60}$ at these years.

The fourth category, (D), includes situations that have life tables in many five-year intervals in 1950-2010. Countries that conducted reliable censuses or large sample surveys that collected data on death, or established accurate death registration, after the 1960s or the 1970s, belong to this category. Moreover, countries that have life tables estimated based on a mixture of different data sources, including, eventually, even model life tables for some, but not all, the five-year intervals in 1950-2010, also belong to this category. For these countries, the coherent Lee-Carter model (Li and Lee, 2005) is recommended:

$$\log[m_x(t)] = A_{x,s} + B_x K(t). \qquad (8)$$

Hereafter, CLC is used to represent the coherent Lee-Carter model. In (8), $m_x(t)$ is the death rate at age x and interval t; $A_{x,s}$ represents the sex-specific corresponding coefficient vector in the original Lee-Carter (1992) model; and B_x and $K(t)$ are the two-sex-combined corresponding coefficient and parameter vectors in the original Lee-Carter model. In the coherent Lee-Carter model, the coefficients and the values of *K(t)* can be estimated utilising the data at the intervals for which life tables exist, using a method developed by Li, Lee, and Tuljapurkar (2004) that does not require original data to be in successive equal intervals. Noting that *K(t)* changes linearly over the periods during which socioeconomic conditions appear normal, the values of *K(t)* for the intervals without data can be obtained from linear interpolations or extrapolations of the estimated *K(t)*. Subsequently, for the intervals with missing data, the values of $m_x(t)$ are computed by (8) using estimated coefficients and the interpolated or extrapolated values of *K(t)* to derive abridged life tables for any time (t) (e.g., annually or for five year intervals).

The above discussion included typical, but not all the models and methods, in estimating life tables for WPP. For example, special models applied to countries with high prevalence of HIV/AIDS are excluded.

3. Case study of Indonesia

For many developing countries, life tables have been created on the basis of model life tables, because death registrations are unreliable or do not exist. Indonesia, a country with 240 million people in 2010, provided a typical example of a country that lacked the necessary resources to register deaths accurately (Muhidin, 2002). Although there were surveys collecting mortality information, such as the Indonesian Family Life Survey (IFLS) and the National Social and Economic Survey (SUSENAS), these surveys were incapable of producing life tables and their sample bias often caused questions. As a result, “Indonesia mainly relied on model life tables to estimate its mortality schedule” (Hidajat, Hayward and Best, 2004).

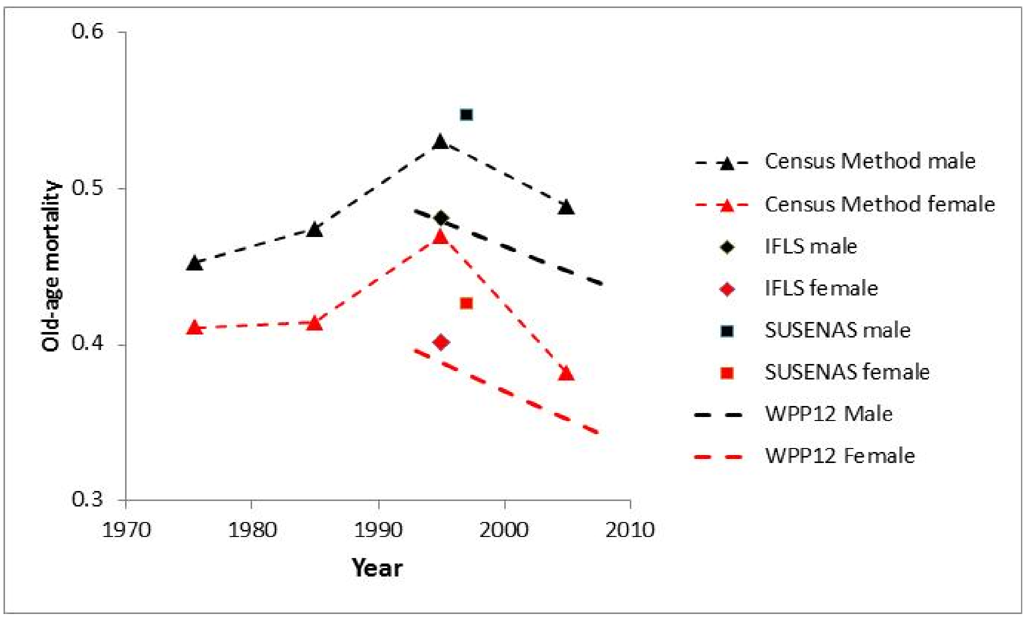
For Indonesia, data on child mortality for successive five-year intervals from 1950 to 2010 were estimated by the UNICEF/IGME (see www.childmortality.org), using indirect estimates from censuses (1971, 1980, 1990, 2000, 2010); and direct estimates from birth histories from the 2003-2010 SUSENAS; 1987, 1991, 1994, 1997, 2002-03, 2007, and 2012 Demographic and Health Surveys (DHS); and 1976 World Fertility Survey (WFS). These estimates are displayed as the triangles in figure 2.

The estimates of adult mortality for Indonesia from 1970 to 2010 were provided by the IHME (see http://vizhub.healthdata.org/mortality/), using recent household death data from the 1996 and 1998 SUSENAS; and 1993-1994, 1997, and 2007-2008 IFLS; and sibling histories from the 1991, 1994, 1997, 2002-03, and 2007 DHS. These estimates are shown as the triangles in figure 3

Data on old-age mortality, however, cannot yet be found from international sources such as the UNICEF/IGME or IHME, and must be collected and estimated on a country-specific basis.

To estimate old-age mortality using census data on population, the numbers of population by sex and five-year age group are available for 1971, 1980, 1990, 2000, and 2010 from censuses, in the UN Demographic Yearbook (e.g., United Nations Statistics Division, 2013). The first census of Indonesia was conducted in 1961, of which the numbers of population aged 60 years and over were aggregated in 10-year rather than 5-year age group, and hence cannot be utilized by the Census method (Li and Gerland, 2013). Applying the Census method to the data of all the censuses from 1971 through 2010, values of old-age mortality are shown in figure 1[8].

Figure 1. Estimates of old-age mortality ($_{15}q_{60}$) of Indonesia.



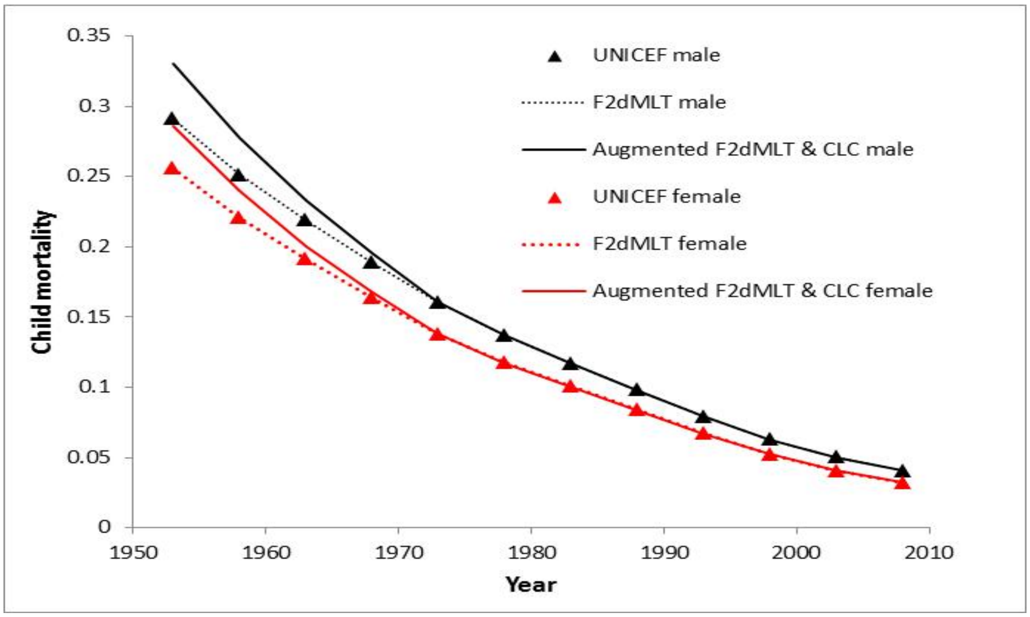
For 1975-1995, the Census method offered over-time increasing estimates, which are inconsistent with the practicality that mortality should decline under normal socioeconomic conditions. These increases could be explained by the improvement of completeness of the censuses. Using the Census method, an increase of the completeness from the first to the second census will result in underestimating old-age mortality, because the increased counts of population in the second census will be treated as additional survivors. From this point of view, the Census method is not recommended for early years when the completeness might improve remarkably between censuses. Between 1995 and 2005, the estimates declined, and they were close to the results of the IFLS and SUSENAS (Hidajat, Hayward, and Best, 2004). Although the two surveys were evaluated as less reliable in general, they could be accurate at old ages, because the population at old ages are less mobile and deaths at old ages are not rare. Considering the potential sample bias and the different completeness of the censuses, the estimates of the

[8] Hereafter the middle point is used to represent a corresponding time interval.

surveys and the Census method could be regarded as close rather than disperse. This closeness can hardly be understood as a coincidence; and a plausible explanation would be that these estimates are close to the true values. In comparison to the estimates of the Census method that uses data on old-age population, and of the surveys that count old-age deaths, the 2012 revision of the *World Population Prospects* (WPP12) is more likely to have underestimated old-age mortality for Indonesia, because it used child and adult mortality to infer old-age mortality according to some relationship found from other countries.

Having obtained the estimates of child, adult and old-age mortality, the next step is to use the F2dMLT on child and adult mortality. The results of this step are shown in figures 2 and 3. As the input data of the F2dMLT, the UNICEF/IGME estimates of child mortality are modelled perfectly as can be seen in figure 2.

Figure 2. Estimates and model values of child mortality (${}_5q_0$) of Indonesia



In the context of model life table, the progress of extending the input from only child mortality to include adult mortality is fundamental. When the data on adult mortality are unavailable for intervals earlier than 1970 and the default $k=0$ is used, however, the F2dMLT produced discontinuous changes between 1968 and 1973, which are described by the dotted curves in figure 3. These discontinuities could be removed in the following way. First, extrapolate the values of adult mortality according to the two-sex combined average decline rate between 1973 and 1988, which covers the same length of the extrapolation. Second, use the extrapolated values of adult mortality as the input of the F2dMLT. Results of the extrapolation, which are also the model values of adult mortality of the F2dMLT, are depicted by the dashed curves in figure 3.

Even if the available data were only child and adult mortality, the estimation would not have finished here, because the time trend and gender difference of old-age mortality still need to be checked. The reason is indicated below.

In comparing the estimates of old-age mortality in figure 4, it can be seen that the F2dMLT inferred values, shown as the dotted curves, are lower than all the estimates for both males and females. Under this condition, augmenting the F2dMLT to better fit the estimates of old-age mortality is a reasonable way to go.

Figure 3. Estimates and model values of adult mortality ($_{45}q_{15}$) of Indonesia

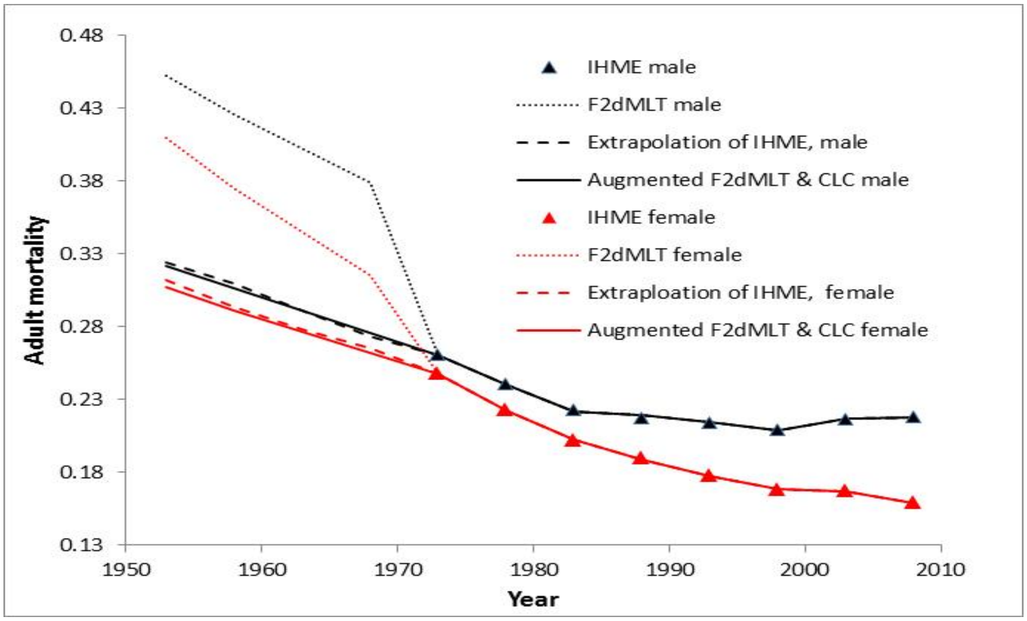
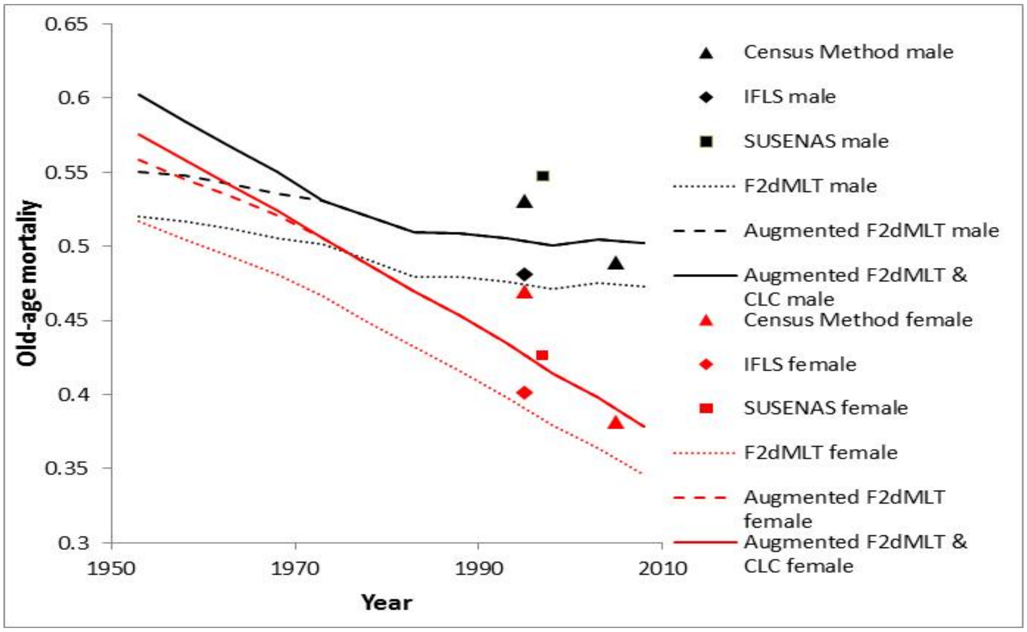


Figure 4. Estimates and model values of old-age mortality ($_{15}q_{60}$) of Indonesia



Applying (2) and (7) could fit any single estimate perfectly, or fit multiple estimates with minimal error. For the situation of Indonesia, an average estimate of 1995 is computed from using the estimates of 1995 of the Census method, the 1995 IFLS survey and the 1997 SUSENAS survey, because treating them equally will give year 1995 too large weight. Equation (7) is then applied to the average estimate of 1995 and the single estimate of 2005. The modifications of the F2DMLT $\hat{a}_x$ are shown in figures 5 and 6 below. It can be seen that the modifications are gentle and smooth.

The results of the augmented F2dMLT are indicated by the dashed curves in figure 4, which overlapped with the solid curves that will be indicated soon. The augmentation is successful for intervals after 1970, and does not affect the perfect model values at child and adult ages. In 1950-1954, however, a crossover between the curves of male and female is visible. Moreover, such a crossover also occurred in the results of the original F2dMLT, as the dotted curves in figure 4 indicated. This is why the time trend and gender difference of old-age mortality still need to be checked, after applying the F2dMLT to child and adult mortality.

Figure 5. Original a_x and modified $\hat{a}_x$ of males of Indonesia

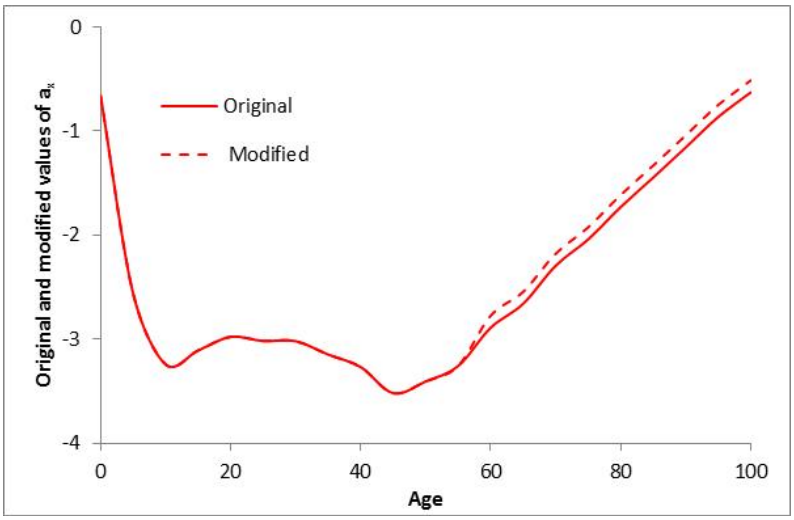
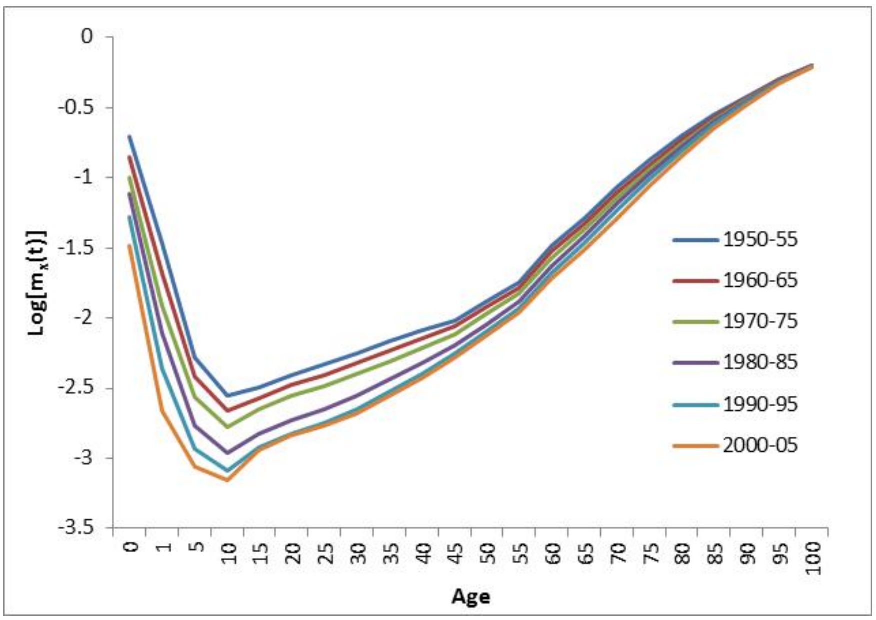


Figure 6. Original a_x and modified $\hat{a}_x$ of females of Indonesia

Crossovers can hardly be observed from countries with reliable data, and therefore should be removed. Noting that life tables after 1970 have now been successfully estimated, the coherent Lee-Carter model (CLC) can be used to extrapolate the death rates for intervals before 1970. Since the CLC is developed to prevent divergent or convergent trend in the projections of the original Lee-Carter model, and since the extrapolation is a backward projection, the crossover is expected to be eliminated by applying the CLC. Results of applying the CLC are depicted by the solid curves in figure 4. It can be seen that the CLC's extrapolation indeed removed the crossover. The CLC, however, cannot guarantee the maintenance of the perfect model values of child mortality, as can be seen in figure 2. Nonetheless, the difference between the estimates of the UNICEF/IGME and the CLC are moderate, and could be partly explained by the big standard errors of the UNICEF/IGME estimates for the 1950s and 1960s, which are based on only one survey that was conducted in 1976, and hence may not be as reliable as those of later intervals.

To conclude, death rates at all ages for successive five-year intervals from 1950-2010 can be computed for Indonesia, as can be seen in figure 7 for female as example, which fit all empirical data either perfectly or closely, and are smooth across age and over time.

Figure 7. Estimates of $Log[m_x(t)]$ for selected 5-year intervals from 1950-2010, Indonesia female

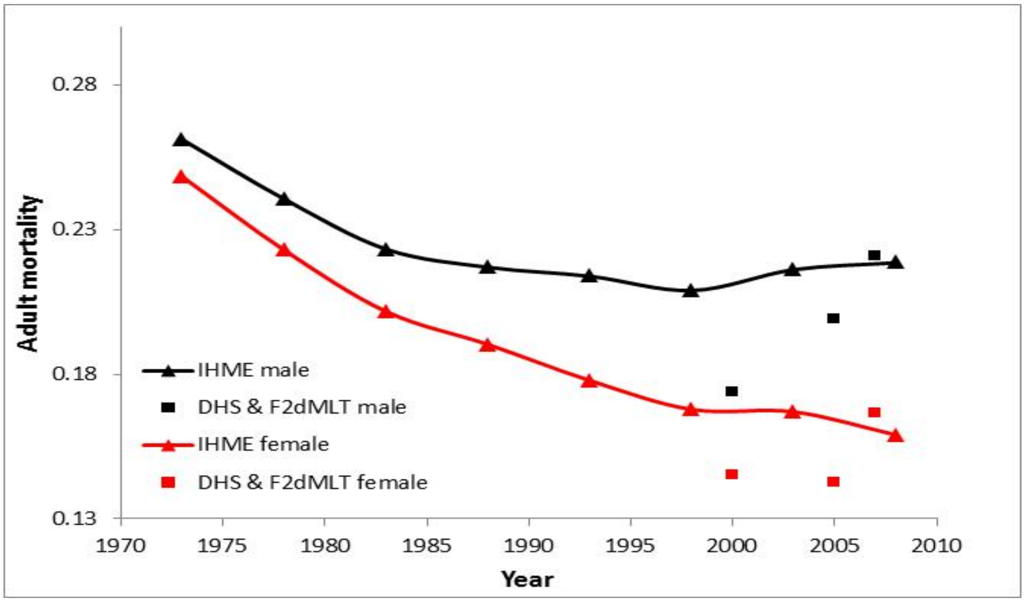


4. DISCUSSION

The case study of Indonesia provided an opportunity to illustrate the situations that require the various types of methods and models indicated above, but it does not include all potential issues in estimating life tables in WPP, some of which are discussed below, following the steps used for estimation.

The first step in estimating life tables involves utilising data on child and adult mortality. For countries that do not have reliable death registration and do not collect data on deaths in censuses and large-sample surveys, the simplest situation in this first step is to input the UNICEF/IGME child mortality and IHME adult mortality, with extrapolations from the 1950s, to the F2dMLT to yield the required life tables. In reality, however, the first step could be more complicated than this simplest situation. Taking also Indonesia as an example, the increase in adult mortality after 2000 may look suspicious, because in general, mortality should decline over time. Thus, finding data other than that of the IHME to check the unusual increase in adult mortality is desirable. Fortunately, the DHS later than 2000 included death rates at reproductive ages, which could be used for the purpose. Using death rates at reproductive ages, namely 15 to 49 years, the values of ${}_{35}q_{15}$ are computed directly, but how to obtain the values of ${}_{45}q_{15}$ is still an issue. Using the F2dMLT, the *k* in (3) can be solved to fit the ${}_{35}q_{15}$. Subsequently, the values of ${}_{45}q_{15}$ are estimated by the F2dMLT using the solved *k* in (3), and are compared with that of the IHME in figure 8.

Figure 8: Independent estimates of adult mortality (${}_{45}q_{15}$) for Indonesia



It can be seen that the DHS data also showed increases in adult mortality after 2000, not only for males but also for females. The reason why IHME data did not depict an increase for females is, perhaps, because the increase in the DHS data was revealed only by the 2012 survey, which was not available when IHME published its estimates in 2012. Although an explanation of the increase in the IHME data remains yet to be explored, the DHS data suggest that the increase is possible. Of course, if the DHS data showed a smooth decline trend, then they would be used to replace the IHME data. Moreover, a similar check would be required for child mortality, if the UNICEF/IGME data showed unusual trend.

The second step is to estimate old-age mortality, which starts with collecting and analysing survey and census data. This step could be time consuming and can result in nothing. For example, it may turn out that proper surveys cannot be found, or that the Census method does not produce reasonable results. Nonetheless, this step is worth trying, because it could extend the empirical basis of estimating life tables to include old-age mortality.

The third and last step is to produce life tables for successive five-year intervals from 1950 to 2010, of which the death rates should change smoothly across age and over time. The examples in the case study of Indonesia include augmentation of the F2dMLT and an application of the CLC. For other countries, the situation could be either simpler or more complex. For some countries, for instance, the F2dMLT may fit the estimated old-age mortality properly and may not lead to a crossover in early intervals. In this situation, neither an augmentation of the F2dMLT nor the CLC is needed. For other countries, the augmentation of the F2dMLT and the CLC may have been used, yet the model values of child mortality are far away from the UNICEF/IGME estimates that are highly reliable. In this situation, the constrained CLC, in which the *K(t)* in (2) is no longer linear but solved to fit the target child mortality, could be applied to fix the problem.

Finally, some countries may have reliable life tables for a few scattered years obtained from censuses or large-sample surveys. In this situation, these life tables could provide more empirical data in addition to the F2dMLT life tables for 1970-2010, using the CLC interpolation if the scattered years are after 1970, or extrapolation if the scattered years are before 1970. Moreover, for countries that have many but not all life tables for 1950-2010, the CLC interpolation and extrapolation can be used to yield life tables for successive five-year intervals in 1950-2010.

5. SUMMARY AND CONCLUSIONS

In estimating life tables for WPP, the target is to accurately describe the real levels and age patterns of mortality. For developing countries that do not have the resources to establish reliable death registration, this goal could be approached gradually, through collecting more empirical data and improving the methods and models to utilise these data.

For most developing countries, national population surveys were rare in the early years and data on child mortality were first sampled by international surveys in the 1970s, through which estimates could be extended to the 1950s. In the case of Indonesia, the estimates of child mortality began from the 1976 World Fertility Surveys. Data on adult mortality were collected by international surveys only in the late 1990s or after the year 2000. For Indonesia, death rates at reproductive ages were first estimated by DHS in the survey of 2002-2003. Currently, data on old-age mortality could only be estimated indirectly, or found only in censuses or large-sample surveys conducted by specific countries in scattered years.

With the availability of data, estimating life tables in the early revisions of WPP relied on child mortality and a one-dimensional model life table for most developing countries. For Indonesia, the revisions of WPP before 2010 were based on child mortality and the West Family of the Coale-Demeny (see Coale, Demeny, and Vaughan, 1983) model life tables (CDW). In the 2012 revision of WPP, an effort was made to take into account adult mortality. This was done through three steps, of which the first was to produce the life tables of 2005 using the values of ${}_{35}q_{15}$ estimated from the DHS 2007 and the UNICEF/IGME ${}_{5}q_{0}$, and the COMBIN procedure in MORTPAK for Windows using the CDW (United Nations Population Division, 2013a). The second step was to apply the CLC to interpolate and extrapolate all the life tables, using the life tables of 2005 and 1953 that is yielded by CDW with UNICEF/IGME ${}_{5}q_{0}$. Finally, the third step was to check whether the ${}_{5}q_{0}$ of these life tables was close to the estimates of UNICEF/IGME from 1950 to 2010, and if not, the constraint CLC was applied. The estimates of life tables for Indonesia presented in this paper have been used to prepare the 2015 revision of WPP, namely WPP15. The major similarities and differences between WPP12 and WPP15 estimates are discussed below.

Figure 9 indicates that the two revisions are similar with regards to child mortality and both are close to the UNICEF/IGME estimates. For adult mortality, however, the two revisions are close only after 2000, in which both revisions stand on empirical basis. Using the F2dMLT, WPP15 fits the IHME estimates perfectly after 1970, and remarkably improved the estimation of adult mortality of WPP12 before 2000, as shown in figure 10.

Figure 9: Two estimates of child mortality (${}_5q_0$) for Indonesia

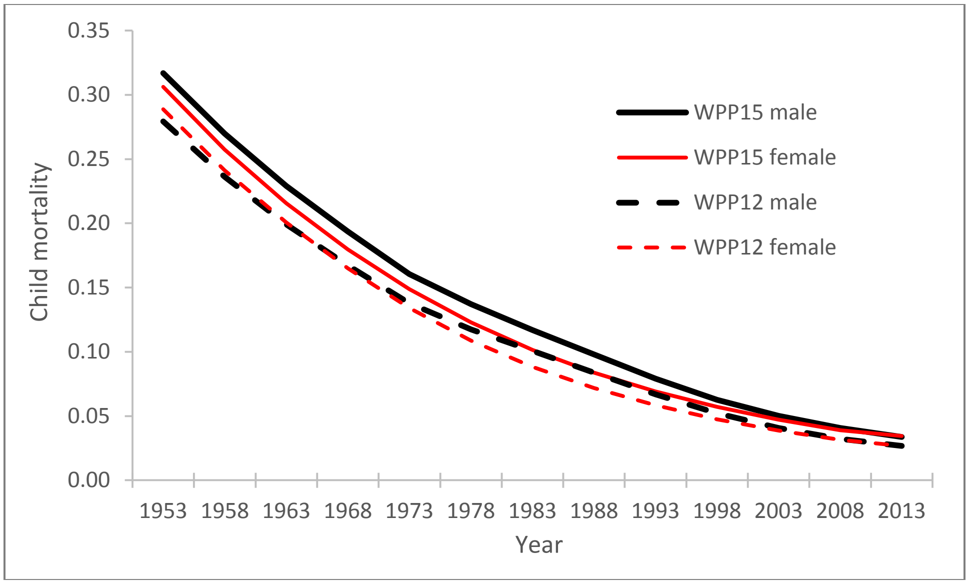
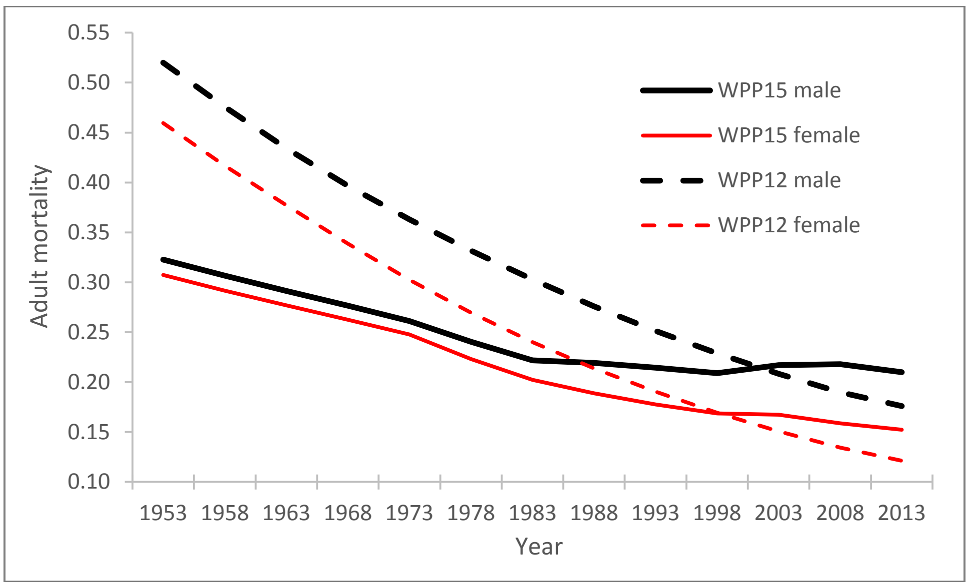
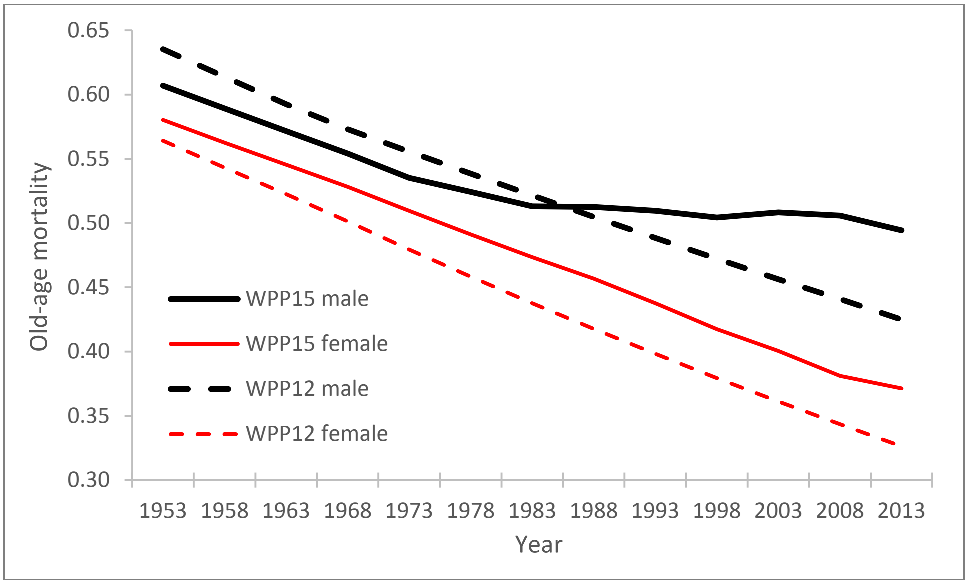


Figure 10: Two estimates of adult mortality (${}_{45}q_{15}$) for Indonesia



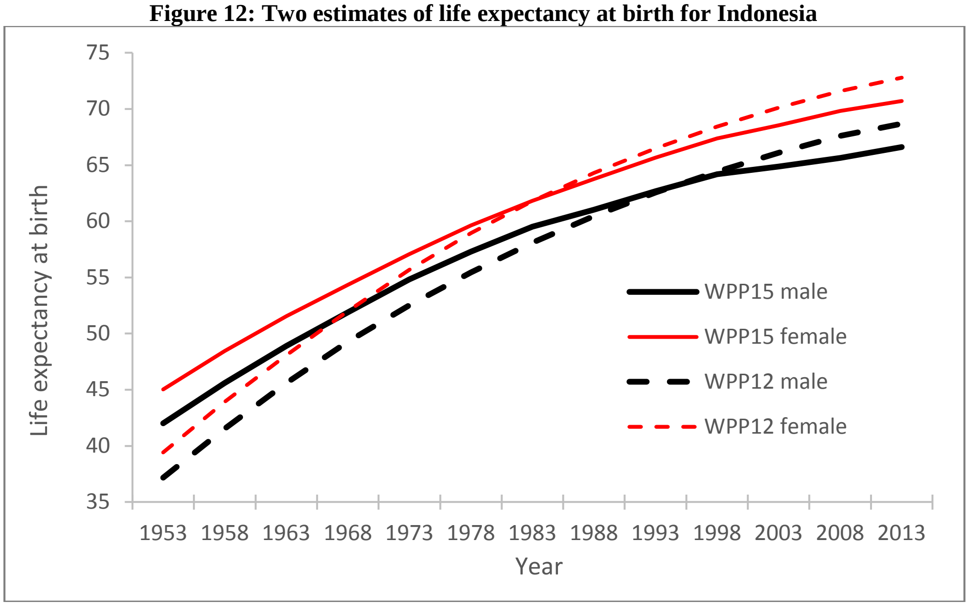
Estimates of old-age mortality are depicted in figure 11. It can be seen that the two revisions differ remarkably in all the intervals. Using child and adult mortality, WPP12 inferred old-age mortality according to the relationships found in other countries. Based on census and survey data, WPP15 produced life tables for years after 1970 using the augmented F2DMLT, and extrapolated life tables at earlier years using the CLC. Compared to the estimates of WPP15, WPP12 is likely to underestimate old-age mortality for years later than 1990 and to overestimate the mortality difference between males and females for years before 1990. More important than the quantitative differences, WPP15 enlarged the empirical basis to include the estimates of old-age mortality.

Figure 11: Two estimates of old-age mortality ($_{15}q_{60}$) for Indonesia



The importance of estimating old-age mortality is even greater, when looking at the difference between the estimates of life expectancy at birth in figure 12. For 2003, the two revisions used identical child mortality and similar adult mortality as can be seen in figures 9 and 10; but WPP15 estimated higher old-age mortality than that of WPP12, as is shown in figure 11. As a result, figure 12 indicates that the estimates of life expectancy at birth in WPP15 are notably lower than that of WPP12. Old-age mortality is more important to estimating life expectancy in recent years, in which the probability of surviving from birth to an old age is much higher than that in earlier years. For Indonesia, the probability of surviving from birth to age 60 years increased from 0.43 in 1953 to 0.77 in 2003.

Figure 12: Two estimates of life expectancy at birth for Indonesia



Life expectancy at birth is widely adopted as a key indicator of the well-being of a country. For instance, the Human Development Index of the United Nations lists life expectancy at birth as the first of its three components. Given that old-age mortality is increasingly important to estimating life expectancy, and that life expectancy is a key indicator of a country's development level, greater attention should be given to estimating old-age mortality.

In estimating mortality for developing countries in WPP, data on child mortality have become available in international surveys and other sources in the 1970s; and since then they have been utilised to produce life tables using one-dimensional model life tables. Starting from around 2000, reliable data on adult mortality have been collected by international and national surveys; and this paper suggests the use of child and adult mortality estimates and the F2dMLT to produce life tables. Utilising both child and adult mortality will certainly make the estimates better than using only child mortality, assuming the methods used for the estimation of adult mortality are reliable. To further improve the estimates of mortality for developing countries, this paper augmented the F2dMLT to model old-age mortality and indicated that it has become timely for researchers to develop more suitable methods and models to utilise empirical data on old-age mortality and, more importantly, for international and national surveys to collect empirical data on old-age mortality.

It is also worth noting that while all the analytical methods presented in this paper and summarized in section 2 and Table 1 have been used to derive abridged life tables for five-year periods, they can also be used to derive time series of abridged life tables for any relevant time periods for which mortality estimates are available. For instance, they can easily be used to produce annual time series of abridged life tables using at the minimum an annual time series of child mortality estimates, complemented as much as possible by annual estimates of adult mortality (and old-age mortality when available).

As this paper has demonstrated, the log-quadratic model based on data from the Human Mortality Database developed by Wilmoth and colleagues (2012) is remarkably flexible and can be extended to take into account additional information when availability and circumstances allow. The augmentation presented therein focused on the incorporation of additional information on old-age mortality for countries with deficient vital registration data, but further research and potential additional extensions are conceivable to extend further the log-quadratic model to deal with mortality shocks and crises due to natural disasters, famines and conflicts, and eventually pandemics like HIV/AIDS (if adult HIV prevalence and ART coverage are known). These potential extensions, however, are beyond the scope of this paper and will require further research and additional empirical evidence.

REFERENCES

Bennett, N.G.and S. Horiuchi (1981). Estimating the completeness of death registration in a closed population. *Population Index* Vol. 47, No. 2, pp. 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.

Demographic and Health Surveys (2008). *Indonesia Demographic and Health Survey 2007*. Maryland USA: Macro International Calverton*.*

Hidajat, M.M., M.D. Hayward and L.E. Best (2004). The life table without the vital registration system: Alternative models of Indonesian mortality. (Manuscript).

Lee, R. D. and L. Carter (1992). Modeling and forecasting the time series of U.S. mortality. *Journal of the American Statistical Association* Vol. 87, pp. 659-671.

Li, N. and P. Gerland (2013). Using census data to estimate old-age mortality for developing countries. Paper prepared for and presented to Session 17-05: Indirect methods of mortality and fertility estimation: new techniques for new realities of the XXVII IUSSP International Population Conference. (Tuesday, August 27th 2013, 13:30pm - 15:00pm) Busan, Korea. Available from http://iussp.org/en/event/17/programme /paper/1959.

_____ (2011). Modifying the Lee-Carter method to project mortality changes up to 2100. Paper presented at the Annual Meeting of the Population Association of America, Washington, DC.http://paa2011 .princeton.edu/download.aspx?submissionId=110555.

Li, N. and R. Lee (2005). Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method. *Demography,* Vol. 42, No. (3), pp. 575-594.

Li, N., R. Lee and S. Tuljapurkar (2004). Using the Lee-Carter method to forecast mortality for populations with limited data. *International Statistical Review,* Vol. 72, No. 1, pp. 20-36.

Muhidin, S. (2002). *The Population of Indonesia: Regional Demographic Scenarios Using a Multiregional Method and Multiple Data Sources*. Amsterdam: Rozenberg Publishers.

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.

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

United Nations Children's Fund (2015). *Levels and Trends in Child Mortality: Report 2015*. New York: United Nations Children's Fund. Available from http://www.childmortality.org.

United Nations Population Division (1982). *Model Life Tables for Developing Countries*. New York: United Nations. Available from http://www.un.org/esa/population/publications/Model_ Life_ Tables/Model_Life_Tables.htm.

_____ (2013a). *MortPak–The United Nations Software Package for Mortality Measurement*. New York: United Nations. Available from http://www.un.org/en/development/desa/population/publications/ mortality/mortpak.shtml.

_____ (2015). *World Population Prospects: The 2015 Revision*. New York: United Nations. Available from http://esa.un.org/wpp/.

United Nations Statistics Division (2004) *Handbook on the Collection of Fertility and Mortality Data, Statistical Papers Series F,* No. 92, New York, NY: United Nations. Available from http://unstats. un.org/unsd/demographic/standmeth/handbooks/default.htm.

_____ (2008) *Principles and Recommendations for Population and Housing Censuses, Revision 2, Statistical Papers Series M, Rev. 2nd ed.,* New York, NY: United Nations. Available from http://unstats.un.org/unsd/demographic/sources/census/census3.htm.

_____ (2013). *Demographic Yearbook 2012, 63rd issue*. New York: United Nations. Available from, https://unstats.un.org/unsd/demographic/products/dyb/2010_round_latest.htm.

Wang H. and others (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,* Vol. 13, No. 380, pp. 2071-2094.

GBD 2013 Mortality and Causes of Death Collaborators (2015). Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet, 385(9963), pp. 117-171.

Wilmoth, J.R. and others (2012). A flexible two-dimensional mortality model for use in indirect estimation. *Population Studies,* Vol. 66, No. 1, pp. 1-28.

APPENDIX

This appendix describes the Census method that uses census data on population to estimate old-age mortality, or the probability of dying between ages 60 and 75:

$$ {}_{15}q_{60} = 1 - \frac{l_{75}}{l_{60}}, \qquad \text{(a.1)} $$

where l_x represents the number of survivors at age *x*. The first step of the Census method is to use the variable-r method (Bennett and Horiuchi, 1981) to convert the age pattern of a real population into the age pattern of a stationary population. Let *p(x, t)* be the observed number of population in age group *[x,x+5)* in a census at time *t*, where *x*=60, 65, 70. The growth rates at age *x* are computed as

$$ r(x) = Log[\frac{p(x,t_2)}{p(x,t_1)}]/(t_2 - t_1), \quad x = 60,65,70. \qquad \text{(a.2)} $$

And he accumulated growth rates are

$$ \begin{aligned} s(60) &= 2.5r(60), \\ s(65) &= 5r(60) + 2.5r(65), \\ s(70) &= 5[r(60) + r(65)] + 2.5r(70). \end{aligned} \qquad \text{(a.3)} $$

Further, the middle-point population in age group *[x,x+5), N(x),* are computed as

$$ N(x) = \sqrt{p(x,t_1)\,p(x,t_2)}, \quad x = 60,65,70. \qquad \text{(a.4)} $$

Furthermore, the corresponding person-years in the underlying stationary population, L_x, are obtained as

$$ L_x = N(x)\exp[s(x)], x = 60,65,70. \qquad \text{(a.5)} $$

When *r(x)* is constant over age and *N(x)* is the age-specific number of a stable population, the L_x in equation (5) is obviously the age-specific number of the underlying stationary population. When *r(x)* changes over age and *N(x)* represents a non-stable population, the variable-r method indicates that the L_x in (5) is still the person-years of the underlying stationary population.

The variable-r method assumes that the errors in enumerating population are constant with age, and identical in the two successive censuses. But in developing countries the errors often occur unevenly across age. A typical example is age heaping. When such errors are severe, the L_x resulted from variable-r method, namely the variable-r L_x, 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.

It is hard to find a proper basis to adjust the errors of age misreporting in a real population, which is affected by historical fertility, mortality and migration. But a stationary population is determined by mortality only. Thus, it is possible to find a proper basis to adjust the errors of age misreporting in stationary populations; and the second step of the Census method is to find the common age-pattern relationship of stationary populations to adjust the errors of age misreporting. Denoting the ratios of surviving from ages 60-64 to 65-69 by S_{60}, and from ages 65-69 to 70-74 by S_{65}, and using the Coale-Demeny (Coale, Demeny, and Vaughan, 1983) and the United Nations model life tables (United Nations, 1982; Li and Gerland, 2011), the common relationship between S_{60} and S_{65} is found as:

$$S_{65} = a + b \cdot S_{60} = -0.29 + 1.27 \cdot S_{60}. \qquad \text{(a.6)}$$

Using (a.6) and S_{60} to predict the S_{65} for all the families of the Coale-Demeny and the United Nations model life tables and at the levels of life expectation from 20 to 100 years, the explanation ratios are all bigger than 0.97. In other words, (a.6) is the common relationship between S_{60} and S_{65}, and can be used to adjust the variable-r L_x.

The third step of the Census method is to adjust errors of age misreporting. When the observed survival-ratio point, (S_{60}, S_{65}), is above the model line (a.6), the difference between the survival-ratio point and the model line is caused mainly by age heaping. When there are Δ persons aged 65-69 years reported their age as 70, the adjustment on L_{70} should be a reduction of Δ and on L_{65} should be an increase of Δ. Assuming that the heaping ratio at age 60 is the same as that at age 70, the adjustment can be written as:

$$\begin{aligned}
\hat{L}_{60} &= L_{60} - \frac{L_{60}}{L_{70}}\Delta, \\
\hat{L}_{65} &= L_{65} + \Delta, \qquad \text{(a.7)} \\
\hat{L}_{70} &= L_{70} - \Delta.
\end{aligned}$$

Using the common relationship (a.6), the Δ in (a.7) is solved as

$$\begin{aligned}
\Delta &= \frac{-B + \sqrt{B^2 - 4AC}}{2A}, \\
A &= b - a\frac{L_{60}}{L_{70}} - \frac{L_{60}}{L_{70}}, \\
B &= a(L_{60} - \frac{L_{60}}{L_{70}}L_{65}) + 2bL_{65} + L_{60} + \frac{L_{60}}{L_{70}}L_{70}, \qquad \text{(a.8)} \\
C &= L_{65}(aL_{60} + bL_{65}) - L_{60}L_{70}, \\
a &= -0.29, \quad b = 1.27.
\end{aligned}$$

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 reasons. Accordingly, the adjustment is to move the survival-ratio point to the model line through minimal distance, which is done by choosing an $\hat{S}_{60}$ to minimize $[(\hat{S}_{60} - S_{60})^2 + (a + b\hat{S}_{60} - S_{65})^2]$. The result of the minimization is:

$$\hat{S}_{60} = \frac{-ab + S_{60} + bS_{65}}{1+b^2},$$
$$\hat{S}_{65} = a + b\hat{S}_{60}. \tag{a.9}$$

$$\hat{L}_{60} = w\frac{L_{60} + \hat{S}_{60}L_{65} + \hat{S}_{60}\hat{S}_{65}L_{70}}{1+\hat{S}_{60}^2 + \hat{S}_{60}^2\hat{S}_{65}^2} + (1-w)L_{60},$$
$$\hat{L}_{65} = w\hat{S}_{60}\frac{L_{60} + \hat{S}_{60}L_{65} + \hat{S}_{60}\hat{S}_{65}L_{70}}{1+\hat{S}_{60}^2 + \hat{S}_{60}^2\hat{S}_{65}^2} + (1-w)L_{65}, \tag{a.10}$$
$$\hat{L}_{70} = w\hat{S}_{65}\hat{S}_{60}\frac{L_{60} + \hat{S}_{60}L_{65} + \hat{S}_{60}\hat{S}_{65}L_{70}}{1+\hat{S}_{60}^2 + \hat{S}_{60}^2\hat{S}_{65}^2} + (1-w)L_{70},$$

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

Using $\hat{L}_x$, the Gompertz law that the force of mortality, $\mu(x)$, rises exponentially with age as $\mu(60)\cdot e^{g\cdot(x-60)}$, is extended to describe $\hat{L}_x$ as

$$\hat{L}_{60} = \int_{60}^{65} l_x dx = l_{60}\int_{60}^{65} e^{-\int_{60}^{x}\mu(60)\exp[g\cdot(y-60)]dy} dx,$$
$$\hat{L}_{65} = l_{60}\int_{65}^{70} e^{-\int_{60}^{x}\mu(60)\exp[g\cdot(y-60)]dy} dx, \tag{a.11}$$
$$\hat{L}_{70} = l_{60}\int_{70}^{75} e^{-\int_{60}^{x}\mu(60)\exp[g\cdot(y-60)]dy} dx.$$

Solving the values of g, $\mu(60)$, and l_{60} in (a.11), ${}_{15}q_{60}$ is estimated as:

$${}_{15}q_{60} = 1 - \frac{l_{75}}{l_{60}} = 1 - \exp\{-\frac{\mu(60)}{g}[\exp[g(75-60)] - 1]\}. \tag{a.12}$$

Why do we estimate only ${}_{15}q_{60}$ but not the death rates for the three age groups? Estimating errors of death rates could cancel each other over age; and hence ${}_{15}q_{60}$ is more robust.