This paper is published in Demography of Population Health, Aging and Health Expenditures (p133-149) edited by Christos H Skiadas and Charilaos Skiadas, *The Springer series on demographic methods and population analysis 50*, Springer.

Using the Developing Countries Mortality Database (DCMD) to Probabilistically Evaluate the Completeness of Death Registration at Old Ages

Nan Li[1], Hong Mi[2], and Xiaotong Tang[3]

1. Population Division, Department of Economic and Social Affairs, United Nations, 2 UN Plaza, 233 East 44th Street, Room DC2-1938, New York, NY 10017, USA (Email: li32@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#, Mengminwei 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, Room 409-1#, Xiyi Building, Zijin'gang Campus, Hangzhou, Zhejiang, P. R. China. (E-mail: 1048259552@qq.com).

Like the ratio of registered deaths to total deaths, the deterministic completeness of death registration (DR) cannot be exactly 1 in practice. Consequently, it is impossible to use deterministic completeness to check whether a DR is complete, which is a problem for developed countries. We propose a probabilistic completeness whose samples are the values of deterministic completeness. When the difference between 1 and the mean of probabilistic completeness is statistically insignificant, the DR is probabilistically complete. But using intercensal population change to estimate deaths and deterministic completeness is still an issue, because it requires unrealistic assumptions about migration and census errors. Focusing on old age and the level of mortality rather than the number of deaths, the effects of migration and census errors are largely reduced in the Developing Countries Mortality Database (DCMD, www.lifetables.org), which is used in this paper to provide applications of the probabilistic evaluation.

(1) Introduction

The deterministic completeness of deaths registered in the census or death registration (DR) is usually defined as the ratio of registered deaths to total deaths. Normally, a DR can be judged as complete when the deterministic completeness is 1. In practice, however, the deterministic completeness cannot be exactly 1 and it is impossible to judge whether a DR is complete. Consequently, even for the developed countries that are commonly believed to have reliable DRs, such as those in the Human Mortality Database, there is no quantitative evaluation to conclude that their DRs are complete.

To solve this dilemma, we propose a probabilistic completeness, of which the samples are deterministic completeness. When the difference between 1 and the mean of probabilistic completeness is statistically insignificant, the DR is probabilistically complete.

Estimating the deterministic completeness of the DR is an important and longstanding issue. The analytical evaluations of the completeness of the DR originate from stable population models. In stationary populations, the number of deaths over a given age is the number of persons at the given age. Therefore, the number of registered deaths over a certain age could be evaluated by the number of persons at this age if the population is stationary. In a stable population, the number of deaths over a given age is the number of persons at the given age minus an additional term, which is the product of the number of persons over the given age and the growth rate of the stable population. This relationship was first utilized by Brass (1975) to evaluate the completeness of the DR. The evaluation was extended to non-stable populations, using two successive censuses and at synthetic extinct generation, seg (Bennett and Horiuchi 1981; 1984), a generalized growth balance, ggb (Hill, 1987), and a combination of ggb, seg, and ggbseg (Hill, You, and Choi, 2009). These methods are convincing methodologically but require unrealistic assumptions, which typically include the absence of migration and that the errors of the census population obey special relationships (Li and Gerland, 2017).

At old ages such as 60 years and over, migrants are negligible compared to deaths. Furthermore, using census population could accurately estimate the levels of mortality at old ages, which provides the basic estimates of the Developing Countries Mortality Database (DCMD, www.lifetable.org, see Li, 2014; Li, Mi and Gerland, 2017; Li, Mi, Gerland, Li, and Sun, 2018). In DCMD, the basic estimates of old-age mortality (or the probability of dying between ages 60 and 75) are obtained using (1) the variable-r method with adjustment of age heaping (Bennett and Horiuchi 1981; Li and Gerland, 2013), (2) the survival model (Li, Mi, Gerland, Li, and Sun, 2018), and (3) the two-input-parameter model life table with child and adult mortality (Wilmoth, Zureick, Canudas-Romo, Inoue, and Sawyer, 2012). The DCMD estimates of old-age mortality are calculated as the averages of the above three basic estimates and extended to single years using local regressions (Li and Mi, 2018).

In DCMD, the unrealistic assumptions are largely reduced. Consequently, the DCMD estimates of old-age mortality should be more reliable than that of the previous methods. Taking the DCMD estimates of old-age mortality as accurate, we applied the probabilistic evaluation to all the developing countries included in DCMD and obtained encouraging results. Because the socioeconomic conditions of some developing countries are similar to those of developed countries, the probabilistic evaluation should also work for developed countries, including those in the Human Mortality Database.

(2) The Method

(2.1) The fundamental difference

Using the numbers of the population aged 60 to 75 years (hereafter all age ranges for both population and deaths are 60 to 75 years) enumerated from two successive censuses could estimate the number of deaths occurred, and the prevailing mortality levels between the two censuses. Big estimating errors could occur in the number of deaths but not in the mortality levels. This fundamental difference may indicate the success of evaluating the completeness of the DR based on estimating the levels of mortality in the DCMD and this paper, and may also explain the failure in evaluations based on estimating the number of deaths in previous studies (Li and Gerland, 2017).

(2.1.1) ***Estimating the numbers of death***

Let the number of persons in an age interval enumerated in the first census be p_1 and the number of the survivors in the next census be p_2. Furthermore, let the net undercount rates[1] be u_1 and u_2 for the first and second censuses, respectively. Neglecting intercensal migration at old ages, the number of deaths (d) and estimated deaths ($\hat{d}$) are:

$$\begin{aligned} d &= p_1 - p_2, \\ \hat{d} &= \hat{p}_1 - \hat{p}_2 = p_1(1-u_1) - p_2(1-u_2) = d - (p_1u_1 - p_2u_2). \end{aligned} \qquad (1)$$

Furthermore, let the survival ratio be

$$s = p_2 / p_1. \qquad (2)$$

Then, the relative error in estimating the number of deaths is

$$e_d = \frac{\hat{d}-d}{d} = -\frac{p_1u_1 - p_2u_2}{d} = -\frac{p_1(u_1 - s\cdot u_2)}{p_1(1-s)} = -\frac{u_1 - s\cdot u_2}{1-s}. \qquad (3)$$

Eq (3) indicates that, except for two special cases ($u_1 = s\cdot u_2$ and $u_1 = u_2$), the effect of the survival ratio could lead to large errors in estimating the number of deaths, especially when the survival ratio is close to 1.

(2.1.2) ***Estimating mortality levels***

The estimated survival ratio can be written as

$$\hat{s} = \hat{p}_2 / \hat{p}_1 = \frac{p_2(1-u_2)}{p_1(1-u_1)} = s\frac{1-u_2}{1-u_1}. \qquad (4)$$

[1] The net undercounting rate represents the relative difference between the reported and the true numbers of population, which could be the result of misreporting of people or misreporting of age. A positive net undercounting rate indicates net under-counting or that the reported number is smaller than the true number. The net undercounting rate could also be negative to reflect net over-counting that may or may not be caused by misreporting of age.

Using (2), the relative errors in estimations *s* is

$$e_s = \frac{\hat{s} - s}{s} = \frac{1-u_2}{1-u_1} - 1 = \frac{u_1 - u_2}{1-u_1}. \qquad (5)$$

It is clear that the estimation errors of the survival ratio (or mortality level) are not affected by the survival ratio itself, because the true value of the survival ratio (*s*) cancels itself in the estimation. Eq (5) differs fundamentally from Eq(3), indicating that large errors are not likely to occur in the estimation of mortality levels, and provides the basis of DCMD and this paper.

(2.2) A deterministic model of completeness

The CDMD provides the estimates of single-year life tables in developing countries, which should be more accurate than the estimates of the number of deaths, as per the discussion above. In this paper, death rates *m(x,t)* obtained from these life tables are assumed to be accurate and are used to evaluate the completeness of the DR at census years, where the numbers of deaths and of the population are available to compute the registered death [2]:

$$m_r(x,t) = \frac{d_r(x,t)}{p_r(x,t)}, \qquad (6)$$

where $d_r(x,t)$ and $p_r(x,t)$ are the registered numbers of deaths and the reported numbers of the population aged *x* in a census, in the corresponding period and at time *t*. Although the evaluation can be carried out for individual ages in a single year, it is more robust and convenient to calculate the completeness of the DR for the age group 60-75 years, because the deaths could be rare at single-year age in some countries, and because $\{1-\exp[\sum_{x=60}^{75} m(x,t)]\}$ equals old-age mortality (the probability of dying between ages 60 and 75). To evaluate the completeness of the DR for the age group 60-75 years, however, it adopts an implicit assumption is adopted that the completeness is constant over the age group 60 to 75 years.

The completeness of the DR at time *t* can be defined as $c(t) = \sum_{x=60}^{75} d_r(x,t) / \sum_{x=60}^{75} d(x,t)$, where $d(x,t)$ represents the accurate number of deaths in the corresponding period and age. The accurate number of deaths, $\sum_{x=60}^{75} d(x,t)$, however, is difficult to estimate. Nonetheless, a reliable estimate of the completeness of the DR can be proposed as below

[2] The middle-point of the death period may differ from the date of the census enumeration of population. The difference could be reduced by adjustments using various assumptions. Data sources: the estimated death rates are from the DCMD(www.lifetables.org), the registered death rates are from the Demographic Yearbook (https://unstats.un.org/unsd/demographic/products/dyb/default.htm)

$$c_r(t) = \frac{\sum_{x=60}^{75} m_r(x,t)}{\sum_{x=60}^{75} m(x,t)} = \frac{\log[1-q_{or}(t)]}{\log[1-q_o(t)]}, \qquad (7)$$

where $q_{or}(t)$ and $q_o(t)$ are the reported and accurate old-age mortality at time *t*. The values of reported old-age mortality are calculated using the number of deaths from the DR, and the population number from the census.

Further, Cauchy's mean value theorem indicates that

$$c_r(t) = \frac{\sum_{x=60}^{75} m_r(x,t)}{\sum_{x=60}^{75} m(x,t)} = \frac{\sum_{x=60}^{75} d_r(x,t)/p_r(x,t)}{\sum_{x=60}^{75} d(x,t)/p(x,t)} = \frac{d_r(y,t)/p_r(y,t)}{d(y,t)/p(y,t)}, \quad 60 < y < 75. \ (8)$$

Furthermore, noticing that the census enumeration may suffer undercount ($p_r(y,t) \le p(y,t)$), and that the completeness is implicitly assumed to be constant over age, the reliable estimate of completeness is an upper limit of the true completeness:

$$c_r(t) = \frac{\sum_{x=60}^{75} m_r(x,t)}{\sum_{x=60}^{75} m(x,t)} = \frac{d_r(y,t)}{d(y,t)} \cdot \frac{p(y,t)}{p_r(y,t)} \ge \frac{d_r(y,t)}{d(y,t)} = c(t). \qquad (9)$$

When the census undercounts are small or negligible compared to the incompleteness of the DR, the reliable estimate of completeness, $c_r(t)$, would be close to the true completeness, $c(t)$. On the other hand, when the census undercount rates are larger than that of the DR ($\frac{p(y,t)-p_r(y,t)}{p(y,t)} > \frac{d(y,t)-d_r(y,t)}{d(y,t)}$ or $\frac{p_r(y,t)}{p(y,t)} < \frac{d_r(y,t)}{d(y,t)}$), $c_r(t)$ will be bigger than 1.

Finally, the average completeness of the DR is obtained as:

$$\bar{c}_r = \frac{\sum_{t=1}^{n} c_r(t)}{n}. \qquad (10)$$

Complete DRs can be defined as $\bar{c}_r = 1$. This definition, however, is not useful in practice, because even if the DR was indeed complete, random errors in registration and census would make $\bar{c}_r$ departure from one. To take random errors into account, a probabilistic model is proposed below.

(2.3) A probabilistic model of completeness and hypothesis test

Among the population enumerated in a census, each person may or may not die within the census year, and each death may or may not be registered. The surviving and registering processes of each person can be described by a random variable that takes value 1 for dead and registered and 0 otherwise. These individual random variables are assumed to be independent from each other and follow the same probability distribution. Denoting the sum of these individual variables (which represents the registered deaths, also being a random variable) by *Dr(x,t)* (hereafter all random variables are expressed in capital letters), the probabilistic completeness of the DR can be defined as

$$C_r(t) = \frac{\sum_{x=60}^{75} D_r(x,t) / p_r(x,t)}{\sum_{x=60}^{75} d(x,t) / p(x,t)} = \frac{\sum_{x=60}^{75} M_r(x,t)}{\sum_{x=60}^{75} m(x,t)}. \qquad (11)$$

Further, when number of deaths is about 30 or larger, the central limit theorem guarantees that the probability distribution of $C_r(t)$ will be approximately normal (see Agresti and Finlay 1997: 104). Assuming that, for *t=1,2,…n,* $C_r(t)$ are independent from each other and follow the same normal distribution, it is intuitive that the average of $C_r(t)$, namely $\bar{C}_r$, is more robust than $C_r(t)$. Denoting the mean and variance of $C_r(t)$ by μ and σ^2, the unbiased statistics of the mean and variance of $C_r(t)$ are

$$\begin{aligned} \bar{C}_r &= \frac{\sum_{t=1}^{n} C_r(t)}{n}, \quad E(\bar{C}_r) = \mu, \\ S^2 &= \frac{\sum_{t=1}^{n} [C_r(t) - \bar{C}_r]^2}{n-1}, \quad E(S^2) = \sigma^2. \end{aligned} \qquad (12)$$

If the data on death registration is available only at one census year, or *n=1*, the variance of $C_r(t)$ cannot be estimated.

Since the variance of the $\bar{C}_r$ is

$$\mathrm{var}(\bar{C}_r) = \mathrm{var}\left(\frac{\sum_{t=1}^{n} C_r(t)}{n}\right) = \frac{\sum_{t=1}^{n} \mathrm{var}[C_r(t)]}{n^2} = \frac{\sum_{t=1}^{n} \sigma^2}{n^2} = \frac{\sigma^2}{n}, \qquad (13)$$

the unbiased statistic of the variance of $\bar{C}_r$ is

$$S_C^2 = \frac{S^2}{n}, \quad E(S_C^2) = E(\frac{S^2}{n}) = \frac{\sigma^2}{n} = \mathrm{var}(\bar{C}_r). \qquad (14)$$

Using $\bar{C}_r$ and S_C, a statistic for the hypothesis test is obtained:

$$T(n) = \frac{1-\bar{C}_r}{S_C} = \frac{1-\bar{C}_r}{S\sqrt{n}}. \qquad (15)$$

The sample values of $C_r(t)$ and $\bar{C}_r$ are the $c_r(t)$ and $\bar{c}_r$ in (9) and (10), respectively. Using these sample values, the sample value of *T* can be calculated by (12)-(15), and the hypothesis test below can be carried out.

If the mean of *T(n)* is 0, *T(n)* should follow a t-distribution with *(n-1)* degrees of freedom (see Agresti and Finlay 1997: 180). Accordingly, the hypotheses can be set as

$$Ho: \ mean(\bar{C}_r) = 1, \quad Ha: \ mean(\bar{C}_r) < 1, \qquad (16)$$

of which the null hypothesis implies that the completeness is 1 or perfect, and the alternative hypothesis represents that the completeness is smaller than 1. When the sampled value of *T(n)* is bigger than the t-score value with the corresponding *n* in Table 1, the probability for such a sample to occur is smaller than 0.05. Subsequently, the null hypothesis should be rejected, inferring that the completeness is significantly smaller than 1.

Table 1. The t-score with right-trail probability equal to 0.05

n	2	3	4	5	6	7	8	9	10
t-score	6.31	2.92	2.35	2.13	2.02	1.94	1.9	1.86	1.83

When the sample value of *T(n)* is smaller than the t-scores in Table 1 and the sample value of $\bar{C}_r$ is close to 1, the null hypothesis cannot be rejected, inferring that the difference between the mean of $\bar{C}_r$ and 1 is not statistically significant, or that the DR is probabilistically complete. In other words, a probabilistically complete DR is defined as one that the deterministic completeness is close to 1, and the difference between 1 and the mean of probabilistic completeness is statistically insignificant. A deterministic completeness is a sample of the probabilistic completeness and contains a random error. Probabilistic completeness takes random errors into account. When the difference between 1 and the mean of probabilistic completeness is statistically insignificant, the DR is judged as probabilistically complete.

When the sample value of *T(n)* is smaller than the t-score in Table 1, but the sample value of C_r is not close to 1, the null hypothesis cannot be rejected. This situation could be caused by two factors. The first is that *n* is small, leading to large $\mathrm{var}(C_r)$. The second is that the difference between sample values of $C_r(t)$

is large, which implies assuming that the same distribution of $C_r(t)$ is improper, and which also leads to large $\text{var}(C_r)$. In this context, the hypothesis test fails.

When the sample value of *T(n)* is negative or the sample value of $\bar{C}_r$ is bigger than 1, the alternative hypothesis should be set as $Ha: \quad mean(\bar{C}_r) \neq 1$, which requires the t-score of two-tail probability in Table 2.

Table 2. The t-score with two-trail probability equal to 0.05

n	2	3	4	5	6	7	8	9	10
t-score	12.71	4.31	3.18	2.78	2.57	2.45	2.37	2.31	2.26

If the absolute sample value of *T(n)* is smaller than the t-score with corresponding *n* in Table 2, the null hypothesis cannot be rejected, inferring that the difference between the completeness and 1 is not statistically significant, or that the DR is probabilistically complete.

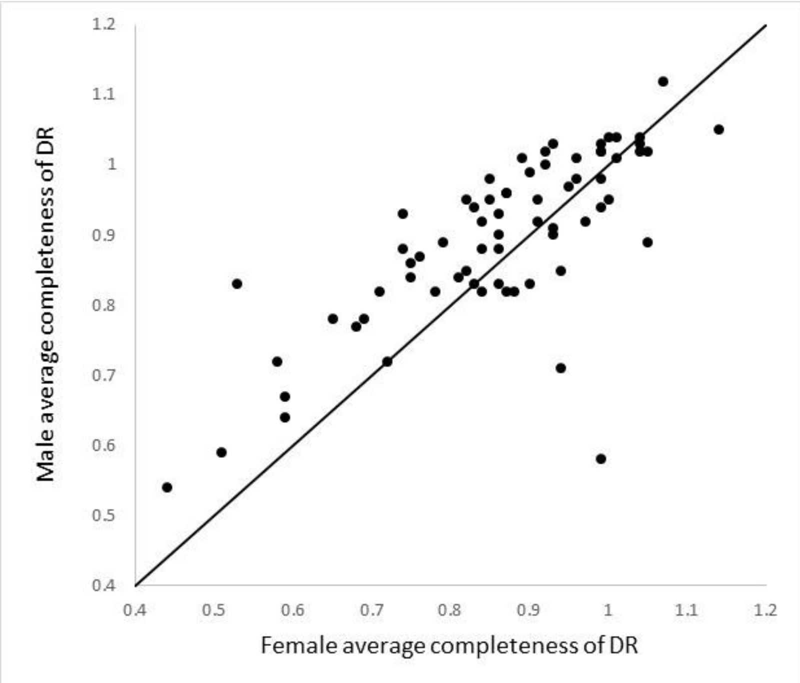
However, there could be a situation in which the sample value of *T(n)* is negative, and the absolute sample value of *T(n)* is bigger than the t-score with corresponding *n* in Table 2. A plausible reason for this situation to occur is that the numbers of the census population are less accurate than the number of deaths registered, which would lead to an overestimate of $m_r(x,t)$. In this situation, the DR is still probabilistically complete, but the undercount of the census population is nonnegligible.

To summarize the probabilistic model, it is necessary to indicate that the survival process is uncertain. When the population (*p(x,t)*) survives according to the unbiased death rate (*m(x,t)*), $D_r(x,t)$ is a random variable with a mean of *d(x,t)=m(x,t)p(x,t)* (Li, 2015). The number of registered deaths, $d_r(x,t)$, is a sample value of the registered deaths, $D_r(x,t)$. Thus, $d_r(x,t)$ can be either larger or smaller than the mean of $D_r(x,t)$. Also, a $\bar{c}_r$ is a sample value of $\bar{C}_r$, and could be either larger or smaller than 1, even if the mean of $\bar{C}_r$ is 1. To illustrate this phenomenon quantitatively, the hypothesis test is used as a proper tool. Setting the null hypothesis as $mean(\bar{C}_r) = 1$, meaning that the DR is probabilistically complete, $\bar{C}_r$ should follow a t-distribution. When the $\bar{c}_r$ does not imply the occurrence of a rare event under the t-distribution, the null hypothesis cannot be rejected, and a conclusion would be that the DR is probabilistically complete.

(3) Applications

The probabilistic evaluation requires the data on deaths for more than one census year. Among the 122 countries in the DCMD, 69 have the data of the DR for two or more census years. Using (10), the values of the average completeness are calculated for these countries and are shown in Figure 1. These values refer to the periods that are within the year group 1970-2015 and are indicated in the DCMD (www.lifetables.org). The average completeness for men is higher than that for women for 50 out of the 69 countries, indicating that the deaths of men are more likely to be registered. About 70% of the average completeness of men and women of the 69 countries is below 0.95, justifying the necessity of establishing the DCMD.

Figure 1. Average completeness by sex



Note: Estimated death rates are from the DCMD (www.lifetables.org), and the census population and registered deaths are from the Demographic Yearbooks and the National Statistics Offices

Setting 0.95 as a threshold above which the deterministic completeness could be regarded as high, and below which the deterministic completeness could be regarded as moderate or low, the results of the probabilistic evaluation can be grouped into 4 categories as are shown in Table 3.

Table 3. Results of the probabilistic evaluation

Deterministic completeness	Category	Difference between 1 and the mean of probabilistic completeness	Interpretation
Bigger than 0.95	I	Significant	Evaluation failure
Smaller than 0.95	II	Significant	Probabilistically incomplete
Smaller than 0.95	III	Insignificant	Evaluation failure
Bigger than 0.95	IV	Insignificant	Probabilistically complete

Category I includes cases of $\overline{c}_r > 0.95$; the difference between the mean of $\overline{C}_r$ and 1 is statistically significant. This is the case when the probabilistic evaluation fails, and the result of undercounted census population enlarges $\overline{c}_r$.

Category II includes cases of $\bar{c}_r \le 0.95$; the difference between the mean of $\bar{C}_r$ and 1 is statistically significant. This is the case when the probabilistic and deterministic evaluations are consistent: the deterministic completeness is moderate or low, and the DR is probabilistically incomplete.

Category III includes cases when the difference between 1 and the mean of probabilistic completeness is insignificant, while the deterministic completeness is moderate or low. This is also the case when the probabilistic evaluation fails. The reasons could be that *n* (the number of data on death at census years) is small, or that $c_r(t)$ differ remarkably and are not samples of the same distribution. Among the 69 countries with $n \geq 2$, only 22 countries have data on deaths at census years ($n=2$). For the 44 male and female populations with $n=2$, 82% have an average completeness below of 0.95, but the difference between their average completeness and 1 is statistically insignificant. In other words, the probabilistic evaluation fails for 82% of the cases with $n=2$. Because using only two samples to calculate variance is unreliable, this result makes sense. For this reason, we apply probabilistic evaluation to cases with data on deaths for three or more census years, which include 47 countries, or 94 cases of men and women.

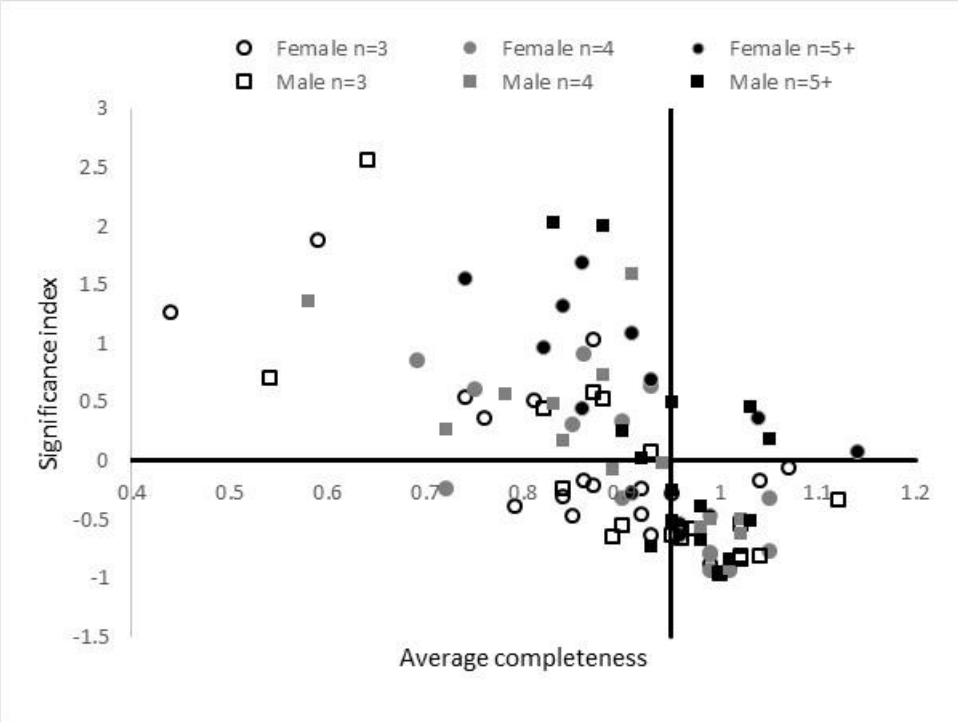
Category IV includes cases of $\bar{c}_r > 0.95$; the difference between the mean of $\bar{C}_r$ and 1 is statistically insignificant. These are probabilistically complete cases.

To describe the results of the probabilistic evaluation, we introduce a significance index as

$$s_i = \frac{t(n) - t_s}{t_s}, \qquad (17)$$

where $t(n)$ and t_s represent sample value of $T(n)$ in (15) and the corresponding t-score in Tables 1 and 2, respectively. Statistical significance corresponds to $s_i > 0$, and vice versa. Moreover, a bigger $|s_i|$ indicates a smaller probability of mistakenly rejecting the null hypothesis, and therefore the more robust the conclusion of the hypothesis test. Using the significance index, the above four categories can be described by the four quadrants of a completeness-significance figure, in which the vertical axis (significance index) crosses the horizontal axis (average completeness) at 0.95, as can be seen in Figure 2. In other words, in the completeness-significance figure of this paper, the origin is (0.95,0), not (0,0). In a completeness-significance figure, quadrants II and IV include successful probabilistic evaluations, while quadrants I and III contain probabilistic evaluations that do not provide reasonable conclusion and require further analysis. The probabilistic evaluation is carried out for the men and women of the 47 countries (94 cases) that have DR data on three or more census years. The results of these probabilistic evaluations are summarized in Figure 2 and are shown in detail in the table in Annex.

Figure 2. Average completeness and statistical significance by sex

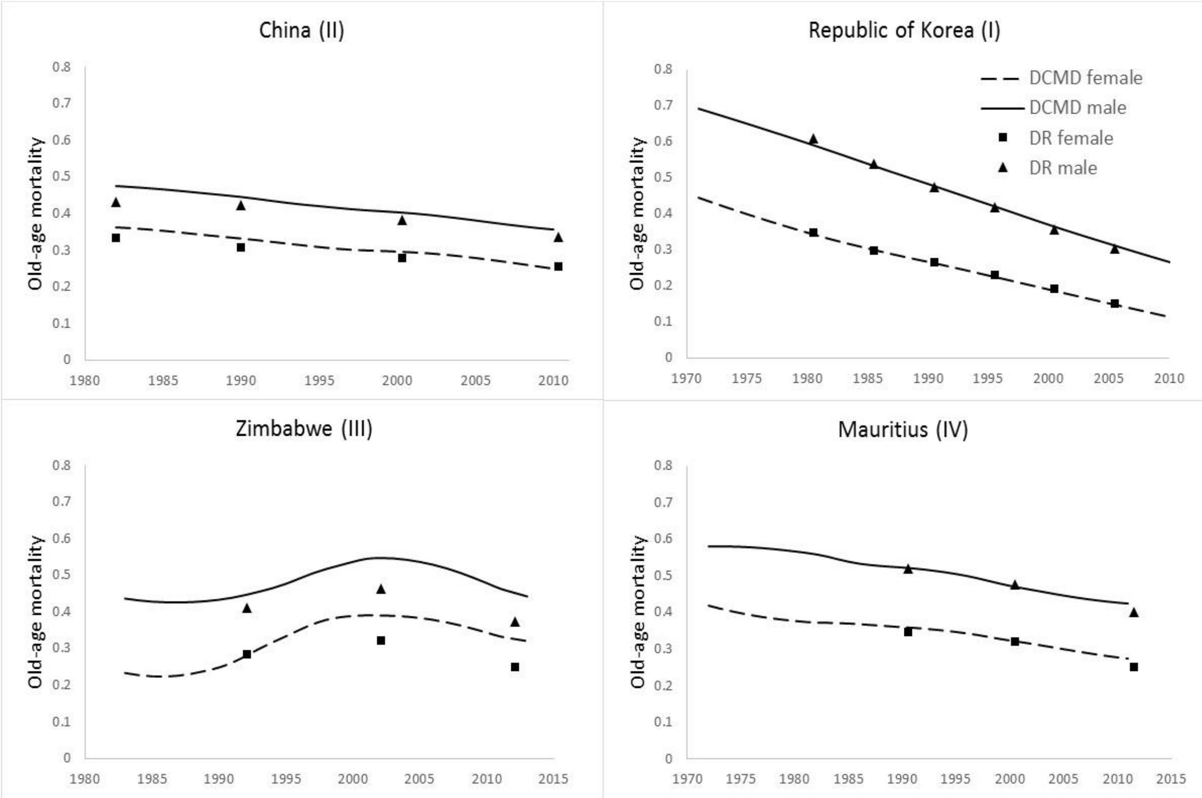


(4) Discussion

In the quadrant I of Figure 2, the average completeness is close to or unreasonably bigger than 1 but the difference between the mean of $\bar{C}_r$ and 1 is statistically significant (category I). Thus, the probabilistic evaluation fails to deliver reasonable conclusions. The 4 cases in quadrant I are for the men and women of the Republic of Korea (see Figure 3) and Seychelles. These cases can be explained as follows: the undercounting rates of old-age population in census are larger than the incompleteness of the DR, which could be common among developed countries. Undercounting old-age population by the same rate at successive censuses does not change the DCMD estimates (the denominator of $c_r(t)$), but it will raise the registered old-age mortality (the numerator of $c_r(t)$). This analysis immediately explains the nonsensical $c_r(t) > 1$, which may still be acceptable if though not statistically significant. Noticing that an enlarged $c_r(t)$ will raise the $|T|$ in (15) to be bigger than its t-score, this analysis further explains why the nonsensical $c_r(t) > 1$ is statistically significant. Assuming that the censuses of the two countries undercounted old-age population by 1.5%, and correspondingly adjusting upward old-age populations by 1.5%, the results would be in quadrant IV in which the DR is probabilistically complete, because the difference between the mean of $\bar{C}_r$ and 1 is statistically insignificant.

There are 37 (40% of the 94) cases in quadrant II (or category II), of which the average completeness is smaller than 0.95 (differs from 1 notably) and the difference is statistically significant. Thus, the probabilistic evaluation concludes that the DR is probabilistically incomplete. China is chosen to provide an example for Category II as is shown in Figure 3. The average completeness of China is 0.93 for females and 0.91 for males.

Figure 3. Old-age mortality and the result of probabilistic evaluation for selected countries



Notes: Same as figure 1.

Despite limiting the probabilistic evaluation to *n>2*, there are still 17 (18% of the 94) cases in quadrant III (category III). For 12 of the 17 cases, the *n* is 3, indicating *n=3* is not big enough to guarantee successful probabilistic evaluation. The other 5 cases in category III are perhaps caused by large differences between sample values of completeness at various census years. In other words, assuming the same probability distribution for the completeness at different census years is not proper. The requirement of having DR data for three or more census years is valid, but it could be removed if the DR and reliable data on old-age population exist in single years between two successive censuses. Zimbabwe provides the values of old-age mortality as an example for quadrant III in Figure 3. The average completeness of Zimbabwe is 0.84 for females and 0.82 for males.

There are 36 (38%) cases in quadrant IV category. These are the cases of probabilistically complete: $\bar{c}_r \geq 0.95$, where the difference between the mean of $\bar{C}_r$ and 1 is statistically insignificant. These results indicate that setting 0.95 as a threshold of deterministic completeness ($\bar{c}_r$) is meaningful. When $\bar{c}_r \geq 0.95$, all conclusions are probabilistically complete. Of course, this observation is based on limited observations (94 cases). When investigating data from countries excluded here, a threshold different from 0.95 may be more suitable. Mauritius is selected to give an example in Figure 2 in category IV. The average completeness of Mauritius is 0.95 for females and 0.97 for males. Putting quadrants II and IV together, the probabilistic evaluation is successfully applied for 78% of the 94 cases. If small undercounting rates existed in the censuses of the Republic of Korea and Seychelles, successful applications would contain 82% of the 94 cases.

The concept of probabilistic completeness is useful. Because of random errors, deterministic completeness cannot be exactly 1 in practice. It is therefore impossible to conclude whether a DR is complete using deterministic completeness. Probabilistic completeness solves this dilemma: when deterministic completeness is high, and when the difference between 1 and the mean of probabilistic completeness is statistically insignificant, the DR is probabilistically complete. Furthermore, when the deterministic completeness is moderate or low, and when the difference between 1 and the mean of probabilistic completeness is statistically significant, the DR is probabilistically incomplete.

(5) References

Agresti, A, and B Finlay (1997). *Statistical Methods for the Social Sciences.* Upper Saddle River, NJ: Prentice Hall.

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

Bennett N G, and Horiuchi S (1984). Mortality estimation from registered deaths in less developed countries. *Demography*, 1984(21), 217-33. doi: 10.2307/2061041

Brass, W (1975). *Methods for Estimating Fertility and Mortality from Limited and Defective Data*. Chapel Hill: International Program of Laboratories for Population Statistics.

Hill, K (1987). Estimating Census and Death Registration Completeness. *Asian and Pacific Population Forum* 1(3):8-13: 23-24.

Hill, K, You, D, and Choi, Y (2009). Death distribution methods for estimating adult mortality: Sensitivity analysis with simulated data error. *Demographic Research* 21(9): 235-252. doi: 10.4054/DemRes.2009.21.9

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.

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 (2015). The probabilistic life table and its applications to Canada. *Canadian Studies in Population* 42 (1-2):117-129

Li, N and P Gerland (2017). Evaluating the Completeness of Death Registration for Developing Countries at Old Ages. Paper presented at the XXVIII IUSSP International Population Conference, 29 Oct-4 Nov 2017, Cape Town, South Africa.

Li, N, H Mi and P Gerland (2017). Using Child, Adult, and Old-age Mortality to Establish a Developing Countries Mortality Database (DMD). 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).

Li, N, H Mi, Gerland, C Li, and L Sun (2018). Establishing a Developing Countries Mortality Database (DCMD) on the Basis of Child, Adult, and Old-age Mortality. Paper presented at the 2018 annual meeting of the Population Association of America, 26-28 April 2018, Cape Town, South Africa.

Li, N and H Mi (2018). Single-year Estimates in Developing Countries Mortality Database (DCMD). To appear in Proceedings of the 18th Applied Stochastic Models and Data Analysis International Conference with Demographics Workshop.

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.

Annex table. Average completeness and results of probabilistic evaluation

Female					Male				
	Number of DR data at census years, n	Average completeness	Sample T	Category		Number of DR data at census years, n	Average completeness	Sample T	Category
Seychelles	7	1.14	-2.62	I	Seychelles	7	1.05	-2.93	I
Republic of Korea	8	1.04	-3.22	I	Republic of Korea	8	1.03	-3.48	I
El Salvador	3	0.81	4.43	II	Kazakhstan	3	0.93	3.16	II
Kuwait	5	0.74	5.42	II	Zimbabwe	3	0.82	4.24	II
Azerbaijan	3	0.76	3.98	II	Paraguay	4	0.72	3	II
Kyrgyzstan	3	0.74	4.5	II	Sri Lanka	4	0.83	3.51	II
Mali	3	0.44	6.64	II	Bahamas	5	0.92	2.2	II
Peru	3	0.59	8.43	II	Azerbaijan	3	0.87	4.64	II
China	4	0.93	3.87	II	Kyrgyzstan	3	0.88	4.49	II
Cuba	4	0.86	4.5	II	Mali	3	0.54	4.99	II
Dominican Republic	4	0.69	4.38	II	Peru	3	0.64	10.42	II
Ecuador	4	0.75	3.78	II	China	4	0.91	6.11	II

Panama	5	0.84	4.93	II	Cuba	4	0.88	4.08	II
Thailand	5	0.86	5.73	II	Dominican Republic	4	0.78	3.7	II
Philippines	7	0.86	2.8	II	Ecuador	4	0.84	2.77	II
Brazil	5	0.91	4.44	II	Panama	5	0.88	6.4	II
Viet Nam	3	0.87	5.97	II	Thailand	5	0.83	6.46	II
Argentina	4	0.9	3.14	II	Philippines	7	0.9	2.44	II
Egypt	4	0.85	3.08	II	Qatar	4	0.58	5.57	II
South Africa	6	0.82	3.97	II	Brazil	5	0.95	3.21	II
India	5	0.93	3.61	II	Burkina Faso	3	0.89	1.05	III
Burkina Faso	3	0.79	1.82	III	Maldives	3	0.9	1.31	III
Maldives	3	0.93	1.1	III	El Salvador	3	0.84	2.25	III
Kazakhstan	3	0.86	2.45	III	Kuwait	5	0.93	0.58	III
Zimbabwe	3	0.84	2.03	III	Colombia	4	0.94	2.31	III
Paraguay	4	0.72	1.82	III	Bahrain	4	0.89	2.18	III
Sri Lanka	4	0.9	1.61	III	Botswana	3	0.95	1.1	IV
Bahamas	5	0.91	1.53	III	Tunisia	3	1	0.13	IV
Botswana	3	0.85	1.57	III	United Republic of Tanzania	3	0.96	1.24	IV
Tunisia	3	0.92	1.59	III	Viet Nam	3	0.96	1	IV
United Republic of Tanzania	3	0.87	2.34	III	Argentina	4	0.99	1.21	IV
St. Vincent and the Grenadines	3	0.92	2.23	III	Egypt	4	0.98	1.04	IV
Colombia	4	0.99	0.5	IV	South Africa	6	0.95	1.01	IV
Qatar	4	0.99	0.18	IV	Mauritius	3	0.97	1.23	IV
Mauritius	3	0.95	2.13	IV	Singapore	5	0.95	1.61	IV
Singapore	5	1	0.08	IV	Venezuela	5	0.98	1.31	IV
Venezuela	5	0.99	1.13	IV	Mexico	6	0.98	0.67	IV
Mexico	6	0.96	0.74	IV	St. Vincent and the Grenadines	3	1.02	-0.81	IV
Malaysia	3	0.99	0.38	IV	India	5	1.03	-1.37	IV
Chile	4	0.99	0.53	IV	Malaysia	3	1.02	-2.01	IV
Uruguay	5	0.96	0.98	IV	Chile	4	1.02	-1.6	IV
Bahrain	4	1.05	-0.74	IV	Uruguay	5	1.01	-0.47	IV
Barbados	3	1.04	-3.59	IV	Barbados	3	1.02	-0.7	IV
Mongolia	3	1.01	-0.35	IV	Mongolia	3	1.04	-0.83	IV
Saint Lucia	3	1.07	-4.08	IV	Saint Lucia	3	1.12	-2.91	IV
Costa Rica	4	1.01	-0.23	IV	Costa Rica	4	1.01	-0.29	IV
Trinidad and Tobago	4	1.05	-2.2	IV	Trinidad and Tobago	4	1.02	-1.24	IV

Note: Same as Figure 1.