# Using the Equivalent Construction to Estimate Complete Life tables for the Developing Countries Mortality Database (DCMD)

Nan Li[1] and Hong Mi[2]

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).

*The Developing Countries Mortality Database (DCMD, www.lifetables.org) provided abridged life tables that use five-year age group. But most population-related statistics and programs are using single-year age. To extend an abridged life table to a complete life table that uses single-year age, earlier methods choose a life-table function to extend to single-year age, and then calculate the complete life table. The essential problem of these methods is the arbitrariness of choosing a life-table function to extend. A recently proposed method of equivalent construction extends a whole life table instead of an individual life-table function. This paper reports an application of equivalent construction, which extends all DCMD abridged life tables to complete life tables. In this application, the targets of preserving the age patterns of the abridged life table and satisfying the smoothness of the complete life table are reached simultaneously for more than 97% life tables in DCMD.*

## Introduction

Most population-related statistics and programs now are using single year of age and time. Examples may include, but not limited to, birth and death registrations, immunization programs, school enrollment and employment statistics, healthcare programs, and pension systems. In most population estimates and projections, however, the time interval and age group are still five years (e.g., United Nations Population Division, 2019). Abridged life tables are calculated more often than are complete life tables for many reasons. For example, age heaping in censuses are common, especially in developing countries. To reduce age-heaping errors, grouping single-year age into 5-year age group could be a solution, which leads to calculating abridged life tables. For another example, deaths are rare at some age in populations with low mortality or limited size. Consequently, even if deaths were perfectly registered and populations were ideally counted or estimated, complete life tables would contain large random fluctuations and abridged life tables are preferred. Subsequently, the commonly used model life tables (MLTs) all provide abridged life tables. These MLTs include the one-input-parameter MLTs that use typically child mortality as input (Coale and Demeny, 1966; United Nations Population Division, 1982), the two-input-parameter MLTs that use child and adult mortality as input (Murray and others 2003; Wilmoth, J.R. and others, 2012), and the three-input-parameter MLTs in DCMD that use child, adult and old-age mortality as input. Recently, a two-input-parameter MLT, which directly models complete life tables as input and provides complete life tables as output, is developed by Clark (2019).

The DCMD, although used a three-input-parameter MLT, provided abridged life tables that use five-year age group, and that need to be extended into complete life tables with single-year age.

Previous methods of extending an abridged life table to a complete life table included applying interpolations (e.g., Elandt-Johnson and Johnson,1980) and parameter models (e.g, Heligman and Pollard 1980) on a life-table function of an abridged life table. These methods have been reviewed, for example, by Kostaki and Panousis (2001). A common feature of these methods is they all choose only one life-table function (such as death rate, the probability of death, or the survivors by age) as input to extend to single-year age, and then calculate the complete life table based on the extended life-table function. Arbitrarily choosing a life-table function to extend is the essential problem of these methods.

In a recent study (Li, 2019), an equivalent construction of life table is proposed. The first step of this method is to reduce an observed unsmooth complete life table to a smooth abridged life table, which can use five-year, ten-year, or flexible age groups. In the second step, the abridged life table is used as input to equivalently construct a complete life table, which refers to that, using the complete life table to compute an abridged life table, the result is identical to the original abridged life table for every life-table function at each age. If the equivalently constructed complete life table is not smooth enough, a construction-based graduation is used as the third step. Although this method is not designed to extend abridged life tables, it can be used to do so by using only the second and third steps. Since the equivalent construction extends a whole life table instead of a single life-table function, it avoids the essential problem of earlier methods, and is successfully used to estimate complete life tables for the DCMD by extending abridged life tables.

## Equivalent construction

In (Li, 2019), the computational structure of a life table is analyzed. It concludes that a life table can be exactly computed when two independent life-table functions are known. It chooses, for the reason of simplicity, the function of survivors at age *x* ($l_x$) and the function of person-years between ages x and x+n (${}_nL_x$) as the two independent life-table functions of an original abridged life table, as input, to construct a complete life table that is equivalent to the original abridged life table. Here the 'equivalent' refers to using the constructed life table to compute an abridged life table, the result is identical to the original abridged life table for every life-table function at each age.

To reach the equivalence of construction, it is apparent that the number of survivors of the complete life table must equal to that of the abridged life table at each abridged age ($x_a$),

$$l_x = l_{x_a}, \quad x = x_a. \qquad (1)$$

To determine the values of function $l_x$ at ages between the successive abridged ages, the law of large numbers and the variation of functionals are in need. According to the law of large numbers, function $l_x$ should change with age *x* smoothly if the population size is large. Consequently, $l_x$ can be required as the smoothest function of age *x*. Subsequently, using the variation to minimize functional $\int_{x_a}^{x_a+n} [\frac{d^2 l_x}{dx^2}]^2 dx$ (which is to maximize the smoothness of $l_x$), it specifies $l_x$ as a quadratic function,

$$l_x = a + b \cdot (x - x_a) + c \cdot (x - x_a)^2. \qquad (2)$$

Since for each abridged age group there are three input data ($l_{x_a}$, $l_{x_a+n}$, and ${}_nL_{x_a}$), the three parameters (*a, b*, and *c*) in (2) are solved from using an integral to describe ${}_nL_{x_a}$:

$$\begin{cases} l_{x_a} = a, \\ l_{x_a+n} = a + n \cdot b + n^2 \cdot c, \\ {}_nL_{x_a} = \int_{x_a}^{x_a+n} l_x dx = n \cdot a + \frac{n^2}{2} b + \frac{n^3}{3} c. \end{cases} \qquad (3)$$

In practice, for $x>0$ the ${}_1a_x$ is usually not calculated from using an integral to describe ${}_nL_{x_a}$, but widely approximated as 0.5 (see Human Mortality Database, 2018), of which the accuracy can be indicated by the Greville (1943) formula. To be consistent with the practice procedure, a factor ($\alpha$) is used to proportionally adjust $l_x$ to make the sum of ${}_1L_x$ in an abridged age group equal ${}_nL_x$:

$$0.5 \cdot (l_{x_a} + \alpha \cdot l_{x_a+1}) + 0.5 \cdot (\alpha \cdot l_{x_a+1} + \alpha \cdot l_{x_a+2}) + ... + 0.5 \cdot (\alpha \cdot l_{x_a+n-1} + l_{x_a+n}) = {}_nL_x. \qquad (4)$$

Finally,

$${}_1L_x = {}_1a_x \cdot l_x + (1 - {}_1a_x) \cdot l_{x+1}, \qquad (5)$$

and all other functions of the constructed complete life table can be calculated. Using this complete life table to compute an abridged life table, the result is guaranteed to be identical to the original abridged life table for every life-table function at each age. It is important to note that the construction does not ignore the observed $l_x$ within the abridged age groups, because ${}_nL_{x_a} = \sum_{y=x_a}^{x_a+n} {}_1L_y$ is used.

## Construction-based graduation

When some functions of the equivalently constructed complete life table are not smooth enough, especially at the boundaries of successive abridged age groups, a graduation (or smooth) could be applied to a life-table function, which is a construction-based graduation. The logarithm of death rate by age is chosen to be graduated by using a local regression model, because death rate by age is usually the direct measure and a logarithm is often used to tell the detail changes of death rate by age. The span of the local regression is chosen as 0.2 to emphasize that the death rates by age are of the complete life table that is equivalent to the original abridged life table. Models other than local regression could also be chosen. Papaioannou and Sachlas (2004) investigated the performance of spline, weighted moving averages, the Whittaker and Henderson model, and many other models, and concluded that 'none of them can be thought as better or more correct, as they give almost the same results.'

Whether an equivalently constructed complete life table is or is not smooth enough, however, is a subjective judgment. This paper will not further discuss the judgment; and complete life tables of equivalent construction and construction-based graduation are provided to all the abridged life tables of DCMD. The difference between the life tables of construction-based graduation and equivalent construction is an adjustment on the latter. When the adjustment made by a construction-based graduation is small or moderate, the life table of construction-based graduation is approximately equivalent to the equivalently constructed life table, and thus also approximately equivalent to the original abridged life table.

## Application

An equivalently constructed complete life table is guaranteed to be equivalent to the original abridged life table. A life table of construction-based graduation is an adjustment on the equivalently constructed complete life table. We measure an adjustment using the average relative difference (Ard) between the life expectancies at ages 0, 15 and 60 years of the complete life tables of the equivalent construction and construction-based graduation,

$$Ard = 100 \cdot \left( \left| \frac{e_0^c - e_0^g}{e_0^c} \right| + \left| \frac{e_{15}^c - e_{15}^g}{e_{15}^c} \right| + \left| \frac{e_{60}^c - e_{60}^g}{e_{60}^c} \right| \right) / 3 , \qquad (6)$$

where $e_x^c$ and $e_x^g$ are the life expectancy at age x of the complete life table of equivalent construction and construction-based graduation, respectively. In DCMD, life tables are estimated using empirical data on the probability of dying between ages 0 and 5, 15 and 60, and 60 and 75 years. Thus, the starting ages 0, 15 and 60 years are chosen in (6), which are also relevant to studies such as population ageing.

Applying the equivalent construction and the construction-based graduation to the 7712 abridged life tables in DCMD, the distribution of Ardis descripted in Figure 1. The values of Ard are smaller than 0.3% for 97% of the 7712 life tables in DCMD. Figure 2 displays the logarithms of death rate by age of Malawi men in 2001 with an Ard about 0.3%. In Figure 2, the black circles represent the logarithms of death rate of the original abridged life table, the white circles stand for that of the equivalently constructed life table, and the line displays that of the construction-based graduation. Because the white circles are constructed equivalently to the black circles, all black circles are overlapped with the white circles at the middle point of abridged age groups, and in line with the white circles at other ages. The adjustment, or the difference between the white circles and the line, is large at only one age and small at all other ages, leading to an Ard as small as 0.3%. We describe the life table of the construction-based graduation, with an Ard of 0.3% or smaller, as approximately equivalent to the equivalently constructed complete life table, and thus also approximately equivalent to the original abridged life table.

Figure 1. Number of life tables by the range of the average relative difference (Ard)

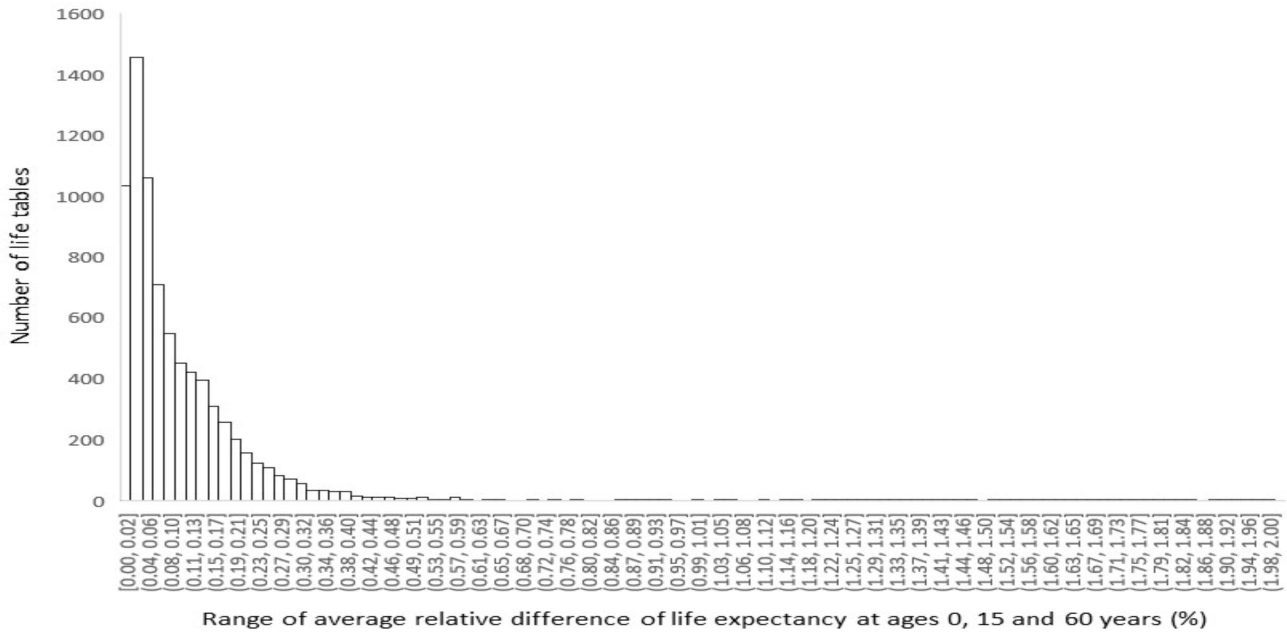
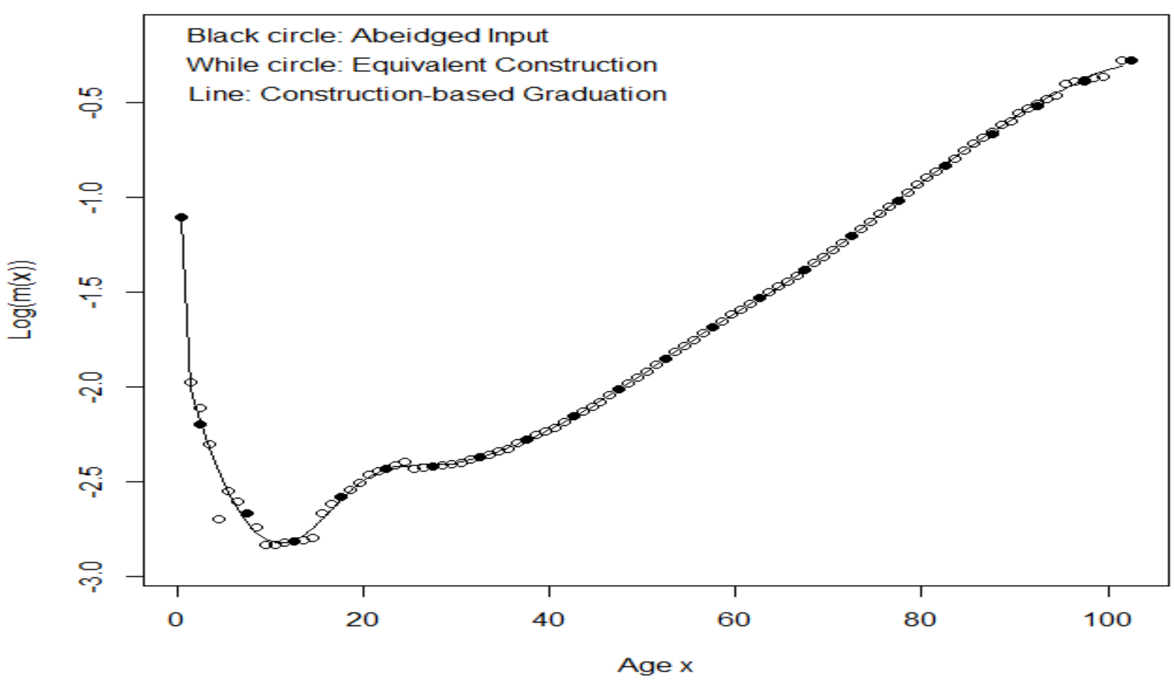
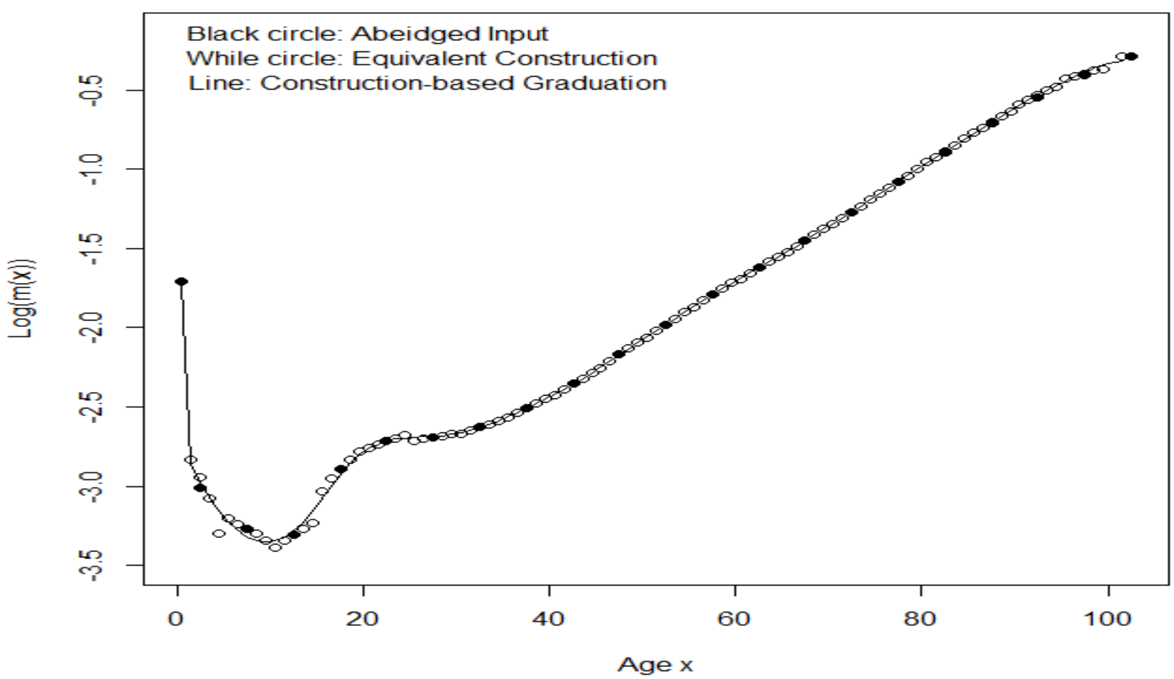



Figure 2. Logarithm of death rate by single-year and five-year age groups, Malawi men, year 2001.



It should be indicated that, when the Ard is remarkably smaller than 0.3%, the adjustment should be notably smaller than that in Figure 2, and the life table of the construction-based graduation should be closer to perfectly equivalent to the original abridged life table. The average Ard value of the 7712 in DCMD life tables is 0.1%, significantly smaller than 0.3%. The Ard of Sri Lanka men in year 1993 is about 0.1%. The logarithms of death rate by age of Sri Lanka men in year 1993 are depicted in Figures 3, in which we see the adjustment is moderate at only one age and small at all other ages, indeed notably smaller than that in Figure 2. Nonetheless, we do not use more complicated terms to distinguish the difference between the adjustments Malawi and Sri Lanka. We choose to concisely describe that 97% life tables of the construction-based graduation in the DCMD are approximately equivalent to the original abridged life tables.

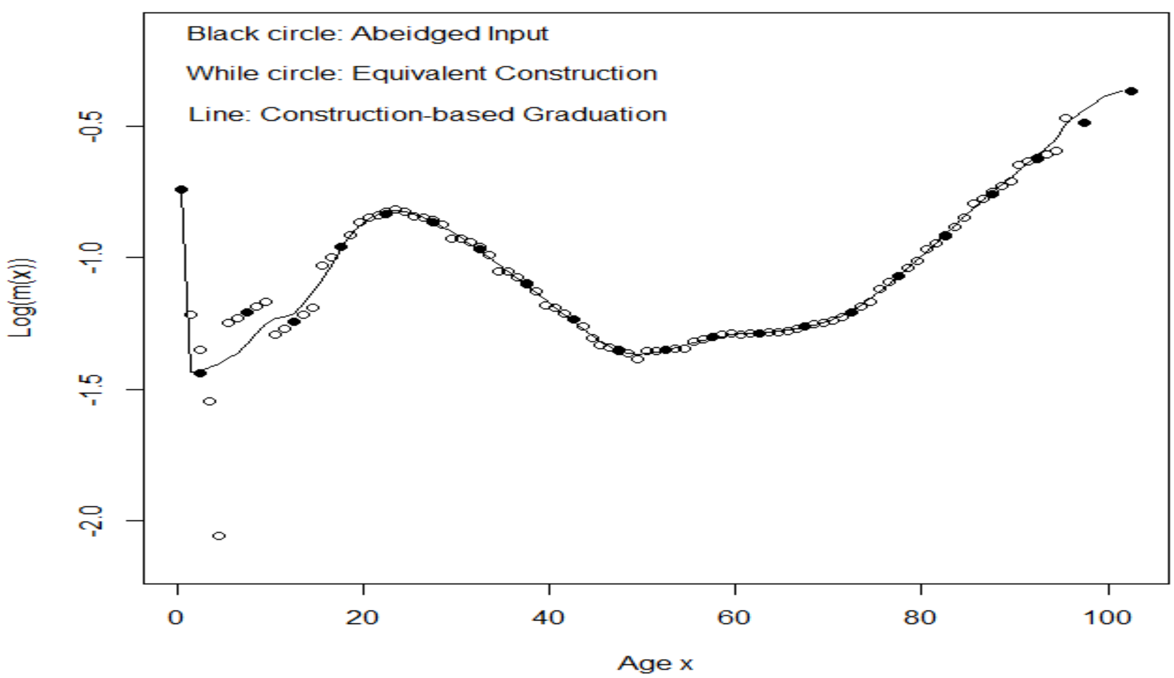
Figure 3. Logarithm of death rate by single-year and five-year age groups, Sri Lanka men, year 1993.




Among the 3% outliers in Figure 1 whose Ard are bigger than 0.3%, the one with the biggest Ap, about 2%, is observed from the female life table of Rwanda in 1994, in which an unprecedented genocide happened. Figure 4 displays the logarithms of death rate by age of

Rwanda women in 1994. Although the age pattern of death rate is extremely unusual, all black circles are still overlapped and in line with the white circles, for the reason same as that indicated for Malawi. The adjustment is still small at most ages, but moderate at ages younger than 15 and extremely large at age 5 years, leading to an Ard as large as 2%. We describe the life table of the construction-based graduation, with an Ard significantly larger than 0.3%, as notably different from the equivalently constructed complete life table. This description, however, apparently does not apply to the life tables that have an Ard slightly bigger than 0.3%. For this reason, it would be more proper to summarize that in the application of using equivalent construction to extend abridged life tables in DCMD, more than 97% of the construction-based complete life tables are approximately equivalent to, and other less than 3% are notably different from, the equivalently constructed complete life table.

Figure 4. Logarithm of death rate by single-year and five-year age groups, Rwanda women, year 1994.

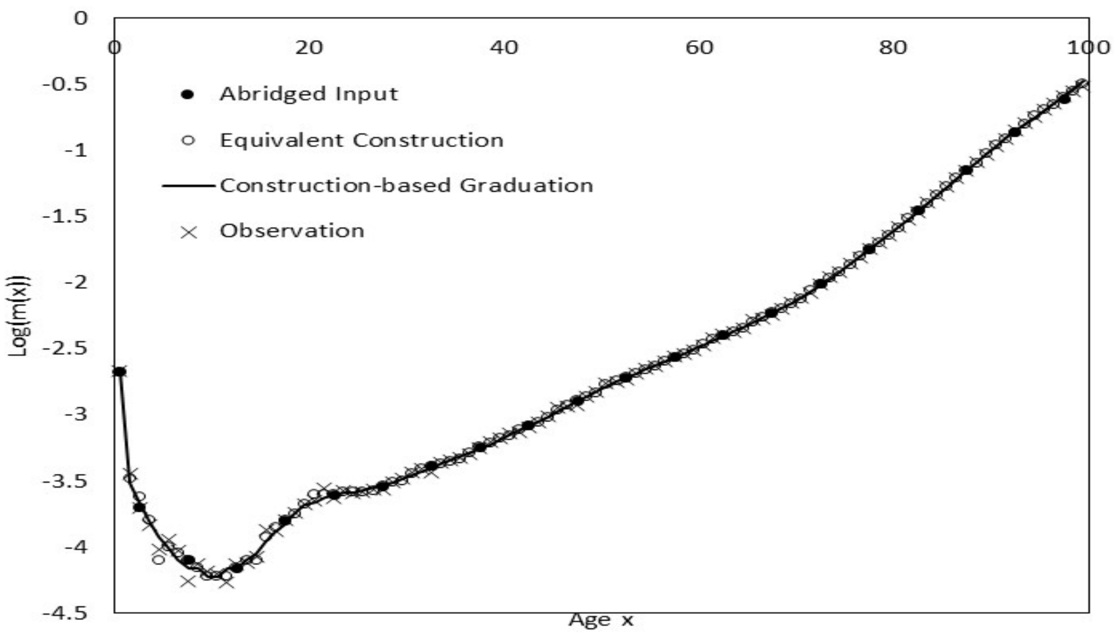


## Discussion and summary

When extending an abridged life table to a complete life table, there are two targets. The first is to preserve the age patterns of the original abridged life table, and the second target is to satisfy the smoothness of the functions of the complete life table. An equivalently constructed complete life table will perfectly reach the first target, which had not been reached by earlier methods. When the second target is not satisfactorily reached, however, a construction-based graduation will apply and will make an adjustment on the equivalently constructed complete life table. When the equivalently constructed complete life table is smooth or when the adjustment small or moderate, in practice the two targets are reached simultaneously.

We should indicate, however, if there is an observed complete life table, it will be different from the equivalently constructed one unless the population size is infinitely large. The difference is understandable in view of the probabilistic life table (Li, 2015) that is built on the uncertain survival process of a hypothetic cohort with a specific size of birth. An example of this difference is given by the data of Japan in figure 5 (Haman Mortality Database, 2019), in which the crosses are the observed single-year values of the Logarithm of death rate that are different from the equivalently constructed values.

Figure 5. Logarithm of death rate by single-year and five-year age groups, Japan women, year 2010.




In a traditional life table, assuming the true value of the probability of dying between ages x and x+1 is ${}_1q_x$, the number of deaths in this age interval is deterministically computed as

$$ {}_1d_x = l_x \cdot (1 - {}_1q_x) . \qquad (7) $$

In reality, however, ${}_1d_x$ is a random number that can be modeled by a binomial distribution B($l_x$, ${}_1q_x$), and that could occur as any integer from 0 to $l_x$ and differ from the deterministically computed ${}_1q_x$. Thus, given the $l_x$ and $l_{x+5}$ of an abridged life table, the ${}_1d_{x+i}$ for $i = 0 : 4$ need not to obey (7), and may take various sets of integers to satisfy

$$ \sum_{i=0}^{4} {}_1d_{x+i} = l_x - l_{x+5} , \qquad (8) $$

In other words, there are multiple complete life tables that correspond to different sets of ${}_1d_{x+i}$ and are all equivalent to the original abridged life table as long as (8) is satisfied; and the observed complete life table is just one sample of the multiple equivalent complete life tables that corresponds a special set of ${}_1d_{x+i}$. The difference between the sample and true value of the probability of death (and other life-table function) is the effect of random disturbance, which is smaller when population size $l_x$ is larger:

$$\text{var}({}_1q_x) = \text{var}(\frac{{}_1d_x}{l_x}) = \frac{\text{var}({}_1d_x)}{l_x^2} = \frac{l_x \cdot {}_1q_x \cdot (1 - {}_1q_x)}{l_x^2} = \frac{{}_1q_x \cdot (1 - {}_1q_x)}{l_x}. \quad (9)$$

The random disturbances could be minimized, and the true values of ${}_1q_x$ could be obtained when the population size is infinitely large according to (9). Since smooth life-table functions are often observed in large populations, smooth complete life-table functions are defined as the true values and are used as the target of the equivalent construction (Li, 2019). Consequently, if there is an unsmooth observed complete life table, it should be a result of random disturbance due to limited population size and should be replaced by that of the smooth equivalent construction. Of course, if the observed complete life table is satisfactorily smooth (such as that of Japan with a big number of population), it would be close to that of the equivalent construction as can be seen in figure 5 (with an Ard of 0.02%) and the replacement would be unnecessary.

Although the equivalent construction aims to provide smooth complete life tables, some of the functions of the complete life table may not be smooth, typically at only a few ages at the boundary of successive abridged age groups because the original abridged life table is unsmooth between these few age groups. In this situation, a construction-based graduation is applied to the equivalently constructed complete life table.

In the construction-based graduation, choosing the logarithm of death rate instead of another life-table function is, in fact, arbitrary. Nonetheless, when the functions of the original abridged life table are smooth, the functions of the equivalently constructed complete life table should also be smooth except at only few ages at the boundary of successive abridged age groups. Consequently, the construction-based graduation could make only small or moderate adjustment at these few ages, on the function that is chosen to graduate, and on other functions that are affected by the graduation. Thus, regardless of which function is chosen to graduate, the adjustment would be small or moderate at only few ages, and the life table of construction-based graduation should be approximately equivalent to the equivalently constructed complete life table. In other words, the effect of the arbitrariness in the construction-based graduation should be small when the original abridged life table is smooth.

For the life table of Zambia men in year 2003, and the other more than 97% life tables in DCMD, the complete life tables of construction-based graduation are approximately equivalent to the original abridged life tables. Thus, the adjustments, and the effects of the arbitrariness in the construction-based graduation, are small for more than 97% life tables in DCMD.

The main reason of reaching approximate equivalence for more than 97% life tables in DCMD is using the equivalent construction, which guarantees perfect equivalence. Another reason is that most abridged life tables in DCMD are smooth, which limits the moderate adjustments to appear at only few ages.

When the original abridged life table is not smooth for unusual reasons, the adjustment could be large at a few ages, and the complete life table of construction-based graduation would differ notably from the original abridged life table. The life table of Rwanda women in year 1994, and the other less than 3% life tables in DCMD, belong to this situation.

The two targets, namely preserving the age patterns of the abridged life table and satisfying the smoothness of the complete life table, are reached simultaneously for more than 97% of the life tables in DCMD. For the other less than 3% life tables in DCMD, the age patterns of the abridged life table could not be preserved when the smoothness is satisfied. Overall, using equivalent construction to extend all DCMD abridged life tables to complete life tables is successful.

## References

Coale, A. J., & Demeny, P. (1966). Regional model life tables and stable populations. Princeton, NJ: Princeton University Press.

Elandt-Johnson, R., and Johnson, N. 1980. Survival Models and Data Analysis. New York, John Wiley.

Greville, T. 1943. Short methods of constructing abridged life tables. The Record of the American Institute of Actuaries, (32)29-43.

Heligman, L., and Pollard, J. H. 1980. The age pattern of mortality. Journal of the Institute of Actuaries 107 (1): 49–80.

Haman Mortality Database, 2019. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de.

Kostaki, A and Panousis, V. 2001. Expanding an abridged life table. Demographic Research 5(1):1-21.

Li, N. 2019. Estimating Complete Life Tables for Populations with Limited Size: From Graduation to Equivalent Construction. North American Actuarial Journal, DOI: 10.1080/10920277.2019.1585879.

Li, N. 2015. The probabilistic life table and its applications to Canada. *Canadian Studies in Population* 42, no. 1–2 (2015): 117–129.

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.

Papaioannou, T and A Sachlas. 2004. Graduation of mortality rates revisited. *Manuscript*. Available at (https://www.researchgate.net/publication/228386032)

Clark, S.J, 2019. A General Age-Specific Mortality Model With an Example Indexed by Child Mortality or Both Child and Adult Mortality. Demography, Demography (2019) Vol. 56, pp. 1131–1159.

United Nations Population Division. 2019. World population prospects: The 2019 revision. New York: United Nations.

United Nations Population Division. 1982. *Model life tables for developing countries*. New York: United Nations

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.