



Max-Planck-Institut für demografische Forschung
Max Planck Institute for Demographic Research
Konrad-Zuse-Strasse 1 · D-18057 Rostock · GERMANY
Tel +49 (0) 3 81 20 81 - 0; Fax +49 (0) 3 81 20 81 - 202;
<http://www.demogr.mpg.de>

MPIDR TECHNICAL REPORT 2010-005
JUNE 2010

**Spreadsheet for calculation of
confidence limits for any life table or
healthy-life table quantity**

Evgeny M. Andreev (andreev@demogr.mpg.de)
Vladimir M. Shkolnikov (shkolnikov@demogr.mpg.de)

© Copyright is held by the authors.

Technical reports of the Max Planck Institute for Demographic Research receive only limited review. Views or opinions expressed in technical reports are attributable to the authors and do not necessarily reflect those of the Institute.

Spreadsheet for calculation of confidence limits for any life table or healthy-life table quantity

by Evgeny M. Andreev and Vladimir M. Shkolnikov

Abstract

In this report, we use Monte Carlo simulations to estimate the confidence limits for survival and death probabilities, life expectancy, and healthy life expectancy, as well as any other quantity based on a conventional life table or on Sullivan's healthy life table. Two Excel spreadsheets for use with this method are provided.

Background

Life-table measures computed from empirical data can be seen as deterministic or random variables. The deterministic approach is by far more popular in demography. It assumes that the population under consideration is a general universe, and that life expectancy and other life table quantities precisely describe the true mortality regime in this population. However, studies on mortality in small populations tend to produce shaky values for life-table functions. These inexplicable fluctuations require researchers to apply the stochastic approach by estimating the likely magnitude of stochastic error.

The growth in the use of Sullivan's (1971) method for estimating health expectancy, which involves combining the life table with survey-based prevalences of poor health, calls for an even greater degree of statistical estimation. Indeed, estimates of health expectancy can be imprecise even in large populations if the prevalence of ill-health origins from samples consist of a few thousand respondents. In this case, stochastic errors in age-specific prevalence of poor health (disability, chronic illness) can be important.

One of the most widely used approaches for addressing statistical inference of life expectancy and other aggregated mortality measures was introduced by Chin Long Chiang (1961, 1984). Chiang considered an unbiased estimate of probability of death $\hat{q} = D/N$, where D denotes death number (number of events) within a certain age interval and N cohort size (number of trials) at the beginning of the same age interval. For estimation of the standard error of the probability of death, Chiang considered the scheme of Bernoulli trials. Then the random death numbers would be distributed binomially around the mean value estimated as $\hat{q} \cdot N$ and with the standard deviation estimated as $\sqrt{\hat{q}(1-\hat{q})} \cdot N$. In this case, standard error of probability of death is

$$S_{\hat{q}} = \sqrt{\frac{\hat{q} \cdot (1 - \hat{q})}{N}}.$$

If the observed death number and population size are higher than 15-20, the normal distribution can be considered as a good approximation of the binomial distribution. In this case, upper and lower confidence limits for the confidence level can be expressed according to the Wald formula

$$CI = \hat{q} \pm z_{\alpha/2} S_{\hat{q}},$$

where $z_{\alpha/2}$ is the $100 \cdot (1 - \alpha/2)$ percentile of the standardized normal distribution. Finally, Chiang deduced formulae linking standard error of survival probabilities and life expectancies with standard error of age-specific probabilities of death (Chiang, 1961 and 1984).

Following the same approach, Mathers derived a formula for estimation confidence limits for healthy life expectancy (Mathers, 1991, Jagger et al., 2006).

Silcocks, Jenner & Reza (2001) proposed an alternative method for the estimation of confidence intervals for life expectancy, especially for the case of an age interval of infinite length (open-age interval).

Unfortunately, the Wald formula is not applicable when the probability of death is close to zero or to one, or if the population size N is too low. Condition $\min(D, N - D) \geq 5$ must be fulfilled for Wald's formula to be applicable. It is also unclear how age-specific death numbers that are equal to zero should be handled.

Toson and Baker (2003) carried out extensive testing of the estimation of the life-table function for small populations. They showed that the common practice of treating this problem by imputing small positive numbers instead of zeros leads to an underestimation of the standard error. Their advice is to assume that standard errors are equal to zero whenever null death counts are observed.

This report provides an instrument that realizes the scheme of Bernoulli trials directly through the use of Monte Carlo simulations.

Mathematics

Input data for calculations of period life tables include the observed death counts ${}_t D_x$ and population exposed to the risk of death ${}_t E_x$ at ages $[x, x + \tau)$. These data are used to calculate age-specific death rates ${}_t M_x = {}_t D_x / {}_t E_x$, as well as further life-table functions, such as ${}_t q_x$, l_x , ${}_t d_x$, ${}_t L_x$, T_x and e_x . Following Chiang (1984), the age-specific numbers of trials resulting in death events is estimated as ${}_t \tilde{N}_x = {}_t D_x / {}_t q_x$. The number of deaths in a population with a death probability equal to ${}_t \hat{q}_x$ and a population size equal to ${}_t \tilde{N}_x$ is considered as a random variable Δ that is distributed binomially with the distribution function

$$\Pr(\Delta = {}_t D_x) = B(\Delta; {}_t \hat{q}_x \cdot {}_t \tilde{N}_x, {}_t \tilde{N}_x) = \frac{{}_t \tilde{N}_x!}{{}_t \tilde{N}_x \cdot ({}_t \tilde{N}_x - {}_t D_x)} \cdot ({}_t \hat{q}_x)^{{}_t D_x} \cdot (1 - {}_t \hat{q}_x)^{({}_t \tilde{N}_x - {}_t D_x)}.$$

The Excel function CRITBINOM(${}_t \tilde{N}_x, {}_t \hat{q}_x, \rho$) returns the smallest death number for which the cumulative binomial distribution $Bcum({}_t \hat{q}_x \cdot {}_t \tilde{N}_x, {}_t \tilde{N}_x)$ is less than ρ . If ρ is an evenly distributed random number, then CRITBINOM(${}_t \tilde{N}_x, {}_t \hat{q}_x, \rho$) is the random death-number variable with the distribution $B(\Delta; {}_t \hat{q}_x \cdot {}_t \tilde{N}_x, {}_t \tilde{N}_x)$. In this way, we are able to generate random numbers of deaths for every age group, and to calculate from the obtained age-specific death numbers the simulated life table.

In the last age interval of infinite length (e.g., 85+, 90+ or 100+), the situation is somewhat different due to the fact that all the survivors at the beginning of this age will die within the interval. This means that the probability of death must be exactly equal to 1. Once ${}_{\infty} q_{\omega} = 1$,

then ${}_{\infty}d_{\omega} = l_{\omega}$. This implies ${}_{\infty}L_{\omega} = l_{\omega}/{}_{\infty}M_{\omega}$, but also ${}_{\infty}L_{\omega} = l_{\omega}e_{\omega}$. Consequently, $e_{\omega} = 1/{}_{\infty}M_{\omega}$. All these expressions are exact equations. This further means that the open-ended age group includes a stochastic component that comes from younger ages, with the randomly fluctuating value l_{ω} , but this age group does not produce its own stochastic error.

In Sullivan's life table, we have to simulate two random variables: the random age-specific death numbers (as in case of the conventional life table) and the random age-specific numbers of people living in poor health. The latter are generated by applying the same Excel function from given age-specific numbers of survey respondents and from the estimated prevalence of ill health among them.

Let us assume now that we need to estimate 95% confidence limits for a chosen quantity Z , such as life expectancy at birth or at age 15, probability of survival from age 60 to age 80, healthy life expectancy at age 50, etc. After completing Monte Carlo simulations K times, K random realizations of Z are obtained ($\hat{Z}^1, \dots, \hat{Z}^K$). The upper and lower limits of the 95% confidence interval are the 2.5% and 97.5% percentiles of the set ($\hat{Z}^1, \dots, \hat{Z}^K$). For example, if $K=1,000$ and the random Z -values are ranked in the ascending order, then the upper and lower limits will have ranks 25 and 975, respectively.

How to use these spreadsheets

ConfidenceInterval-LE.xls

This spreadsheet estimates confidence limits for any life table quantity. In order to launch the calculation procedure, the following steps are necessary:

- 1) Specify sex ("Males" or "Females") in cell A11, lower limits of age groups in cells A12:A35, lengths of age intervals in cells B12:B35 (+ if open ended), population exposures in cells C12:C35 (null exposures are not acceptable) and death counts in cells D12:D35.
- 2) Specify significance level in cell K1.
- 3) Specify number of simulations in cell K2. It is recommended that this number is not lower than 500.
- 4) Insert reference to the quantity of interest in cell K3, and reference to its point estimate in cell H8.
- 5) Click the "Run" button.

ConfidenceInterval-HE.xls

This spreadsheet estimates confidence limits for any quantity of Sullivan's life table. In order to launch the calculation procedure, the following steps are necessary:

- 1) Specify sex ("Males" or "Females") in cell A12, lower limits of age groups in cells A13:A36, lengths of age intervals in cells B13:B36 (+ if open-ended), population exposures in cells C13:C36, death counts in cells D13:D36, numbers of respondents in cells M13:M36 and numbers of respondents who reported poor health (disability, chronic illness, etc.) in cells N13:N36. All these numbers must be positive.
- 2) Specify significance level in cell K1.

- 3) Specify number of simulations in cell K2. It is recommended that this number is not lower than 500.
- 4) Insert reference to the quantity of interest in cell K3 and reference to its point estimate in cell H8.
- 5) Click the "Run" button.

The optimal number of simulations can be determined through several experiments. Usually it is enough to have $K=1,000$. This means that the number of simulations K is high enough to secure robustness of the obtained confidence limits. (They confidence interval values would not change visibly after another run of the program with $K=1,000$).

References

Chiang, C. L. 1961. Standard error of the age-adjusted death rate. Vital Statistics, Special Reports 47:271-285, USDHEW

Chiang, C. L. 1984. The Life Table and its Applications, Malabar (FL), Robert E Krieger Publ Co. P. 153-168

Jagger, C., Cox, B., Le Roy S., EHEMU. 2006. Health Expectancy Calculation by the Sullivan Method. Third Edition. EHEMU Technical Report September.

Mathers C. 1991. Health expectancies in Australia in 1981 and 1988. Technical Report. Australian Institute of Health: AGPS, Canberra.

Silcocks, P. B. S., Jenner, D. A., Reza, R. 2001. Life expectancy as a summary of mortality in a population: statistical considerations and suitability for use by health authorities, Journal of Epidemiology and Community Health 55, 38-43.

Sullivan, D.F. 1971. A single index of mortality and morbidity: HSMHA Health Reports, 86: 347-354.

Toson, B., Baker, A. 2003. Life expectancy at birth: methodological options for small populations. Office for National Statistics (ONS) UK