SHORT REPORT

Improving estimation of the variance of expectation of life for small populations

P B S Silcocks

J Epidemiol Community Health 2004;58:611–612. doi: 10.1136/jech.2003.011338

It has been pointed out (D Eayres, personal communication) that Chiang's formula[1] for estimation of the standard error of the expectation of life (which allows for the fact that mortality rates increase with age *within* each age band—clearly relevant with abridged life tables and advancing age) can be improved by combining it with the fact that the final age band is *open-ended* so that its width is a random variable and therefore contributes to the standard error of expectation of life.[2] However, in simulations Eayres notes (personal communication) that there is still bias in the estimation of expectation of life, especially with small populations. The purpose of this communication is to suggest an improvement to reduce the effect of this bias. Errors in estimation of the population may also cause spurious variation in e^o_x between populations, but these are not considered here.

METHODS

In Silcocks *et al*[2] the expectation of life (e^o_x) is expressed in terms of functions of the age specific mortality rates. The (theoretically) open-ended final age band contributes to the point estimate of e^o_x and its variance by assigning to the final age band a notional fixed width, which in turn is a function of that age band's mortality rate.

Assuming exponential survival, the mean survival time in the final age band is $1/m_n$ years; the total person-time is l_{n-1}/m_n where l_{n-1} is the number alive at the start of the final age band. With exponential survival there is no fixed width to the final age band but the correct person-time is obtained *as if* given by the area of a triangle, of height l_{n-1} and base equivalent to a notional width of the final age band fixed at $2/m_n$ so that

$$z_n = 2/m_n \qquad (1)$$

$$\text{with variance}(z_n) = 4/r_n m^2_n = {z_n}^2/r_n \qquad (2)$$

where m_n = annual mortality rate in the final age band and r_n = number of deaths in one year in the final age band.

The variance of e^o_x is then estimated by applying the delta rule to the function relating e^o_x to the age specific mortality rates.

However, although the estimated mortality rate m_n is an unbiased estimate of the true mortality rate, z_n *overestimates* the width of the age band on average.

An approximation for the bias can be obtained mathematically as follows:

Putting $z_n = g(m_n)$ and denoting the expected value of m_n by μ a Taylor series expansion gives:

$$z_n = g(m_n) = g(\mu) + (m_n - \mu)dg/dm_n + \tfrac{1}{2}(m_n - \mu)^2 d^2g/dm^2_n + \ldots$$

Considering only the first three terms and taking expectations of both sides (noting that the expected value of $(m_n-\mu)$ is 0, because m_n is an unbiased estimate of the true mortality rate) then after rearranging, an approximate estimate of the bias[3] is the expected value of $\tfrac{1}{2}(m_n-\mu)^2 d^2g/dm^2_n = \tfrac{1}{2}\text{var}(m_n)d^2g/dm^2_n$

Using a binomial distribution for the number of deaths (regarding m_n as an annual probability of dying), the bias corrected estimate is z_n-bias, which works out to be:

$$z^*_n = 2/m_n[1 - (1 - m_n)/r_n] \qquad (3)$$

The fewer the number of deaths, r_n, the bigger the difference between z_n and z^*_n Conversely if r_n is large then $1/r_n$ is small and the bias is small.

The variance of z_n under the binomial assumption is given by:

$$\text{variance}(z_n) = 4(1 - m_n)/r_n m^2_n \qquad (4)$$

Alternatively the point estimate could instead be multiplied by a *shrinkage factor* to reduce the mean squared error,[3] which will also reduce bias because the expectation of the mean squared error = bias^2+variance.

$$\theta = {z_n}^2/[{Z_n}^2 + \text{var}(z_n)] \qquad (5)$$

$$\text{with } z^{**}_n = \theta z_n \qquad (6)$$

Table 1 Comparison of point, shrunk, and bias corrected estimates

	Estimated final age band width	Variance	Skewness of simulated values
True point estimate (based on ''known'' rate)	11.69	8.72*	–
Mean of uncorrected simulated values	12.57	14.61	2.66
Mean of simulated shrunk values	11.68	9.83	1.89
Mean of simulated BC values	11.60	9.06	1.55

*That is, the variance that would be estimated if 13 deaths were observed in a population of 76, using (4)

Key points

- The variance of expectation of life must be allowed for in studies of small populations
- With small populations, the small number of deaths in the oldest age band may bias the estimate of the final age band width, and hence the expectation of life
- A simple bias correction is proposed to alleviate this

As a practical example, consider an English electoral ward with a total population of perhaps 6000 people with about 76 people in the 85+ age band, among whom there would be 13 or so deaths per year. The point estimate of the age band width would be 11.69 years. I compared this estimate and its variance (8.72) with the mean and variance of the estimated widths based on 5000 simulations (using MINITAB v10 to simulate a binomial distribution with n = 76 and 13 expected deaths—that is, with binomial parameter 0.171).

RESULTS

In this exercise we know that the ''true'' value for z_n is given by the point estimate $z_n = 2/m_n$. The mean of the simulated values shows the extent to which an uncorrected estimate of the notional width is an overestimate (from table 1 on average about 7.5% greater).

The simulations also show that the variance of the uncorrected estimate of z_n was 68% greater than predicted from the formula (4).

The mean of the bias corrected simulated values was closer to the true value (99.9%) than was the mean of the shrunken simulated values (99.2%). The point estimate of variance was smaller than the variances of the shrunken and bias corrected estimates but was closer to the latter.

The bias correction also improved the approximation to a normal distribution by reducing the skewness of the uncorrected estimate and performed better than the shrunk estimate—although because the width of the age band is a reciprocal transformation of the (approximately normal) mortality rate, zero skewness is an unrealistic target.

Policy implications

- Apparently high expectation of life estimates in small populations may be the result of correctable bias, but errors in estimating the population at risk may also contribute.

DISCUSSION

The limited simulation exercise performed here suggests that a correction based on a Taylor series has the following advantages:

- It removes some of the bias associated with the original estimate for the width of the final age band
- The point estimate of variance is then closer to the variance of the simulated corrected estimates
- The skewness is substantially reduced, improving the overall fit of the estimated expectation of life to a normal distribution.

The temptation to adjust the estimated variance of the age band width by plugging-in the corrected point estimate should be resisted as this will then under-estimate the variance. Instead the original variance estimate should be used.

Funding: none.

Competing interests: none declared.

Correspondence to: Dr P B S Silcocks, Trent Institute for Health Services Research, Queen's Medical Centre, Nottingham NG7 2UH, UK; paul.silcocks@nottingham.ac.uk

Accepted for publication 17 October 2003

REFERENCES

1 **Chiang CL**. *The life table and its applications*. Malabar, FL: Krieger, 1984;chapter 8.
2 **Silcocks PBS**, Jenner DA, Reza R. Life expectancy as a summary of mortality in a population: statistical considerations and suitability for use by health authorities. *J Epidemiol Community Health* 2001;**55**:38–43.
3 **Gilchrist W**. *Statistical modelling*. Chichester: Wiley, 1984:305.