



Estimating life expectancy using an age-cohort model in Taiwan.

W C Lee and R L Hsieh

J Epidemiol Community Health 1996;50;214-217
doi:10.1136/jech.50.2.214

Updated information and services can be found at:
<http://jech.bmj.com/cgi/content/abstract/50/2/214>

These include:

Rapid responses

You can respond to this article at:
<http://jech.bmj.com/cgi/eletter-submit/50/2/214>

Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article

Notes

To order reprints of this article go to:
<http://journals.bmj.com/cgi/reprintform>

To subscribe to *Journal of Epidemiology and Community Health* go to:
<http://journals.bmj.com/subscriptions/>

Methods

Estimating life expectancy using an age-cohort model in Taiwan

W-C Lee, R-L Hsieh

Abstract

Objectives – Life expectation is a valuable summary index in public health and actuarial science. The life expectancies published in the vital statistics, however, are derived from the “current” rather than from the “cohort” life table. The former is based on a strong assumption of constant mortality in the population, whereas the latter calls for a recording of the mortality experience of a group of individuals, which is often an impossible task. Thus, a method of calculating cohort life expectancy without actual follow up is much needed.

Methods – Estimation of cohort life expectancy was based on an age-cohort model. Mortality data for the male population in Taiwan from 1951 to 1990 are used to illustrate the methodology.

Results – The increment of life expectancy over time in Taiwan is actually steeper than was previously thought using the current life table technique.

Conclusions – The method is easy to implement and the data required are the usual age and period cross classified mortality data. It warrants further investigation.

derived from the “current” rather than from the “cohort” life table. A current life table projects the life span of each individual in a hypothetical cohort on the basis of the cross sectional death rates in a given population. When we speak of the life expectancy of an infant born in the current year, for example, we mean the life expectancy that would be obtained if that infant were subject throughout his life to the same age specific mortalities prevailing in the current year – clearly, a bold assumption. Yet, since it is evidently an insurmountable if not an impossible task to record the actual mortality experience of a particular group of individuals (the cohort) from the first birth to the last death, the “current” life expectancy is still widely used.

We present a new method of estimating the cohort life expectancy. The estimation (projection) is based on an age-cohort model. Because the data required for the model are the usual age and period cross classified mortality data in the format of vital statistics, follow up of a group of subjects for a long time is unnecessary. The mortality data for the male population in Taiwan from 1951–90 are used to illustrate the methodology.

(*J Epidemiol Community Health* 1996;50:214–217)

Method

MODELLING MORTALITY DATA USING AN AGE-COHORT MODEL

Assume that the mortality data in a defined population are cross classified by age and period, with the interval widths of both variables being 5 years. The age groups are indexed by i ($i=1,2,\dots,I$), the period groups by j ($j=1,2,\dots,J$), thereby forming an I by J contingency table. We assume that there are 18 groups for the age variables ($I=18$). The youngest age group ($i=1$) represents those died at 0–4 years old and the oldest one ($i=18$) represents those died at 85 years old or older. Groups of birth cohorts are indexed by k ($k=1,2,\dots,I+J-1$). Since it is obvious that we can tell the birth year of a person from knowledge of the calendar year and the age when he or she died, the index of birth cohort in the aforementioned contingency table is related to the age and the period indices by “ $k=j-i+I$ ”.

The mortality rates for the cell with age i and period j are $\lambda_{ij}=m_{ij}/n_{ij}$, where m_{ij} is the expected death number and the n_{ij} represents

The problem of measuring health status in human populations has long been of considerable interest to epidemiologists. The mortality rate, either age specific or age adjusted, is one of the pertinent indices that has often been used for this purpose. In some instances, however, life expectancy is the preferred measure because it has a clear meaning: the life expectancy at a particular age is the mean lifetime of individuals who have lived beyond that age.¹ Because of its application as a population summary measure, the life expectancy at birth is often used to make comparisons between different countries, across different time periods, or between different subgroups in the population. Furthermore, a table with life expectancies at various ages is also used as a basis for the computation of things such as life insurance premiums and annuities.²

It must be remembered, however, that life expectancies published in the vital statistics are

Graduate Institute of Epidemiology, College of Public Health, National Taiwan University, 1 Jen-Ai Rd, 1st Sec, Taipei, Taiwan, ROC
W-C Lee

Department of Physical Medicine and Rehabilitation, Shin Kong Wu Ho-Su Memorial Hospital, Taipei, Taiwan
R-L Hsieh

Correspondence to:
Dr W-C Lee.

Accepted for publication
July 1995

the size of the population in each cell. Although mortality rates in the oldest age category truly represent a weighted average of rates in various birth cohorts, they are, for simplification, fully ascribed to the cohorts aged 85 to 89 years in the given calendar periods which contribute the vast majority of person-years and death in this age group. The observed death number in each cell is assumed arising from a Poisson distribution with the mean m_{ij} .³ And the death rates, λ_{ij} , are modelled as a log-linear function of age and cohort parameters:

$$\log(\lambda_{ij}) = \mu + \alpha_i + \gamma_k, \quad k = j - i + I,$$

where the μ is the intercept term, the age effects are represented by α_i , and the cohort effects by γ_k ($\sum_i \alpha_i = \sum_k \gamma_k = 0$). This is the usual age-cohort model. Readers are referred to the report of Clayton and Schifflers⁴ for more details. The parameters of the age-cohort model can easily be estimated using standard statistical packages such as GLIM⁵ or EGRET⁶.

ESTIMATING THE COHORT LIFE EXPECTANCY

The age-cohort model, with its parameters estimated, can be used to predict the age specific mortality rates for the entire life span of each birth cohort under study, that is, between the first birth-cohort ($k=1$) and the last one ($k=I+J-1$). For each cohort, Chiang's technique¹ can be used to convert the age specific death rates to the probability of dying for each 5 year age group. The cohort life expectancies can then be calculated using the standard life table method.¹

It should be noted that projection of future mortality and retro-projection of past mortality using the age-cohort model, as in this paper, do not require extrapolation of either the age or the cohort parameters.⁷ This is in sharp contrast to the mortality projection scenario using the more sophisticated age-period-cohort model.⁸ The latter calls for an extrapolation of recent period trends which is always hazardous. Also note that our model based life expectancy estimation works well even when the data set contains some missing cells, as shown by the following example.

Results

EXAMPLE

Mortality data for the male population of Taiwan from 1951–90 (table) were analysed to illustrate the methodology. In some earlier calendar periods in Taiwan, the oldest age category used to tabulate mortality and population data was not as old (≥ 85 years) as that used in most recent periods. This explains why some cells in the table are empty. We begin by fitting an age-period and an age-cohort model to this data set.⁴ Since there are 21 birth cohort groups in total, we assume a cubic polynomial for the cohort parameters for the sake of parsimony. We decided to use a cubic degree polynomial because of the observation that it fits "closely" enough to the cohort effects obtained from full parameterisation. In view of the problem of

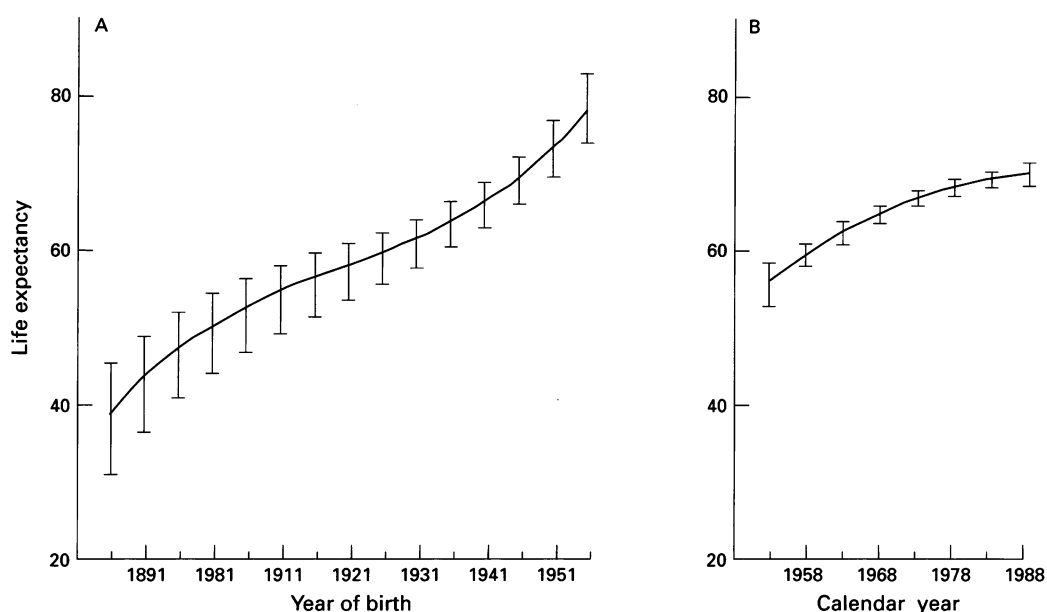
extra-Poisson variation which often jeopardises a formal statistical test in mortality analysis,⁹ we do not intend to build up the model from the lowest degree polynomial to the highest degree possible in a stepwise fashion and rely upon a likelihood ratio test, for example, as the stopping criteria. For comparison, we also assume a cubic polynomial for the period parameters in the age-period model. Therefore, the age-cohort and the age-period models have the same number of model parameters ($18+3=21$) and hence the same degree of freedom (total number of cells – number of model parameters = $132-21=111$).

When we had estimated the relevant parameters in the age-cohort and the age-period models, we projected or retro-projected the mortality rates cross sectionally or longitudinally to complete the life span of the study population. We then applied the standard life table technique¹ to estimate the "cohort" and the "period" life expectancies (both are life expectancies at birth). The results are plotted in the figure. For birth cohorts more recent than 1956, the estimated cohort life expectancies are becoming larger than 80 years. In this study, the oldest age group used for life table calculation is an open ended "85+", which means that the mortality rate for ages over 85 years is assumed to be constant. We therefore refute the results of life expectancy estimates in some recent birth cohorts (note that we did not "discard" the data of recent cohorts) as we believe that a reliable life expectancy estimation, in these circumstances calls for better understanding of elderly mortality, either in terms of more refined mortality grouping in the data set for the elderly or through extrapolation of the age parameters into old ages in the age-cohort model. It is also noted that the more recent the cohort, the stronger the assumption that the future mortality rates conform to the age-cohort model with the same set of parameters.

The figure also presents the 95% confidence intervals of the life expectancy estimates, which are based upon the bootstrap method.¹⁰ The confidence intervals could, alternatively, be derived from the usual variance-covariance matrix of the model parameters. The interest here is the variability of the life expectancy estimates however, not the parameters per se in the age-cohort or the age-period models. Complex matrix manipulation (multiplication and inversion for example) is thus required for the derivation. This is why we prefer the use of the bootstrap method – though computer intensive it is simple in concept and elegant. Briefly, one thousand bootstrap replications are resampled from the original data set and each is analysed using the aforementioned procedures (including the fitting of the age-cohort and the age-period models, the projection and retro-projection of the mortalities, and the life table analysis). The pointwise percentiles of the estimates of these 1000 replicates are then taken as the interval estimates of the life expectancies. How bootstrap samples are generated deserves some attention, since the goodness of fit statistic of either the age-cohort model (likelihood ratio statistic = 32196.0, $df=111$) or the age-

Death number and population number (in parenthesis x 1000) for male population of Taiwan, 1951–90.

Age (y)	1951-55	1956-60	1961-65	1966-70	1971-75	1976-80	1981-85	1986-90
0-4	81776 (3977)	71963 (3856)	53029 (5049)	35001 (4982)	22494 (4713)	18176 (4862)	13340 (5066)	8331 (4357)
5-9	6613 (2611)	5889 (3929)	4896 (4730)	4284 (5098)	3459 (5048)	2863 (4796)	2591 (4935)	2398 (5124)
10-14	3761 (2477)	2612 (2597)	3074 (3924)	3215 (4735)	3143 (5091)	2658 (5037)	2377 (4780)	1318 (4906)
15-19	3722 (2230)	3539 (2468)	3482 (2587)	4946 (3912)	5652 (4711)	7137 (5069)	6204 (5009)	7206 (4739)
20-24	4953 (1879)	4674 (1679)	4978 (2500)	4713 (2578)	6701 (3897)	8122 (4700)	7925 (5049)	7881 (4992)
25-29	5511 (1649)	4779 (1735)	4953 (2216)	5219 (2431)	5040 (2534)	6877 (3850)	7989 (4643)	8661 (4978)
30-34	5840 (1435)	5560 (1749)	5727 (2088)	5627 (2219)	5748 (2397)	5313 (2495)	7635 (3804)	9123 (4568)
35-39	6569 (1225)	6522 (1521)	7039 (1926)	7594 (2246)	7193 (2225)	7102 (2360)	6958 (2467)	10082 (3727)
40-44	8008 (1067)	7855 (1294)	8305 (1613)	9892 (1718)	11568 (2438)	9493 (2181)	9641 (2320)	9608 (2431)
45-49	9464 (853)	9532 (1132)	10287 (1308)	12083 (1712)	14068 (2317)	14636 (2372)	13007 (2126)	13159 (2275)
50-54	10281 (646)	11537 (833)	12862 (1083)	14694 (1350)	16873 (1768)	19807 (2236)	19575 (2289)	17181 (2067)
55-59	11014 (458)	13042 (593)	14836 (756)	17214 (1056)	19880 (1293)	22824 (1656)	26403 (2116)	25897 (2179)
60-64	11867 (318)	13145 (390)	16019 (504)	19683 (689)	21585 (950)	25532 (1174)	30356 (1522)	34797 (1956)
65-69	12226 (226)	13114 (247)	14772 (310)	19327 (429)	23222 (580)	26243 (814)	32988 (1026)	38739 (1345)
70-74	-	-	13291 (178)	16537 (234)	20744 (326)	26165 (455)	33438 (659)	39671 (846)
75-79	-	-	10390 (94)	12198 (111)	15262 (152)	20057 (222)	25606 (322)	35495 (475)
80-84	-	-	-	-	8958 (59)	12144 (85)	15781 (132)	22469 (194)
85+	-	-	-	-	6120 (22)	7693 (31)	9790 (48)	14586 (76)



Model based life expectancy estimates and 95% confidence intervals for the male population of Taiwan ((A) cohort life expectancy; (B) period life expectancy).

period model (likelihood ratio statistic = 75686.8, $df=111$) for this example indicates the presence of extra-Poisson variation (over dispersion).⁹ Although this lack of fit is not an uncommon phenomenon, especially when dealing with rates in large populations, and may not necessarily imply a systematic departure from the models,¹¹ the confidence intervals of the life expectancy estimates should reflect this properly. In generating the bootstrap samples, therefore, we included a random effect term to accommodate this extra-variation, in addition to the usual assumption that the death

number in each cell follows the Poisson distribution. Readers are referred to the paper of Breslow⁹ for technical details.

The figure shows that the increment of life expectancy over time for the male population in Taiwan is actually steeper (figure (A)) than was previously thought using current life table technique (figure (B)).

Discussion

Although the model based life expectancy estimation illustrated here has some desirable

features, several points need to be clarified before it is accepted. First of all, we do not take into account the differences in the "fraction of last age interval of life"¹ in each age category. Rather, we assume that deaths occurred uniformly throughout each year of age. This assumption is especially unrealistic for the first age category, since a large proportion of deaths below 5 years old occurs in the first year, and a large proportion of infant deaths occurs in the first few weeks. To overcome the problem, a complete (one year age group) life table is needed instead of an abridged one. The period grouping should, consequently, be based on each calendar year instead of five, so that an age-cohort model can be applied. Although, in such a circumstance, the large number of parameters might inevitably compel us to use a smoothing technique¹² on the parameters, the basic principle of life expectancy estimation remains the same.

Secondly, it should be pointed out that our method relies greatly on mortality forecast (and backcast), and that the basis of forecasting is the assumption that the prevailing trends will continue to operate in a predictable way (that is, according to the presumed age-cohort or the age-period model). Such an assumption surely cannot be justified beforehand. Uncertainty about future trends is unfortunately difficult, if not impossible, to incorporate into the confidence intervals of the life expectancies. It should therefore be realised that the confidence intervals based on the forecasting model we report are over optimistic and are better interpreted as stability indicators of the life expectancy estimates rather than in a literal sense. A so called "process model" was recently proposed for forecasting chronic disease risks. This takes into account the information of time trajectories of mortality risk factors in the population under study¹³ and seems a step towards predicting future mortality trends. That particular model, however, calls for an integration of multiple sources of information beyond the age and period cross classified mor-

tality data and the calculation is also much more involved.

Finally, our method can be extended and applied to such scenarios as estimation of the potential gains in (cohort) life expectancy when mortality from some specific causes is assumed reduced or eliminated. To do this, information on cause specific mortality rates in the population under study is needed and a multiple decrement life table analysis¹⁴ has to be performed. In addition, to get a better fit for the "all cause" data which we analysed in this paper, disaggregation of the mortality rates into their component causes, modelling them separately using the age-cohort or the age-period model and then recombining the results to a "better" estimation of the life expectancy can be tried. Further studies to examine the magnitude of bias or efficiency loss if the "crude" estimation illustrated in the present paper is carried out instead are warranted.

The authors wish to thank the anonymous referees for helpful comments.

- 1 Chiang CL. *The life table and its applications*. Florida: Robert E Krieger Publishing Company, 1984.
- 2 Bowers NL, Gerber HU, Hickman JC, Jones DA, Nesbitt CJ. *Actuarial mathematics*. Schaumburg: Society of Actuaries, 1986.
- 3 Frome EL. The analysis of rates using Poisson regression models. *Biometrics* 1983;39:665-74.
- 4 Clayton D, Schifflers E. Models for temporal variation in cancer rates. I: age-period and age-cohort models. *Stat Med* 1987;6:449-67.
- 5 Baker RJ, Nelder JA. *Generalized linear interactive modeling (GLIM)*. Release 3. Oxford: Numerical Algorithms Group, 1978.
- 6 Statistics and Epidemiology Research Corporation. EGRET user's manual. Copyright 1985-1988. Seattle: Statistics and Epidemiology Research Corporation.
- 7 Brenner H, Ziegler H. Monitoring and projecting cancer incidence in Saarland, Germany, based on age-cohort analyses. *J Epidemiol Community Health* 1992;46:15-20.
- 8 Osmond C. Using age, period and cohort models to estimate future mortality rates. *Int J Epidemiol* 1985;14:124-9.
- 9 Breslow NE. Extra-Poisson variation in log-linear models. *Applied Statistics* 1984;33:38-44.
- 10 Efron B. Bootstrap methods: another look at the jackknife. *Annals of Statistics* 1979;7:1-26.
- 11 Holford TR. The estimation of age, period and cohort effects for vital rates. *Biometrics* 1983;39:311-24.
- 12 Becker NG, Marschner IC. A method for estimating the age-specific relative risk of HIV infection from AIDS incidence data. *Biometrika* 1993;80:165-78.
- 13 Dowd JE, Manton KG. Forecasting chronic disease risks in developing countries. *Int J Epidemiol* 1990;19:1019-36.
- 14 Tsai SP, Lee ES, Hardy RJ. The effect of a reduction in leading causes of death: potential gains in life expectancy. *Am J Public Health* 1978;68:966-71.