

AUTOREGRESSIVE AGE–PERIOD–COHORT MODELS

WEN C. LEE AND RUEY S. LIN

Graduate Institute of Epidemiology, College of Public Health, National Taiwan University, No. 1, Jen-Ai Road, 1st Sec., Taipei Taiwan, Republic of China

SUMMARY

Age–period–cohort analysis of vital data has received much attention recently, and it is already well known that the exact linear relation of the three time factors creates a non-identifiability problem. Previous studies have shown that the curvature terms of these factors are estimable but the linear trends are not. However, little attention has been paid to the possibility that the effects due to cohort and/or period might change through time stochastically rather than deterministically and hence display a stochastic trend. In this paper, we model the cohort effects as an AR(1) process and use lung cancer mortality data from 1966 to 1990 for males in Taiwan as an example. The parameters are identifiable in the proposed model and the estimates are found to be stable. However, the assumption made in the model should be carefully considered before using our methodology.

1. INTRODUCTION

Birth cohort analysis for disease incidence and mortality has a long history in epidemiology.¹ Originally, the technique employed a graphical approach in which the contributions of various temporal factors were determined visually. Recently, the statistical age–period–cohort (APC) analysis has received much attention in the literature.^{2,3} By introducing a three-factor multiplicative model, separate effects due to age, period and cohort for incidence or mortality rates of a particular disease can be quantified and the significance of each can be gauged.^{2,3} The results from the model provide epidemiologists with clues or hypotheses for testing the etiology of disease. The model is of the form:

$$\log \lambda_{ij} = \mu + \alpha_i + \beta_j + \gamma_k, \quad \begin{array}{l} i = 1, 2, \dots, I \\ j = 1, 2, \dots, J \\ k = j - i + I \end{array} \quad (1)$$

where λ_{ij} are the unknown true disease rates cross-classified by age (i) and period (j), μ is the intercept term, and the α_i , β_j , and γ_k represent effects due to age, period and cohort factors, which are indexed by i , j and k , respectively. In this form the model resembles the analysis of variance, and the usual constraints apply, namely $\sum_i \alpha_i = \sum_j \beta_j = \sum_k \gamma_k = 0$. The equation $k = j - i + I$ merely states that knowing the age at death (or onset), i , and the date of death (or onset), j , of a particular disease for a subject determines the date of birth, k .

Although the APC model (1) has been applied to studies on secular trends of various diseases,^{4–10} it is well known that the exact linear relationship of the three factors ($k = j - i + I$) creates a non-identifiability problem.^{2,3} Because of this, it is necessary to impose one or more

constraints or assumptions on the parameters. For example, Osmond and Gardner⁵ introduced the constraint of minimal-weighted distances between the APC model and the three two-factor models (AP, AC and PC). On the other hand, Robertson and Boyle⁷ assumed a constant effect within each age, period and cohort cross-classified cell. However, the constraints used have been criticized for lacking a sound biological basis, leading to parameters where interpretation is in doubt.^{3,8} Recently, Holford *et al.*¹⁰ incorporated biological information about the functional form of age effects derived from multistage carcinogenesis models for cancers into the APC analysis. They showed that it is possible to determine an identifiable set of parameters in this way. However, the parameter estimates were highly unstable, and four of the five different functional forms which they tried for age resulted in biologically implausible estimates.

For clearer understanding of the source of non-identifiability in the APC model, we can partition each of the three effects into a component that describes linear trend (slope), and a component for curvature (departure from the overall linear trend).⁶ It has been shown that the curvature terms are all estimable but the linear trends of age, period and cohort effects are hopelessly entangled with each other.⁶ This led Clayton and Schifflers⁸ to introduce the concept of 'drift', which is a common linear trend for the period and cohort parameters. This drift is attributable neither to period nor to cohort effects and can be thought of as an overall measure of change in relative risk over time.

Little attention has been paid by previous investigators to the fact that, in the context of time series analysis, a distinction should be made between two types of trend for temporal variables, namely deterministic and stochastic.¹¹ The ideas described above, of partitioning effects into linear trends and curvature components and introducing the drift parameter, implicitly assume that the linear trends for the three temporal variables, if they exist, are deterministic. However, this may not always be so, since it is possible that the effects due to cohort and/or period change through time stochastically rather than deterministically. We may therefore model them as stationary time series. Further support for doing so lies in the fact that there is usually some dependency in the cohort and/or period effects. For example, with birth-cohort, it is conceivable that the effect of one particular birth-cohort may be interrelated with the effects of its predecessors, since the lifestyles, dietary and cultural habits of previous birth-cohorts may be preserved and passed on to the succeeding ones. It is also possible that subjects of different but close birth-cohorts may experience similar environmental exposures and hence share similar cohort effects.

In this paper we model cohort effects as an autoregressive process of order one, that is AR(1). An identifiable set of parameters can be obtained from this 'autoregressive' APC model. Lung cancer mortality data from 1966 to 1990 for men in Taiwan are analysed as an example. The relation between our approach and the recently proposed 'Bayesian APC models'¹²⁻¹⁴ will also be discussed.

2. THE AUTOREGRESSIVE AGE-PERIOD-COHORT MODEL

As before, the unknown true disease rate is modelled as a log-linear function of age, period and cohort parameters. However, in our formulation the cohort effects are considered to be derived from an AR(1) process. That is,

$$\log \lambda_{ij} = \mu + \alpha_i + \beta_j + \gamma_k \quad (2)$$

$$\gamma_k = \phi \gamma_{k-1} + \delta_k,$$

where the δ_k are independently and identically normally distributed with mean zero and variance, σ^2 . In the model described above, μ is the intercept term, and the α_i and β_j represent the age and period effects as before. The usual constraints are applied, namely $\sum_i \alpha_i = \sum_j \beta_j = 0$. Two new parameters are introduced to model the autoregressive nature of the cohort effects, namely ϕ and σ^2 . The parameter ϕ describes the sign and strength of dependency of the autoregressive process and the condition $|\phi| < 1$ is imposed for stationarity; ϕ can be interpreted as the (auto)correlation of the cohort effects between two adjacent birth-cohorts. When $\phi > 0$ there is positive correlation, when ϕ is zero the effects of two birth-cohorts are independent, and when $\phi < 0$ there is negative correlation. In practice, the last situation is rarely encountered, since it is hard to imagine that the lifestyles, dietary and cultural habits of one birth-cohort would tend to be different and at variance with those of the previous one. Beside the dependency of one birth-cohort on its predecessor, subjects in each individual birth-cohort are assumed exposed to a random shock, δ_k , which can be thought of as environmental in origin. These random shocks are modelled as a white noise process with parameter σ^2 .

3. ESTIMATION OF THE PARAMETERS

We assume that disease occurrence or death, y_{ij} , in each cell, cross-classified by age and period, follows a Poisson distribution with mean $Z_{ij} = n_{ij}\lambda_{ij}$, where n_{ij} is the person-years in each cell and λ_{ij} is the event rate specified in (2) above. Then the likelihood function of observing y_{ij} is given by

$$L(\mu, \alpha, \beta, \phi, \sigma^2 | \mathbf{y}) = \int_{\gamma} L_1(\mu, \alpha, \beta, \gamma | \mathbf{y}) L_2(\phi, \sigma^2 | \gamma) d\gamma \tag{3}$$

where $L_1(\mu, \alpha, \beta, \gamma | \mathbf{y})$ is the Poisson likelihood,¹⁵ namely

$$L_1(\mu, \alpha, \beta, \gamma | \mathbf{y}) = \prod_{i=1}^I \prod_{j=1}^J Z_{ij}^{y_{ij}} e^{-Z_{ij}/\gamma_{ij}!},$$

and $L_2(\phi, \sigma^2 | \gamma)$ is the likelihood function of AR(1),¹¹ namely

$$L_2(\phi, \sigma^2 | \gamma) = \left[\frac{1 - \phi^2}{2\pi\sigma^2} \right]^{\frac{1}{2}} \exp \left[\frac{-\gamma_1^2(1 - \phi^2)}{2\sigma^2} \right] \left[\frac{1}{2\pi\sigma^2} \right]^{(I+J-2)/2} \exp \left[-\frac{1}{2\sigma^2} \sum_{k=2}^{I+J-1} (\gamma_k - \phi\gamma_{k-1})^2 \right],$$

and the integration is over γ . The likelihood function (3) above recognizes the observable y_{ij} as data but considers the cohort components, γ_k , as random unobservable effects. We refer to this as the ‘marginal’ or ‘unconditional’ likelihood. A closed analytical form may not exist. Also, in the context of APC analysis, cohort effects are considered equally, if not more, important than the age and period effects, and thus treating them as unobservable may not be desirable. In the following, therefore, we turn to the ‘conditional’ maximum likelihood approach. That is, instead of maximizing the integrated likelihood (3), we search in the parameter space for γ in addition to that of $(\mu, \alpha, \beta, \phi, \sigma^2)$, for the point at which $L_1 L_2$ is maximized. A similar approach has been adopted in mixture models.¹⁶ The ‘conditional’ log-likelihood function of the proposed model (2) is as follows.

$$\begin{aligned} \log L(\mu, \alpha, \beta, \gamma, \phi, \sigma^2 | \mathbf{y}) &= -\frac{I+J-1}{2} \log 2\pi + \frac{1}{2} \log(1 - \phi^2) - \frac{I+J-1}{2} \log \sigma^2 \\ &\quad - \frac{S}{2\sigma^2} + \sum_{i=1}^I \sum_{j=1}^J (y_{ij} \log Z_{ij} - Z_{ij} - \log y_{ij}!), \end{aligned}$$

where

$$S = \gamma_1^2(1 - \phi^2) + \sum_{k=2}^{I+J-1} (\gamma_k - \phi\gamma_{k-1})^2.$$

The first and second partial derivatives of this likelihood function with respect to the parameters can be easily derived (available from the authors upon request). We then adopt the Levenberg-Marquardt algorithm¹⁷ for maximization, which ensures that we always move in an ascending direction on the likelihood surface (a computer program written in SAS/IML language¹⁸ is available on request). We may use, for example, the method of Osmond and Gardner⁵ to obtain initial values for the parameter estimates. In the following example, however, we have found that, irrespective of the starting values, the algorithm converges to the same parameter values. We also found that despite the high correlation of the cohort effects and the high dimension of the search space, numerical difficulties did not occur and the fitting of the model required no more than 40 iterations and was completed within 10 minutes on an IBM PC-486.

Since the likelihood function contains non-linear parameters and some of them are range-restricted, using the usual asymptotic properties of maximum likelihood estimates for determining precision may not be appropriate. In the following example, we resort to the bootstrap method¹⁹ instead, by means of which stability of the model can also be checked. Alternatively, we can calculate the profile likelihood-based confidence intervals, which do not depend heavily on whether the likelihood function can be approximated by a quadratic function. However, it should be noted that the theory behind the profile-likelihood method is at best an asymptotic one and furthermore it does not have the range-preserving property.²⁰ For our implementation of the bootstrap method, 1000 bootstrap replications are re-sampled from the original dataset, assuming that the number of deaths in each cell follows the Poisson distribution. Each is analysed using the aforementioned procedure, and the pointwise percentiles of the parameter estimates of these 1000 replicates are taken as our interval estimates. Note that, in the following example, since goodness-of-fit using either the chi-square test ($p = 0.27$) or the likelihood ratio test ($p = 0.27$) indicates a good fit to the data, there seems no need to consider the problem of extra-Poisson variation (overdispersion).²¹

4. LUNG CANCER MORTALITY

To illustrate the calculations, we consider data on lung cancer mortality from 1966 to 1990 for men in Taiwan. We restrict our analysis to ages between 30 and 79 years, since there are too few cases in younger ages and the quality of data is less reliable in older ones. Table I, which was compiled from vital statistics in Taiwan, presents the number of lung cancer deaths and the population of males in these ages.

The autoregressive APC model (2) was applied to this dataset using the estimation procedure just described; parameter estimates and the bootstrap 95 per cent confidence intervals are presented in Table II. The cohort effects ($\mu + \gamma_k$) are also plotted in Figure 1. It can be seen that the effects initially increase and then decrease in recent birth-cohorts. The auto-correlation of the cohort effects is rather high ($\hat{\phi} = 0.90$) in this example, which conforms with our expectations. If desired, some of the asymmetry in the confidence intervals tabulated and plotted can be adjusted for by the bias-correction method discussed by Efron and Tibshirani.²⁰ However, it can be seen from the narrow confidence intervals that our proposed model is rather stable, especially in comparison with those of Holford *et al.*¹⁰

Table I. Number of lung cancer deaths and population for men in Taiwan, by age and period

Age (years)	Period				
	1966-1970	1971-1975	1976-1980	1981-1985	1986-1990
<i>Lung cancer deaths</i>					
30-34	36	47	32	74	68
35-39	57	61	74	81	116
40-44	125	151	135	167	195
45-49	215	297	350	329	358
50-54	289	450	635	786	662
55-59	393	595	887	1329	1416
60-64	394	733	1027	1653	2182
65-69	308	559	1082	1702	2333
70-74	192	360	662	1379	2104
75-79	95	160	355	665	1320
<i>Population ($\times 1000$)</i>					
30-34	2218.59	2396.95	2494.82	3803.88	4568.25
35-39	2246.12	2225.08	2360.36	2467.23	3726.53
40-44	2129.07	2438.46	2181.02	2320.28	2430.55
45-49	1711.67	2317.11	2372.19	2126.07	2275.05
50-54	1349.95	1667.89	2235.97	2289.36	2066.92
55-59	1056.38	1293.32	1656.05	2115.72	2178.59
60-64	689.28	949.68	1173.96	1521.83	1955.82
65-69	428.74	579.64	814.09	1025.52	1344.75
70-74	234.23	325.98	454.82	659.03	846.04
75-79	110.71	152.27	222.13	321.51	474.56

It should be emphasized that we have taken this dataset only to demonstrate our methodology. We make no claim that using such a model is the best or even an appropriate way to analyse lung cancer mortality in Taiwan (we will discuss this issue later). However, we discovered that the increasing/decreasing cohort pattern of lung cancer mortality in this example is consistent with the patterns of other cancers previously studied in Taiwan,^{22,23} although whether such cohort effects can be linked to specific events occurring in the male populations residing in this island is an issue which deserves additional scrutiny.

5. COMPARISON WITH THE 'BAYESIAN AGE-PERIOD-COHORT MODELS'

It is of interest to compare our approach with the 'Bayesian APC models',¹²⁻¹⁴ which have just appeared in the literature. These papers propose a Bayesian model in which the effects along each of the three temporal scales are subject to a stochastic second difference prior. Such a prior acts much in the same way as our AR(1) process. However, there are differences between the two approaches which need discussion. First, in the Bayesian approach, the autoregressive error structure is introduced into the models mainly for the purposes of 'smoothing' and 'prediction', whereas we are more interested in studying the mechanism that drives the cohort effect. The autocorrelation parameter (ϕ) in our model serves such a purpose, from which inferences can be drawn regarding the strength of dependency between adjacent cohorts and hence the macroscopic cultural transmission processes that may occur in the population. In the Bayesian APC

Table II. Parameter estimates and 95 per cent bootstrap confidence intervals derived from model (2)

Parameters	Estimates	95% bootstrap confidence intervals
α_i (age effects)		
32	- 2.60	- 2.74- - 2.48
37	- 2.10	- 2.20- - 2.00
42	- 1.36	- 1.44- - 1.28
47	- 0.67	- 0.74- - 0.61
52	- 0.03	- 0.08- 0.02
57	0.54	0.50- 0.58
62	1.06	1.02- 1.10
67	1.46	1.42- 1.51
72	1.77	1.72- 1.83
77	1.93	1.86- 2.01
β_j (period effects)		
1968	- 0.30	- 0.35- - 0.25
1973	- 0.15	- 0.18- - 0.11
1978	- 0.01	- 0.04- 0.01
1983	0.19	0.16- 0.21
1988	0.27	0.23- 0.30
$\mu + \gamma_k$ (cohort effects)		
1891	- 8.62	- 8.77- - 8.53
1896	- 8.57	- 8.69- - 8.48
1901	- 8.40	- 8.49- - 8.32
1906	- 8.27	- 8.34- - 8.21
1911	- 8.10	- 8.15- - 8.05
1916	- 8.06	- 8.10- - 8.02
1921	- 8.07	- 8.11- - 8.03
1926	- 8.12	- 8.17- - 8.07
1931	- 8.15	- 8.21- - 8.09
1936	- 8.28	- 8.35- - 8.20
1941	- 8.33	- 8.42- - 8.24
1946	- 8.41	- 8.50- - 8.30
1951	- 8.51	- 8.60- - 8.42
1956	- 8.62	- 8.77- - 8.52
ϕ	0.90	0.86- 0.90
σ^2 ($\times 100$)	0.92	0.66- 1.78

models, there is, by contrast, only the σ^2 parameter serving to measure the 'roughness' of the three temporal effects.

Secondly, in the Bayesian approach of Berzuini and Clayton,^{12,13} there is no explicit prior specification for 'level' and 'trend' of the effects of the three temporal variables (only the second difference prior is specified). The models are therefore unidentifiable and an arbitrary linear constraint still has to be imposed on the log-linear trend components of age, period and cohort effects. Such an arbitrary constraint indeed has no effect whatsoever on the predictions of the model.²⁴ Yet in instances when knowing the separate trends of age, period, as well as cohort variables is important for understanding the etiology of the disease under study, the credibility of such a constraint should be carefully checked. In our approach, the assumption of a stationary

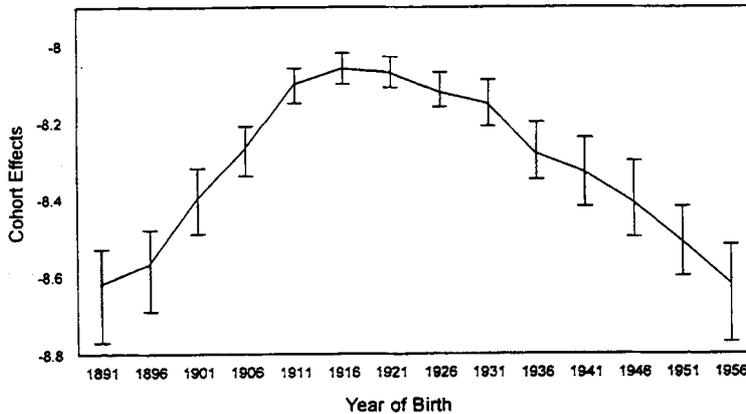


Figure 1. Cohort effects ($\mu + \gamma_k$) and 95 per cent bootstrap confidence intervals derived from model (2)

autoregressive process for the cohort effects can be viewed as the 'constraint' we impose. Its credibility will be examined and discussed in the following.

6. DISCUSSION

In this paper we model cohort effects as an AR(1) process, and we have demonstrated that by doing so identifiability of the parameters and stability of the model are both ensured. However, several points should be clarified before accepting this methodology. First, we choose cohort effects in our experiment for time-series modelling, although the same opportunity was available for period effects (or even for age effects). Although it is rather straightforward to extend our methodology to model simultaneously the period and cohort effects as stationary time series, we have some reservations about doing so. This is because in APC modelling the number of period groups available is often too small for meaningful time-series modelling, whereas this is usually not the case for the relatively larger number of cohort groups. In addition, the period trend probably more often behaves in a deterministic fashion than in a stochastic way, since the effects due to period often result, for example, from improvement of diagnostic technology or from better survivorship due to better treatment, which certainly involve some deterministic components.

Secondly, we may question application of only the AR(1) process, the simplest time-series model,¹¹ for the cohort effects. Indeed, models with autoregression terms of higher order and/or with moving average parameters can be attempted. However, in view of the additional parameters and further complexity which these factors would introduce, we feel that this approach would only improve marginally the fit of the model and is unlikely to shed further light on the problem of identifying model parameters.

Thirdly, a technical difficulty which often jeopardizes APC analysis is that the data may not be tabulated on a square grid (5-year for age and 1-year for calendar period, for example), since in this case the cohort variable cannot be easily defined. The problem can be avoided by using a more refined grid, say by 1-year group on both axes. In such circumstances, however, the large number of parameters may inevitably compel resort to some sort of smoothing technique²⁵ or even to a Bayesian approach,¹²⁻¹⁴ since crude maximum likelihood estimation, in this case, calls for inversion of a large matrix, and the estimates, even if obtainable, will be very unstable.

Finally, and most importantly, our proposed model, though identifiable and stable, is definitely not assumption-free. The assumption of some dependency in the cohort effects, we believe, requires relatively little scrutiny, but the assumption that the cohort trend is stochastic rather than deterministic is probably violated more often and thus should be carefully considered. One example is the lung cancer mortality in Taiwan just described. This latter assumption can be violated if the cohort effects of this disease result largely from the industrialization or Westernization process, which, in a developing country like Taiwan, is likely to be deterministic in nature. Accordingly, only information about the curvature of the cohort effects (the quadratic pattern in Figure 1) can be given a biological interpretation. Although data with a deterministic trend can be easily 'detrended' before time-series analysis,¹¹ such a technique can, unfortunately, not be applied to the APC analysis due to its inherent non-identifiability in the (deterministic) trend components.⁶ Hence epidemiologists should have, for the particular disease under study, at least some knowledge regarding the possible determinants of the cohort effects and their possible modes of action before considering use of our proposed model. This may result in our methodology being less widely used. Nevertheless, since the assumptions underlying the proposed model are explicitly made clear, suitable places for its application should be, we believe, not too hard to find.

ACKNOWLEDGEMENTS

We are grateful to the two reviewers for their valuable advice and for bringing to our attention some important recent references.

REFERENCES

1. Frost, W. H. 'The age selection of mortality from tuberculosis in successive decades', *American Journal of Hygiene(Section A)*, **30**, 91–96 (1939).
2. Kupper, L. L., Janis, J. M., Karmous, A. and Greenberg, B. G. 'Statistical age-period-cohort analysis: a review and critique', *Journal of Chronic Diseases*, **38**, 811–830 (1985).
3. Holford, T. R. 'Understanding the effects of age, period, and cohort on incidence and mortality rates', *Annual Review of Public Health*, **12**, 425–457 (1991).
4. Fienberg, S. E. and Mason, W. M. 'Identification and estimation of age-cohort models in the analysis of discrete archival data', in Schuessler, K. F. (ed), *Sociological Methodology*, Josey-Bass, San Francisco, 1979, pp. 1–67.
5. Osmond, C. and Gardner, M. J. 'Age, period and cohort models applied to cancer mortality rates', *Statistics in Medicine*, **1**, 245–259 (1983).
6. Holford, T. R. 'The estimation of age, period, cohort effects for vital rates', *Biometrics*, **39**, 311–324 (1983).
7. Robertson, C. and Boyle, P. 'Age, period, and cohort models: the use of individual records', *Statistics in Medicine*, **5**, 527–538 (1986).
8. Clayton, D. and Schifflers, E. 'Models for temporal variation in cancer rates II: age-period-cohort models', *Statistics in Medicine*, **6**, 469–481 (1987).
9. Levi, F., La Vecchia, C., Decarli, A. and Randriamiharisoa, A. 'Effects of age, birth cohort and period of death on Swiss cancer mortality, 1951–1984', *International Journal of Cancer*, **40**, 439–449 (1987).
10. Holford, T. R., Zhang, Z. and McKay, L. A. 'Estimating age, period and cohort effects using the multistage model for cancer', *Statistics in Medicine*, **13**, 23–41 (1994).
11. Wei, W. W. S. *Time Series Analysis, Univariate and Multivariate Methods*, Addison-Wesley Publishing Company, Redwood City, California, 1990.
12. Berzuini, C., Clayton, D. and Bernardinelli, L. 'Bayesian inference on the Lexis diagram', *Bulletin of the International Statistical Institute*, **LV**, 149–165 (1993).
13. Berzuini, C. and Clayton, D. 'Bayesian analysis of survival on multiple time scales', *Statistics in Medicine*, **13**, 823–838 (1994).
14. Breslow, N. E. and Clayton, D. G. 'Approximate inference in generalized linear mixed models', *Journal of the American Statistical Association*, **88**, 9–25 (1993).

15. Frome, E. L. 'The analysis of rates using Poisson regression', *Biometrics*, **39**, 665–674 (1983).
16. Symon, M. J., Grimson, R. C. and Yuan, Y. C. 'Clustering of rare events', *Biometrics*, **39**, 193–205 (1983).
17. Thisted, R. A. *Elements of Statistical Computing*, Chapman and Hall, London, 1988.
18. SAS Institute Inc. *SAS/IML User's Guide*, Release 6.03 Edition, Cary, NC, U.S.A., 1988.
19. Efron, B. 'Bootstrap methods: another look at the jackknife', *Annals of Statistics*, **7**, 1–26 (1979).
20. Efron, B. and Tibshirani, R. J. *An Introduction to the Bootstrap*, Chapman and Hall, London, 1993.
21. Breslow, N. E. 'Extra-Poisson variation in log-linear models', *Applied Statistics*, **33**, 38–44 (1984).
22. Lee, W. C. and Lin, R. S. 'Age-period-cohort analysis of pancreatic cancer mortality in Taiwan, 1971–1986', *International Journal of Epidemiology*, **19**, 839–847 (1990).
23. Lee, W. C. and Lin, R. S. 'Interactions between birth cohort and urbanization on gastric cancer mortality in Taiwan', *International Journal of Epidemiology*, **23**, 252–260 (1994).
24. Osmond, C. 'Using age, period and cohort models to estimate future mortality rates', *International Journal of Epidemiology*, **14**, 124–129 (1985).
25. Becker, N. G. and Marschner, I. C. 'A method for estimating the age-specific relative risk of HIV infection from AIDS incidence data', *Biometrika*, **80**, 165–178 (1993).