

## **Modeling Exposure Intensity and Susceptibility-Latency Effect from Retrospective Cohort Data – Application to Lung Cancer Mortality in Blackfoot Disease Endemic Area in Taiwan**

WEN-CHUNG LEE and RUEY-SHIUNG LIN

Graduate Institute of Epidemiology  
College of Public Health  
National Taiwan University  
R.O.C.

### *Summary*

It is challenge in epidemiology to characterize the temperol aspect of exposure-disease association. The authors propose a stochastic model to deal with exposures that are time-dependent and exhibit susceptibility and latency effects. The model is applied to a retrospective cohort data on lung cancer mortality in the blackfoot disease endemic area in Taiwan. The authors compare the proposed model with the multistage model, the back-calculation model, the catalytic model, and the age-period-cohort models.

**Key words:** Age-period-cohort model; Back-calculation model; Catalytic model; EM algorithm; Lung cancer, Multistage model; Smoothing.

### **1. Introduction**

Characterizing the relationship between an exposure and a disease is a fundamental issue in epidemiology. To this end, epidemiologists often use indices such as relative risk or odds ratio to represent strengths of association or dose-response gradient. However, such representations do not pay due respect to the temporal aspect of exposure-disease associations. For example, the intensity of an exposure may vary with time and its effects on humans may depend on the ages at which a subject is exposed. Furthermore, the effect of the exposure may not be simultaneous. Rather, it may be latent for years after the first exposure, becomes full-fledged, and then fades away. It becomes a challenge then to give a full account on these complex relationships.

In this paper, the authors propose a stochastic model to deal with exposures that are time-dependent and exhibit susceptibility and latency effects. A retrospective cohort data on lung cancer mortality in the blackfoot disease endemic area in Taiwan was used to illustrate the method.

## 2. The proposed models

### 2.1 Notations and assumptions

We begin by assuming that people in different birth cohorts have different early exposures (experiences) and hence have different disease rates in their later life-times. For cancers of environmental origins which are of present concern, the disease rates may not be simply related to a single carcinogenesis insult which occurred sometime in early life, but rather to the accumulation of them from birth to current age of a subject. Furthermore, in the theory of multistage carcinogenesis (CRUMP and HOWE, 1984), the effect due to the accumulation of such time-varying exposures is to produce an excess risk (rate) in one's later life. Therefore, we may consider the following excess risk model, which was formulated in discrete time using a five-year time scale, i.e.,

$$\lambda_{ij} = \lambda_{ij}^* + \sum_{a=1}^i \beta_{k+a-1} \cdot \alpha_{a,i-a+1}, \quad \begin{aligned} i &= 1, 2, \dots, I; \\ j &= 1, 2, \dots, J; \\ k &= j - i + I. \end{aligned} \quad (1)$$

In this model,  $\lambda_{ij}$  are the unknown true disease rates of the population under study. The first term in the right of the equation,  $\lambda_{ij}^*$ , represents the background disease rates which were assumed known from national vital statistics. These rates were cross-classified by age,  $i$  ( $i = 1, 2, \dots, I$ ) and period,  $j$  ( $j = 1, 2, \dots, J$ ). The second term represents the excess risk due to accumulation of carcinogenesis insults since birth (the summation starts from age group '0-4' ( $a = 1$ )). The exposure intensities of the study population in different time period ( $t$ ) are quantified by the parameters,  $\beta_t$ , with  $t = 1, 2, \dots, I + J - 1$ . These parameters share the same measurements unit as  $\lambda_{ij}^*$ , i.e., occurrence (death) per 100,000 person-year, for example. In view of the possibilities that the same amount of exposure may have different impact on disease rates, depending on the ages of the subjects being exposed and the elapsed time since that exposure, another set of parameters,  $\alpha_{g,d}$  was introduced, with  $g \geq 1$ ,  $d \geq 1$ , and  $g + d - 1 \leq I$ . These parameters are unitless and reflect the relative impact of an exposure experienced at age ( $g$ ) upon the disease rate ( $d - 1$ ) unit of time later. The dependency of disease rates upon the exposed ages is termed the susceptibility effects in this paper, and the dependency of time-since-exposure, the latency effects. Instead of setting a particular  $\alpha_{g,d}$  to one as a reference group for the relative impact parameters, we use the constraint  $5 \cdot \sum_g \sum_d \pi_g \alpha_{g,d} = 1$ , where  $\pi_g$  is the proportion of individuals in the study population who are of age  $g$ . This proportion was assumed known from census tables or from surveys. Using this constraint, the time-dependent exposure intensity,  $\beta_t$ , can be interpreted as average lifetime (up to age  $I$ ) excess disease rates for a subject exposed at time  $t$ . The reason why we construct our model using 5-year time scale rather than using continuous scale or other discrete time lengths is that

the national vital statistics is often presented by 5-year age groups. Using discrete formulation also considerably simplifies the problem.

## 2.2 Incorporating information of out-migration subjects

It is noted that model 1 may not be identifiable when all  $\beta_i$  are equal or when  $J < I$  (see appendix 1). We are often interested in disease rates up to 80 years of age ( $I = 16$ ). And for the model to be identifiable, it requires that we collect mortality data for a time span of 80 years or more ( $J \geq 16$ ) which is hardly possible in practice. However, if the out-migration subjects from the study population can be traced and the corresponding disease rates be collected, the above conditions can be relaxed and the non-identifiability problem may be avoided (appendix 1). In this case, the disease rates are cross-classified by age,  $i$ , period,  $j$ , and time-since-out-migration,  $m$ . The index  $m$  is also expressed in a five-year discrete time scale, with  $m = 1$  standing for those who didn't migrate out or who had out-migrated in less than 4 years. For analyzing such data, the following model can be formulated,

$$\lambda_{ijm} = \lambda_{ij}^* + \sum_{a=1}^{i-m+1} \beta_{k+a-1} \cdot \alpha_{a,i-a+1}, \quad \begin{aligned} i &= 1, 2, \dots, I; \\ j &= 1, 2, \dots, J; \\ k &= j - i + I; \\ m &= 1, 2, \dots, i. \end{aligned} \quad (2)$$

Model 2 above differs from model 1 in only one respect, namely, it is assumed that subjects are no longer exposed to the 'specific' carcinogenesis insults once they migrated out from the study population. This assumption appears reasonable when the population under study are exposed to a 'specific' carcinogen, for example, an occupational cohort in a certain factory or in an underground mine, etc., or the residents living in an environmentally polluted area, such as the blockfoot disease endemic area in Taiwan. The last situation will be described in more details in section 3.

## 2.3 Imposing a smoothness condition

The likelihood function of model 2 (see appendix 2) has a non-diagonal matrix of second partial derivatives, which makes direct maximization difficult. Furthermore, raw maximum likelihood estimates of  $\beta$  and  $\alpha$  tend to be very unstable due to the large number of parameters ( $(I + J - 1) + I(I + 1)/2 - 1$  in total). The same difficulties were previously encountered by the back-calculation models of AIDS (BECKER and MARSCHNER, 1993), which are akin to our model. Here we also use an EMS-like algorithm to overcome the difficulties. The EMS-algorithm (SILVERMAN et al., 1990) is the EM-algorithm (DEMPSTER et al., 1977) with a smoothing step added. The rationale behind the addition of a smoothing step lies in the fact

that there is strong *a priori* expectation that the  $\beta$  and the  $\alpha$  each follow a fairly smooth curve or surface. Note that the EMS-algorithm amounts to solving a penalized likelihood maximization problem (SILVERMAN et al., 1990). The estimation algorithm for the present problem is outlined in appendix 2.

### 3. Lung cancer mortality in blackfoot disease endemic area

Data on lung cancer mortality in the blackfoot disease endemic area in Taiwan was used to illustrate the methodology. Residents of the area have consumed water from artesian wells contaminated with inorganic arsenic since the turn of this century when new drilling techniques began to be employed. They gradually abandoned the use of the artesian wells decades later when a piped water system began to be introduced into the area in the 1950's. Previous studies conducted in this area have demonstrated significant dose-response relations between ingested arsenic and mortality from various cancers (CHEN et al., 1985, 1986; WU et al., 1989). However, the possibility that cancer potentials may be modified by temporal factors, such as birth-cohort, time-since-exposure, and age-at-exposure, etc., has not been investigated. It is therefore of interest to analyze mortality data in this area using model 2. Here, the data is from a retrospectively assembled cohort of 59156 subjects residing in this area in the year 1946. This cohort was followed-up till 1990 by us. The person-year observation and lung cancer mortality cases of this cohort were cross-classified according to age, period, and time-since-out-migration. Due to retrospective nature of the data, we do not have information of smoking behavior of the study subjects, one of the most important risk factors for lung cancer. However, this does not jeopardize the applicability of model 2. Since in model 2, we are examining the 'excess' effects from the 'area-specific' exposure (i.e., the ingested arsenic) relative to the general population. And there seems no reason to suspect that the out-migrated subjects also 'change' their smoking behaviors and contribute (or reduce) additional excess effect for lung cancer (note that in model 2, we do not assume out-migrated subjects 'stop' smoking after they left the study population). On the other hand, it seems reasonable to assume that subjects are no longer exposed to excess carcinogenesis insults (relative to the general population) once they migrated out from the study population, since the exposure of high-level inorganic arsenic from artesian well is very specific to the blackfoot disease endemic area in Taiwan.

Since national mortality for lung cancer in Taiwan was considered not reliable for analysis before 1966, follow-up data before that time was not used either. Thus we have  $I = 16$  5-year age groups (0-4, 5-9, ..., 75-79) and  $J = 5$  5-year period groups (1966-'70, '71-'75, ..., '86-'90). In addition, due to the aging of the cohort, we do not have information of those cells with  $i - j < 4$ . Note specifically that the youngest age group we have is  $i = 5$ . The incomplete nature of our data implies that those parameters,  $\alpha_{g,d}$ , for which  $g + d < 6$ , cannot be estimated, i.e.,

those parameters which describe the exposure accumulating processes for age groups  $i < 5$ . Thus we are left with a total of 145  $((16 + 5 - 1) + 16 \cdot (16 + 1)/2 - 1 - 4 \cdot (4 + 1)/2)$  parameters. The analysis is based on 340 cells with non-zero person-year which, in total, contain 808105.2 person-years of follow-up and 791 lung cancer deaths. Appendix 2 details the estimation algorithm. Convergence was declared when relative changes of parameter values between successive cycles for all the parameters are less than 0.0001. Using more stringent criteria requires more computer time which may take several hours.

Different bandwidth values were tried to smooth the parameters (see appendix 2 for a detailed description about smoothing). The smallest values that leads to biologically plausible estimates were subjectively chosen. Alternatively, one can use cross-validation technique to determine appropriate values of the smoothing parameters. However, for illustrative purposes, we feel our 'intuitive' approach is sufficient. Figure 1 displays estimates of time-dependent exposure intensity when different bandwidth values were used. It can be noted that bandwidth values  $h_\beta = 1.0$  and  $h_\alpha = 2.0$  lead to a smaller peak at around 1973–1983, which was considered implausible since

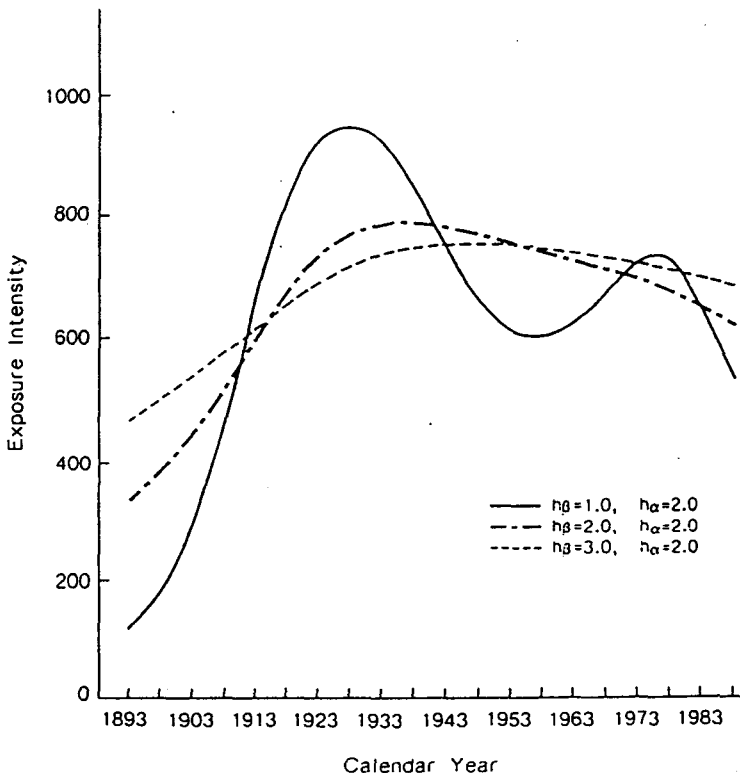


Fig. 1. Time-dependent exposure intensity in the blackfoot disease endemic area. The intensity can be interpreted as average lifetime (up to 80 years) excess lung cancer mortality for an exposed subject (per 100,000 person-year)

there is no evidence that the exposure intensity in the study area rises again in recent years. On the other hand, bandwidth values  $h_\beta = 3.0$  and  $h_\alpha = 2.0$  lead to estimates that are too much smoothed out. Therefore, we choose  $h_\beta = 2.0$  and  $h_\alpha = 2.0$  as reasonable values. From figure 1, it can be seen that the exposure patterns estimated from our model correspond roughly well with what we known about the chronicle of artesian wells usage in that area. The curve, however, started to decline a little bit earlier than was expected. This finding may not necessarily be a counter-example against our model, since other factors not associated with usage of artesian wells, such as nutritional improvement, dietary change, etc., in the study population (CHEN et al., 1988) may also contribute to the decline.

The large number of parameters and the addition of the smoothing steps preclude the use of the usual asymptotic properties of maximum likelihood estimates

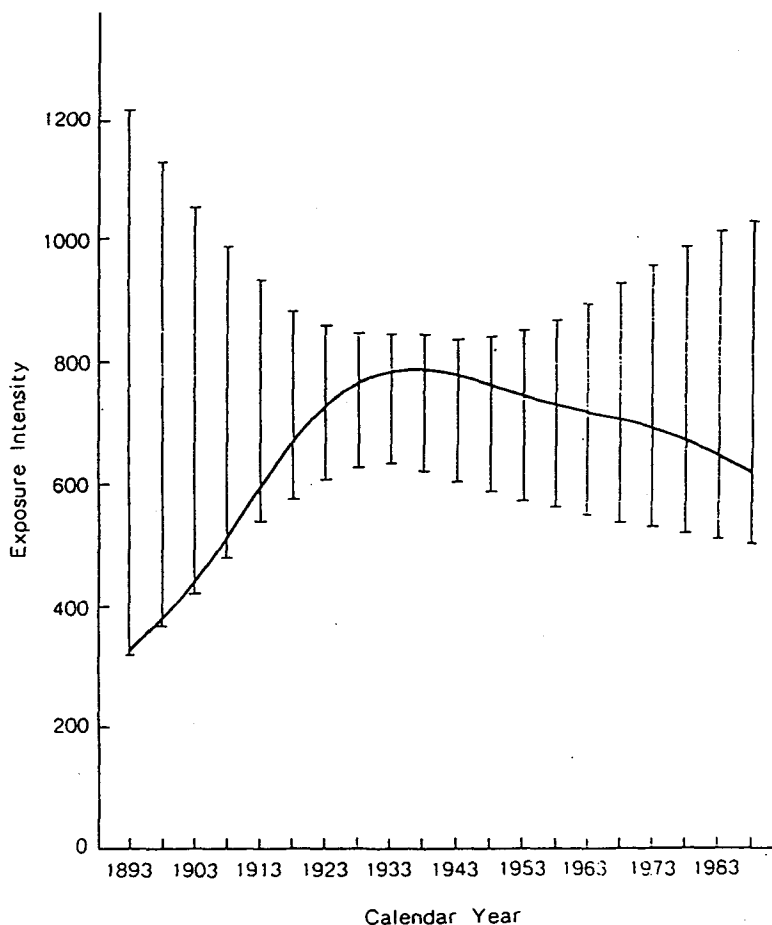


Fig. 2. Time-dependent exposure intensity in the blackfoot disease endemic area, with 90% bootstrap confidence intervals

for determining precision. Here, we resort to the bootstrap method (EFRON, 1979) instead. A total of 500 bootstrap samples were generated assuming that deaths (or events) follow a Poisson distribution with rates specified by model 2. These simulated data were each analyzed using our estimation procedure. Pointwise percentiles of the bootstrap replications were reported. The 90 percent bootstrap confidence intervals for the exposure intensity estimates were displayed in figure 2. It can be seen that there is much variability in estimates in the earlier and in the recent years. There is also a considerable right skew (exposure intensity toward higher level) in the bootstrap estimates in the early years. This may reflect the bias incurred from the smoothing procedures, as one can see from figure 1 that greater smoothing moves the exposure intensity toward higher levels in the early years. At this stage, it seems premature to draw any firm conclusion from the time patterns. Nevertheless this example serves well to illustrate our methodology.

As regards the relative impact effects (susceptibility-latency effects), a contour plot for joint actions of age-at-exposure and time-since-exposure were presented (figure 3). It can be seen that an interaction exists between the two variables. When time-since-exposure is less than 50 years, the relative impact increases as

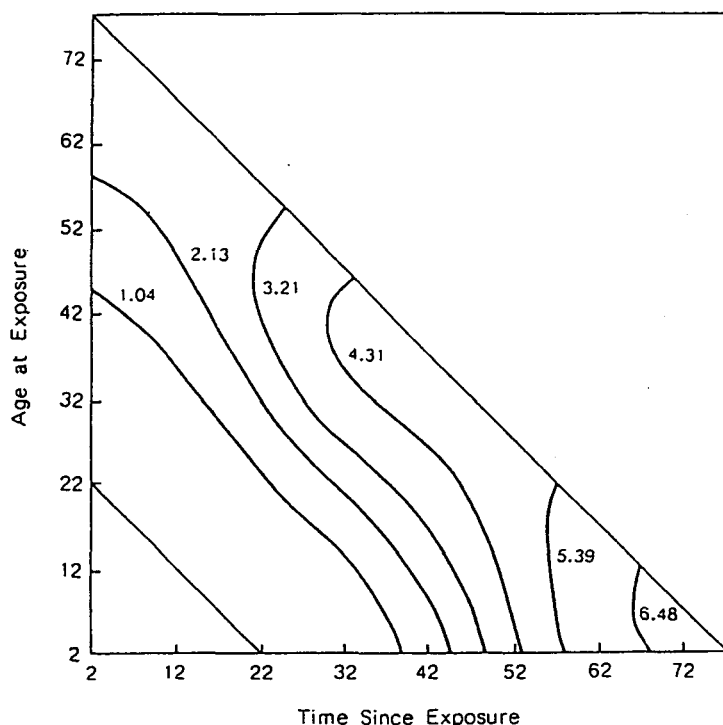


Fig. 3. Contour plot for joint actions of age-at-exposure and time-since-exposure variables. The value accompanying each contour line represents relative impact on lung cancer mortality for a given exposure. The relative impact estimates were subject to the constraint described in the text. Here the estimates were multiplied by 150 before display

either age-at-exposure or time-since-exposure increases. While for time-since-exposure greater than 50 years, the impact effects seem only related to the time-since-exposure variable. It indicates that arsenic may induce lung cancer through two different mechanisms, which is in accordance with present knowledge regarding arsenic carcinogenesis (MAZUMDAR et al., 1989). It should be noted again that the interpretation here is constrained due to low precision of the estimates as well.

#### 4. Discussion

In formulating our models, we assumed that cancers developed as a result of the accumulation of carcinogenesis insults since birth and that the accumulation processes are modified by susceptibility and latency effects. These assumptions seem reasonable for environmental cancers. In addition, since the proposed roles of exposure accumulation are essentially the same as what can be derived from the Armitage-Doll multistage models with a time-dependent exposure pattern (CRUMP and HOWE, 1984), our models also have a biological justification on a cellular basis. We, however, have formulated the relative impact of susceptibility and latency in our models less restrictively than those adopted by the multistage models. The multistage models with a time-dependent exposure pattern can at best accommodate susceptibility and latency effects which are polynomial functions and which combined multiplicatively in joint actions (CRUMP and HOWE, 1984). Therefore our models may also be applied to those cancers, such as breast cancer and childhood cancers, which are known not described well by the multistage models (MOOLGAVKAR, 1986). In the example presented, the interactive roles of age-at-exposure and time-since-exposure variables would not be disclosed had we separated the relative impact into two independent effects of susceptibility and latency. In addition, since the two variables of age-at-exposure and time-since-exposure usually are highly correlated, using our parameterization also avoids the collinearity problem. Other differences between our methodology and the multistage modeling are that, in the latter, the time-dependent exposure intensity was assumed known and that the information conveyed in the strata, cross-classified by age-at-exposure and time-since-exposure variables, was modeled to infer the particular stage(s) in the carcinogenesis processes which is (are) affected by the exposure. In many occupational and/or environmental settings, monitoring of exposure concentrations may not be routinely performed or the records may not be traced back long enough for a multistage modeling. Even if the monitored data is adequate, one may sometimes confront with a situation where no particular kind of exposures that are monitored has been proven causally linked to the disease under study. In such occasions, analysis with our model seems a reasonable alternative. The reason that we do not tailor our model to infer stage(s) affected as in the multistage model is that we believe such simple stochastic stage-transition model may not adequately reflect the biologic reality.



Besides the multistage models, our models are also similar in form with the catalytic models of infectious diseases (ADES and NOKES, 1993) and with the back-calculation models of AIDS (JEWELL and BACCHETTI, 1991; BECKER and MARSCHNER, 1993). It is of interest to investigate further the common properties of such 'accumulation-type' models and to find other possible applications for them. However, in the latter two models, it is the intensity of infection that is accumulating rather than is the intensity of exposure as in the present setting. Also, in the backcalculation approach, only one of the functions which are involved in the convolution is unknown, while in the present model, the two can be simultaneously estimated from the data. What makes this possible hinges on the fact that the information of out-migration subjects can be retrieved in this study. While a proper utilization of this valuable information calls for such a model which at first glance seems over-parameterized, the appendix 1 shows that it is identifiable at least under certain conditions.

It is also of interest to compare our models with the age-period-cohort (APC) models. The APC analysis is a popular tool for analyzing vital data (KUPPER et al., 1985; HOLFORD, 1991). In APC models, the disease rates are hypothesized as being determined by three independent effects of age, period, and cohort. However, in our proposed models, their roles are replaced by the  $\beta$  and  $\alpha$  parameters. The time periods covered by  $\beta$  parameters span from the earliest birth-year of the study population to the most recent period. Thus the  $\beta$  parameters can in a sense be viewed as a combination of the period and cohort effects. However, making epidemiological inferences based on the  $\beta$  parameters is advantageous because these parameters have already quantified the exposure intensities in the study population and therefore we do not have to obtain such information indirectly from the traditionally defined cohort or period variables. The effects of susceptibility and latency are separately accounted for by the  $\alpha$  parameters in our models. These effects are ignored in the APC analysis, the 'excess-risk APC models' (LEE and LIN, 1995) being one exception.

Due to large number of parameters in our models, an additional assumption of smoothness in the two sets of parameters is made. For the relative impact effects ( $\alpha$ ), this assumption seems reasonable since the relative impact is related to biological responses which in most circumstances will not display sudden jumps. There are occasions, however, when time-dependent exposure intensity ( $\beta$ ) is expected to show abrupt changes. The atomic bombing in Hiroshima and Nakasaki provides such an example. In those situations, the assumption of smoothness is no longer valid and our estimation algorithm may not be employed.

In order to improve the precision in the example we presented, one probably would consider increase sample size in the example. This is hardly possible since the example is already based on a very large cohort. One can nevertheless, consider perform a 'pooled' analysis of arsenic-related cancers, i.e., cancers of lung, bladder, kidney, liver, etc. (CHEN et al., 1985, 1986, 1988; WU et al., 1989), assuming a common pattern of time-dependent exposure intensity. Another possibil-

ity would be to resort to a parametric approach. The exponentially damped function as used in some catalytic models (ADES and NOKES, 1993) can be assumed for the exposure intensity, and the polynomial functions as derived from multi-stage models (CRUMP and HOWE, 1984) can be assumed for the age-at-exposure and/or time-since-exposure variables, etc. One may also try a response surface modeling for the  $\alpha$  parameters. With less parameters in the models, the parametric approach can give more precise estimates. Finally, our models can be easily extended to accommodate mortality data of those subjects who migrated into the study area at different ages. This additional information, if can be collected and analyzed in our example, may further improve the precision of the estimates as well.

### Acknowledgements

The authors wish to thank Dr. Kung-Yee Liang for helpful comments.

### Appendix 1

When all the  $\beta_t$  ( $t = 1, 2, \dots, I + J - 1$ ) are equal, namely  $\beta$ . Model 1 becomes

$$\lambda_{ij} = \lambda_{ij}^* + \beta \sum_{a=1}^i \alpha_{a,i-a+1} = \lambda_{ij}^* + \delta_i.$$

That is, the excess rates are only dependent on age  $i$ . In that case, we only have  $I$  degree of freedom but have  $I(I+1)/2$  parameters. Therefore the model is unidentifiable.

When  $\beta_t$  ( $t = 1, 2, \dots, I + J - 1$ ) are not all equal, the excess rates are also dependent on period  $j$ , in addition to be age-dependent. This is conceivable since subjects in different periods but with the same age will accumulate different amounts of exposure. Therefore, we have  $I \cdot J$  degrees of freedom. Now we have in total,  $(I + J - 1) + I(I + 1)/2 - 1$  parameters. And for the model to be identifiable, the number of parameters must not exceed degree of freedom. A little algebra shows that this leads to the condition,  $J \geq I/2 + 2$ . However, this condition is not sufficient. To see why, consider only the oldest age groups ( $i = I$ ). There are in total,  $J$  groups of them. And we use  $I$  parameters ( $\alpha_{a,I-a+1}$ ,  $a = 1, 2, \dots, I$ ) to describe the exposure accumulating processes for them. Therefore, we see that we must have  $J \geq I$  as well. Alternatively, using the method of GOODMAN (1974), models with  $J < I$  can also be shown to be locally unidentifiable.

When we have information of out-migration subjects and apply model 2 for analysis, the degrees of freedom increase but the number of parameters remains the same. Therefore, the above-listed conditions may be relaxed. The extent of relaxation depends on how much out-migration information we have.

# Appendix 2

Assume that disease occurrence or death in each cell, cross-classified by age, period, and time-since-out-migration, follows a Poisson distribution with the mean  $\mu_{ijm} = n_{ijm} \cdot \lambda_{ijm}$ , where  $n_{ijm}$  is the person-year of follow-up in each cell and  $\lambda_{ijm}$  is the disease rate specified in model 2. The likelihood function conditioned on observation of events in each cell,  $y_{ijm}$ , with  $i = 1, 2, \dots, I; j = 1, 2, \dots, J; m = 1, 2, \dots, i$  is

$$L(\beta, \alpha | \{y_{ijm}\}) = \prod_{i=1}^I \prod_{j=1}^J \prod_{m=1}^i \mu_{ijm}^{y_{ijm}} \cdot e^{-\mu_{ijm}}.$$

To apply the EM algorithm, we must first identify a more complete data set for which the maximum likelihood estimation is relatively easier. Here the complete data are  $z_{ijma}$  ( $1 \leq a \leq i - m + 1$ ), the excess case number of occurrences (deaths) due to the exposure which is experienced at age  $a$ , for each  $i, j$ , and  $m$  cross-classified cell. The log-likelihood function for the complete data is

$$\begin{aligned} \log L(\beta, \alpha | \{z_{ijma}\}) \\ = \sum_{i=1}^I \sum_{j=1}^J \sum_{m=1}^i \sum_{a=1}^{i-m+1} [z_{ijma} \cdot \log(n_{ij} \cdot \beta_{k+a-1} \cdot \alpha_{a,i-a+1}) - n_{ij} \cdot \beta_{k+a-1} \cdot \alpha_{a,i-a+1}], \\ k = j - i + I. \end{aligned}$$

The E-step consists of replacing  $z_{ijma}$  in the above equation by  $\hat{z}_{ijma}$ , the conditional expectation of  $z_{ijma}$ , given  $y_{ijm}$  and parameter value  $(\beta, \alpha)$ . That is

$$\hat{z}_{ijma} = E(z_{ijma} | \{y_{ijm}\}; \beta, \alpha) = \left[ \frac{\beta_{k+a-1} \cdot \alpha_{a,i-a+1}}{\lambda_{ij}^* + \sum_{g=1}^{i-m+1} \beta_{k+g-1} \cdot \alpha_{g,i-g+1}} \right] \cdot y_{ijm},$$

$$k = j - i + I.$$

The M-step consists of maximizing  $\log L(\beta, \alpha | \{\hat{z}_{ijma}\})$  with respect to  $\beta$  and  $\alpha$ , subject to  $\sum_g \sum_d \pi_g \alpha_{g,d} = 1$ . Since there are no close-form solutions in this step, some modification of the EM-algorithm should be undertaken. Here we adopt the 'multi-cycle ECM' algorithm (MENG and RUBIN, 1993). In our case, CM (conditional maximization) entails, instead of simultaneously maximizing the two sets of the parameters, maximizing one set while holding the other at its previous values. Multi-cycle here means we incorporate an E-step preceding each CM-step and let the resulting ECM-step cycle between the two sets of the parameters until some convergence criteria are met. Specifically, the CM-step for  $\beta$  and  $\alpha$  are, respectively,

$$\beta_t^{\text{CM}} = \frac{\sum_{k+a-1=t} \sum \sum \sum \hat{z}_{ijma}}{\sum_{k+a-1=t} \sum \sum \sum n_{ij} \cdot \alpha_{a,i-a+1}^{\text{old}}}, \quad k = j - i + I$$

and

$$\alpha_{g,d}^{\text{CM}} = \frac{\alpha_{g,d}^*}{5 \cdot \sum_u \sum_v \pi_u \alpha_{u,v}^*},$$

where

$$\alpha_{u,v}^* = \frac{\sum_{a=u, i=u+v-1} \sum \sum \sum \hat{z}_{ijma}}{\sum_{a=u, i=u+v-1} \sum \sum \sum n_{ij} \cdot \beta_{k+a-1}^{\text{old}}}, \quad k = j - i + I.$$

Now, a smoothing step for each set of the parameters can be added following each corresponding ECM-step. Smoothing here is accomplished through weighted averages of the unsmoothed parameters:

$$\beta_t^s = \sum_{u=1}^{I+J-1} W_{tu} \cdot \beta_u^{\text{CM}},$$

$$\alpha_{g,d}^s = \sum_u \sum_v W_{gduv} \cdot \alpha_{u,v}^{\text{CM}}.$$

Here  $W_{tu} \propto K_\beta \left( \frac{t-u}{h_\beta} \right)$  and  $W_{gduv} \propto K_\alpha \left( \frac{g-u}{h_\alpha} \right) \cdot K_\alpha \left( \frac{d-v}{h_\alpha} \right)$ , where  $k_\beta(\cdot)$  and  $k_\alpha(\cdot)$  represent kernel functions. These weights are suitably standardized such that  $\sum_u W_{tu} = 1$  and  $\sum_u \sum_v W_{gduv} = 1$ . In the example we analyzed,  $K_\beta(\cdot)$  was chosen to be the Gaussian kernel and  $K_\alpha(\cdot)$ , triweight kernel. The whole procedure can thus be described as the 'multi-cycle ECMS' algorithm. Theoretical investigation of the convergence properties of this algorithm is currently underway. At present, we have empirical evidences that, irrespectively of the starting configuration  $(\beta, \alpha)$ , the algorithm converges to the same parameter values. We also found that by using larger bandwidths,  $h_\beta$  and  $h_\alpha$ , the number of iterations required for convergence is reduced.

Implementation of the above algorithm is straightforward because simple explicit expressions exist for the updated parameter values. The procedure also has the advantage of always giving non-negative estimates.

## References

- ADES, A. E. and NOKES, D. J., 1993: Modeling age- and time-specific incidence from seroprevalence: toxoplasmosis. *Am. J. Epidemiol.* **137**, 1022-1034.
- BACCHETTI, P. and JEWELL, N. P., 1991: Nonparametric estimation of the incubation period of AIDS bases on a prevalent cohort with unknown infection times. *Biometrics* **7**, 947-960.
- BECKER, N. G. and MARSCHNER, I. C., 1993: A method for estimating the age-specific relative risk of HIV infection from AIDS incidence data. *Biometrika* **80**, 165-178.

- CHEN, C. J., CHUANG, Y. C., LIN, T. M., and WU, H. Y., 1985: Malignant neoplasms among residents of a blackfoot disease-endemic area in Taiwan: high-arsenic artesian well water and cancers. *Cancer Res.* **45**, 5895–5899.
- CHEN, C. J., CHUANG, Y. C., YOU, S. L., LIN, T. M., and WU, H. Y., 1986: A retrospective study on malignant neoplasms of bladder, lung and liver in blackfoot disease endemic area in Taiwan. *Br. J. Cancer* **53**, 399–405.
- CHEN, C. J., WU, M. M., LEE, H. H., WANG, J. D., CHENG, S. H., and WU, H. Y., 1988: Arterogenicity and carcinogenicity of high-arsenic artesian well water: multiple risk factors and related malignant neoplasms of blackfoot disease. *Arteriosclerosis* **8**, 452–460.
- CRUMP, K. S. and HOWE, R. B., 1984: The multistage model with a time-dependent dose pattern: applications to carcinogenic risk assessment. *Risk analysis* **4**, 163–176.
- DEMPSTER, A. P., LAIRD, N. M., and RUBIN, D. B., 1977: Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Stat. Soc. B* **39**, 1–38.
- EFRON, B., 1979: Bootstrap methods: another look at the jackknife. *Ann. Stat.* **7**, 1–26.
- GOODMAN, L. A., 1974: Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika* **61**, 215–231.
- HOLFORD, T. R., 1991: Understanding the effect of age, period, and cohort on incidence and mortality rates. *Ann. Rev. Publ. Health* **12**, 425–457.
- KUPPER, L. L., JANIS, J. M., KARMOUS, A., and GREENBERG, B. G., 1985: Statistical age-period-cohort analysis: a review and critique. *J. Chron. Dis.* **38**, 811–830.
- LEE, W. C. and LIN, R. S., 1995: Analysis of cancer rates using excess risk age-period-cohort models. *Int. J. Epidemiol.* **24**, 671–677.
- MAZUMDAR, S., REDMOND, C. K., ENTERLINE, P. E., MARSH, G. M., COSTANTINO, J. P., ZHOU, S. Y. J., and PATWARDHAN, R. N., 1989: Multistage modeling of lung cancer mortality among arsenic-exposed copper-smelter workers. *Risk Analysis* **9**, 551–563.
- MENG, X. L. and RUBIN, D. B., 1993: Maximum likelihood estimation via the ECM algorithm: a general framework. *Biometrika* **80**, 267–278.
- MOOLGAVKAR, S. H., 1986: Carcinogenesis modeling: from molecular biology to epidemiology. *Ann. Rev. Publ. Health* **7**, 151–169.
- WU, M. M., KUO, T. L., HWANG, Y. H., and CHEN, C. J., 1989: Dose-response relation between arsenic concentration in well water and mortality from cancers and vascular diseases. *Am. J. Epidemiol.* **130**, 1123–1132.
- SILVERMAN, B. W., JONES, M. C., WILSON, J. D., and NYCHKA, D. W., 1990: A smoothed EM approach to indirect estimation problems, with particular reference to stereology and emission tomography. *J. Roy. Stat. Soc. B* **5**, 271–324.

Received, November 1995

Revised, September 1996

Accepted, September 1996

Dr. WEN-CHUNG LEE  
Graduate Institute of Epidemiology  
National Taiwan University  
No. 1, Jen-Ai Rd, 1st Sec.  
Taipei, Taiwan, R.O.C.