Latency Analysis in Epidemiologic Studies of Occupational Exposures: Application to the Colorado Plateau Uranium Miners Cohort

Bryan Langholz, PhD,* Duncan Thomas, PhD, Anny Xiang, PhD, and Daniel Stram, PhD

Background *Latency effects are an important factor in assessing the public health implications of an occupational or environmental exposure. Usually, however, latency results as described in the literature are insufficient to answer public health related questions. Alternative approaches to the analysis of latency effects are warranted.*
Methods *A general statistical framework for modeling latency effects is described. We then propose bilinear and exponential decay latency models for analyzing latency effects as they have parameters that address questions of public health interest. Methods are described for fitting these models to cohort or case-control data; statistical inference is based on standard likelihood methods.*
Application *A latency analysis of radon exposure and lung cancer in the Colorado Plateau uranium miners cohort was performed. We first analyzed the entire cohort and found that the relative risk associated with exposure increases for about 8.5 years and thereafter decreases until it reaches background levels after about 34 years. The hypothesis that the relative risk remains at its peak level is strongly rejected (P < 0.001). Next, we investigated the variation in the latency effects over subsets of the cohort based on attained age, level and rate of exposure, and smoking. Age was the only factor for which effect modification was demonstrated (P = 0.014). We found that the decline in effect is much steeper at older ages (60+ years) than younger.*
Conclusion *The proposed methods can provide much more information about the exposure–disease latency effects than those generally used.* Am. J. Ind. Med. 35:246–256, 1999. *© 1999 Wiley-Liss, Inc.*

KEY WORDS: cohort studies; case-control studies; protracted exposures; effect modification; relative risk; risk assessment; radon; uranium miners

INTRODUCTION

As part of their investigation of the possible effects of herbicides on the risk of cancer in Vietnam veterans, the National Academy of Sciences Agent Orange and Vietnam Veterans Committee was charged to review the scientific literature for information about latency effects. In particular, they were to summarize the current state of knowledge regarding the following latency-related questions:

- How long does it take after exposure to detect an increase in disease risk?
- How long do the effects of exposure last?
- Are such effects modified by age or other factors?

The second and third questions are of particular importance for Vietnam veterans. While it appears that the veterans generally received very low exposures, there is still concern about more highly exposed subgroups and a perpetual

Department of Preventive Medicine, University of Southern California, School of Medicine, Los Angeles, California.
Contract grant sponsor: NCI; contract grant number: CA 42949; Contract grant sponsor: NIOSH; contract grant number: OH 11869.
*Correspondence to: Bryan Langholz, Department of Preventive Medicine, University of Southern California, School of Medicine, 1540 Alcazar Street, CHP 220, Los Angeles, California 90033.

Accepted 14 September 1998

controversy about exposure assessment. However, the timing of exposure can unequivocally be restricted to the range of dates that herbicides were used during the Vietnam War. Thus, much could be said about the risk to exposed veterans (as well as any herbicide exposed group) today if we knew about the evolution of risk with time since exposure, the latency effects. Of course, since exposure information is not available on the veterans themselves and, in any case, they generally had very low exposures, this information needs to be inferred from other, more highly exposed cohorts. The committee concluded that, although some studies reported latency results, they were of limited value in answering the types of latency-related questions raised above [NAS/NRC, 1996].

The difficulty in using the published reports to assess the role of latency on cancer risk from herbicides was that latency effects were quantified as the (relative) risks as a function of time since first exposure. This is a common practice in studies of occupational or environmental exposures and chronic diseases because there is generally a period after exposure in which there is very little increase in risk due to that exposure, followed by an increase in risk. So evidence of such a pattern is often used to establish a causal relationship between the exposure and disease (e.g., Hill, 1965). Such an analysis is generally sufficient to answer the first question raised above. However, if the exposure is protracted, as is the case with most occupational exposures, the time since first exposure analysis will not be informative about the general evolution of risk due to a given exposure history, and cannot be used to answer other latency-related questions such as the second and third questions raised above. While methods exist for describing latency effects for protracted exposures (e.g., Breslow and Day, 1987, Chapter 5), these have rarely been used in practice. We believe this is because 1) the parameters in these models do not directly address the public health questions, and 2) there are technical difficulties in fitting such models to protracted exposure data. Also, latency models have often been developed in the context of mechanistic models for carcinogenesis that make specific predictions about latency effects, rather than as a descriptive tool.

In this article, we focus on descriptive methods for addressing latency questions and propose some simple flexible latency models that have parameters that are directly relevant to questions of public health. We apply these methods to extensively analyze latency effects in the Colorado Plateau uranium miners cohort.

METHODS

Latency Models for Protracted Exposures

The importance of latency effects, as well as other time-related factors, such as attained age and age at exposure, in determining the risk of cancer and other chronic diseases has long been recognized (Armenian, 1987). Early attempts to characterize the latent period were based on the simple idea of tabulating the time from exposure to disease in those who experienced both in order to get a ‘‘latency distribution.’’ This technique was shown to be seriously flawed. In particular, this distribution is completely dependent on the length of follow-up of the study group. The longer the group was followed, the longer the apparent mean latency since new cases would always have longer time to disease than those previously followed (Enterline and Henderson, 1973; Peto, 1985; Thomas, 1987). This phenomenon is precisely the motivation for methods that accommodate censoring in failure time data. Thus, it is natural to apply the methods for censored survival data that serve as the basis for standard statistical methods used in the study of variation in rates used in epidemiologic studies. In this context, latency will be quantified as the evolution of the rate, or more relevantly, the rate ratio relative to unexposed subjects, as a function of time since exposure (Thomas, 1983, 1987; Finkelstein, 1991).

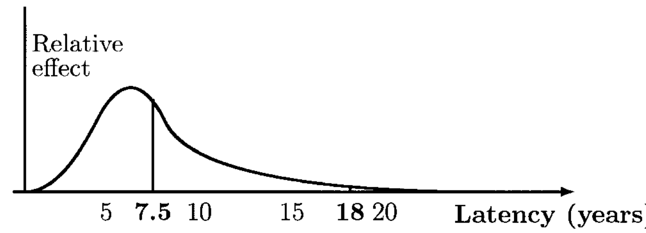

FIGURE 1. The relative effect of exposure as a function of time since exposure (or time in the past): instantaneous exposure.

To start, it is easiest to consider a single ‘‘instantaneous’’ exposure. Exposure to radiation from the atomic bombs at Hiroshima and Nagasaki is an example of this type of exposure. As time passed, leukemia rates among those exposed went up for a time, relative to that expected given their age, then came back down to about ‘‘normal’’ (NAS/NRC, 1990; Curtis and Thomas, 1992). As illustrated in the hypothetical example in Figure 1, the effect of exposure 7.5 years since exposure is about ten times that after 18 years had elapsed. Note that rather than time since exposure, we can equivalently think of the time scale as time in the past and that the latency curve gives the relative effectiveness of an exposure by how long ago in the past it was experienced. This point of view is useful for thinking about protracted exposures. One can think of a protracted exposure as a sequence of instantaneous exposures experienced, say, annually. The total effect of the protracted exposure is then the sum of the annual exposures weighted by the latency curve. For example, consider an individual who is 40 years old and was exposed between ages 29 and 35. Figure 2a gives the annual exposure doses (represented by the vertical lines) for this subject at age 40 as a function of number of years before

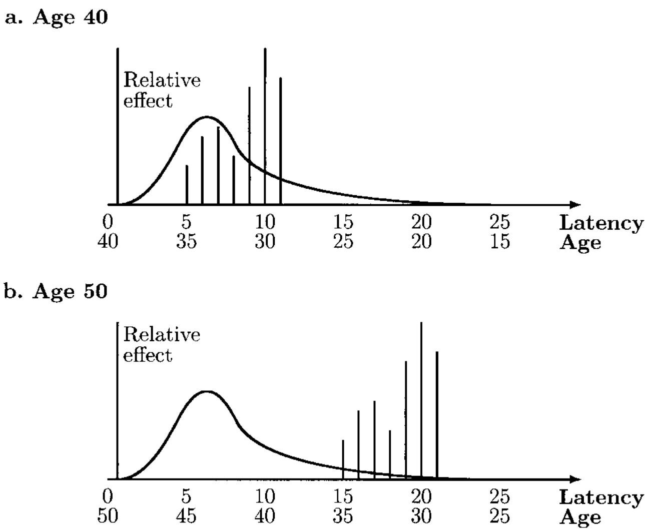

FIGURE 2. The relative effect of exposure as a function of latency (in years) at ages 40 and 50: protracted exposure. Each vertical line represents the one year's exposure level.

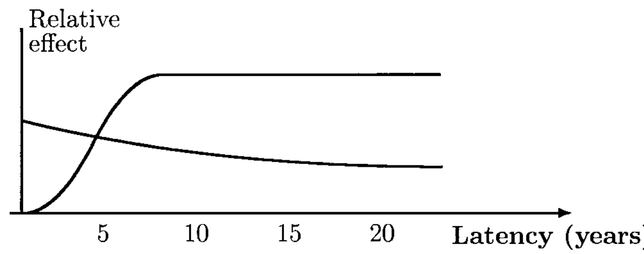

FIGURE 3. Some other possible latency curves.

age 40. On this latency time scale, age decreases with increasing latency with zero latency at 40 years of age. The total effective dose for this exposure history at age 40 would then be the sum of the doses at each age multiplied by corresponding latency curve values. The total effective dose changes with the age of the individual since the overlying latency curve starts at the current age. As illustrated in Figure 2b, 10 years later the total effective dose for this individual is much smaller because of the lower values of the latency curve multiplying the doses. Mathematically, if we let t be the current age and u index the ages at exposure, with d(u) the dose at age u and w $(t - u)$ the latency curve $t - u$ years in the past, then the effective dose at age t from an exposure incurred at age u is $d(u)w(t - u)$. The total effective dose at age t, D(t), is then given by

$$D(t) = \sum_{u} d(u)w(t - u).$$

In one further step, instead of thinking of exposure during increments of time, it is often more natural to express protracted exposure as given continuously in time. The sum above then becomes the integral

$$D(t) = \int_0^t d(u)w(t - u)\, du \qquad (1)$$

with d(u) now the dose rate at age u.

Modeling Disease Rates

The total effective dose summarizes the exposure history into a single value that at each age can be related to disease risk. A form of this relationship must be chosen or determined from the data. For various reasons, theoretical and empirical, radiation effects have been modeled as the excess relative risk as a linear function of total effective dose. This is expressed as

$$\lambda(t, D(t)) = \lambda_0(t)(1 + \beta D(t)) \qquad (2)$$

where $\lambda_0(t)$ is the rate of disease at age t in a (comparable) unexposed population and β is the increase in relative risk per unit effective dose. Combining this dose–response model with the total effective dose formula (1), we note that

$$\beta D(t) = \beta \int_0^t d(u)w(t - u)\, du = \int_0^t d(u)\beta w(t - u)\, du \qquad (3)$$

so that $\beta w(t - u)$ gives the excess relative risk per unit dose ascribed to exposure at $t - u$ years in the past.

Other forms for the rates may be more appropriate in other situations, such as the log linear form (Cox model)

$$\lambda(t, D(t)) = \lambda_0(t) \exp(\beta D(t))$$

or more complex relationships. In one simple alternative which we will use, the parameters are the excess relative risks in categories of exposure. Categorizing D(t) into ''dose intervals'' C_k, the rates of disease for those in dose category k is given by

$$\lambda(t, C_k) = \lambda_0(t)(1 + \beta_k). \qquad (4)$$

Descriptive Latency Models

Having introduced how to conceptualize the effect of latency on disease risk, we now describe the framework for investigating the shape of the latency curve.

First, we note that the shape of the curve given in Figure 1 is not the only possibility. Some alternatives, illustrated in Figure 3, are that the effect of exposure only increases, levels off at a maximum with time since exposure, or jumps immediately after exposure. Note that the use of total cumulative exposure as the effective dose, as is a common practice, is equivalent to assuming a latency curve that is constant, and equal to one, over all time in the past. Thus, in choosing models for the latency curve, we want to have flexibility to accommodate some variation in shape. Further, we want latency models that have parameters that have a simple interpretation and, in particular, are well adapted for addressing the public health questions we have discussed above. This is different from mechanistic modeling, where

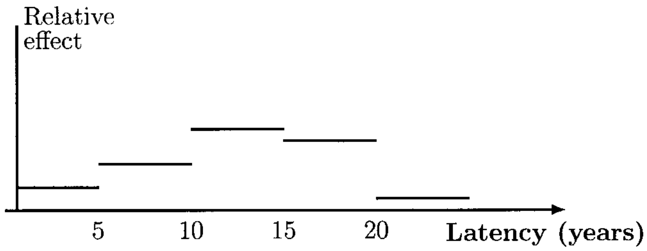

FIGURE 4. Piecewise constant latency model.

the goal is to model latency (and other factors) within the context of a presumed model of carcinogenesis, e.g., the Armitage-Doll multistage model (Armitage and Doll, 1961) or Moolgavkar-Knudsen two-stage model (Moolgavkar and Venzon, 1979). The goal here is to explore latency empirically. Thus, after using the simple step function to get a general idea of the shape of the latency curve, we will favor the bilinear or exponential decay models described below because of their flexibility and because the parameters are interpretable in ways that are informative about the public health questions. A mathematical description of these models as they would be used in the data analysis is given in the Appendix. Here, we give an informal description of the curves and their parameterization.

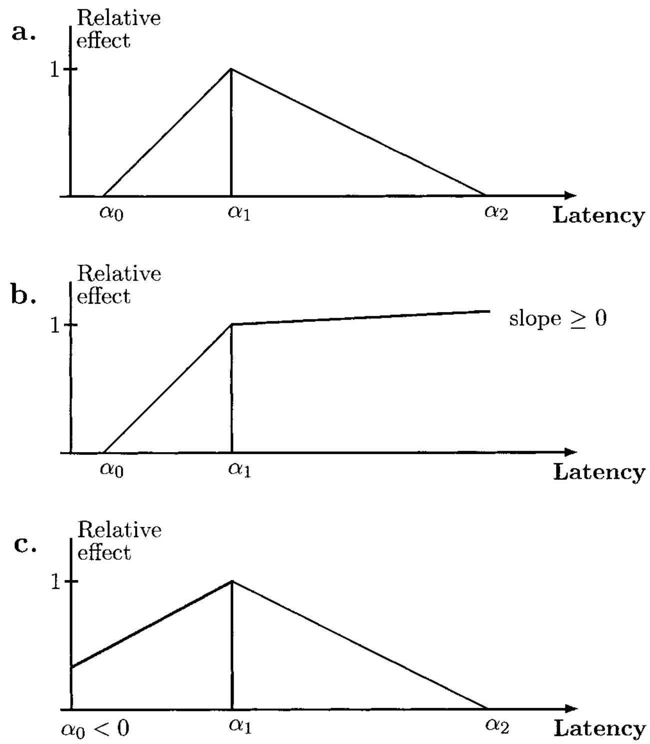

FIGURE 5. Bilinear latency model variations.

Piecewise Constant Model

To get a general sense of the shape of the latency curve, a simple piecewise constant function over ''time windows'' is appropriate (Finkelstein, 1991). This shape is illustrated in Figure 4. This pattern is certainly not a realistic representation of the latency curve, in that changes in risk occur continuously and not in ''jumps.'' Further, we have found that the estimated height of the steps can be statistically very unstable. This is because exposures in adjacent intervals tend to be highly correlated. In spite of this, if there are not too many intervals the steps will give a data-driven overall impression of the shape of the curve. It is intended as a ''first pass'' model to get a sense of the shape of the latency curve. Also, since the height of the steps are unconstrained, the other models are approximately submodels (nested) in this one so it can serve as a base model to test the fit of the others. The step heights are interpretable as the excess relative risks per unit dose received during the respective latency intervals. From formula (3), the total excess relative risk associated with an exposure history is the sum of the excess relative risks for exposure over each constant piece.

Bilinear Model

The bilinear model is a simple descriptive latency model consisting of two attached straight lines, with the height of the curve at the inflection point constrained to be one. Three potential shapes accommodated by the bilinear model are shown in Figure 5. The first, displayed in Figure 5a, is characterized by three times on the latency scale. Up to time α_0 there is no effect of exposure. Then the relative effect increases linearly, reaching a peak α_1 years in the past and decreases linearly thereafter, reaching zero (no effect) at α_2 years in the past. This parameterization has the advantage that it directly addresses the questions of how long does it take before there is an effect of exposure (α_0) and, assuming that the effect eventually decreases with time, how long does the effect last (α_2). Further, α_1 is interpretable as the time at which the effect of a given exposure is maximum. In the excess relative risk model, since the peak value of the bilinear curve is one, β is interpretable as the maximum excess relative risk; this is the excess relative risk attributable to exposures experienced at about α_1 years in the past. This parameterization cannot accommodate latency curves that do not decrease after the peak, as in Figure 5b. However, this can be accomplished by parameterizing the second line by its slope rather than by α_2. In fact, the important question of whether the effect of exposure ever decreases can be formulated as the hypothesis that the slope of the second line is zero or greater. This can be tested using the one degree of freedom chi-square likelihood ratio test comparing the fit of the bilinear model with the slope of the second line set equal to zero to the fit of the bilinear model with the slope (or, equivalently, α_2) unconstrained. An increase in risk immediately after exposure is illustrated in Figure 5c. This shape is characterized by a bilinear model with $\alpha_0 < 0$. Thus, if the

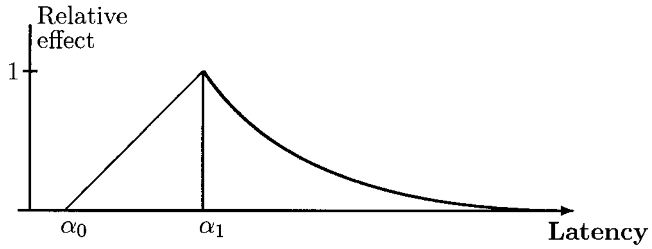

FIGURE 6. Exponential decay of relative risk after peak as a function of latency.

estimate of α_0 is found to be (statistically significantly) smaller than zero, this is evidence that disease risk increases immediately after exposure. On the other hand, if $\alpha_0 > 0$, this is evidence that there is a period of no increased risk.

Exponential Decay Model

While we find the bilinear model very appealing because it directly addresses the public policy questions, it may be reasonable to object to the presumption that the effect of exposure would ever entirely disappear. For example, the herbicide contaminant 2,3,7,8-tetrachlorodibezo-p-dioxin (TCDD) is retained in fat tissue and is very slowly released and eliminated over time. Thus, even if the TCDD exposure ended immediately, there are years of continued (low) exposure even after the original exposure has ceased. Thus, we propose an alternative model which simply replaces the second line by an exponential decay curve. This is illustrated in Figure 6. The parameters α_0 and α_1 are as in the bilinear model. The exponential part is parameterized by the ‘‘half life,’’ the number of years required for the effect to be reduced by half. Thus, given a presumed maximum exposure and the β, one could easily calculate the residual effect after a given amount of time.

Fitting the Models

Nested Case-Control Sampling

In principle, any of the standard methods for analysis of cohort data could be used to perform the latency analyses. For instance, the ‘‘person-years approach’’ using Poisson regression may be used by finely grouping person-year ‘‘cells’’ on year of birth, age, time since exposure, and exposure level and then properly weighting the exposure levels by the latency curve. We prefer to use the Cox regression approach in which case-control sets are formed at each lung cancer death (Cox, 1972). The latency-weighted exposures are then individual-specific, thus avoiding the imprecision introduced by the grouping done in the person-years method. Further, this approach provides a natural way to reduce the size of the analysis dataset by random sampling from the ‘‘controls’’ for each case (e.g., Lubin and Gail, 1984; Langholz and Goldstein, 1996). Nested case-control sampling, as this method is called, is particularly useful in our analysis of the Colorado Plateau uranium miners because the latency model computations take an inconvenient amount of time for the entire cohort. Standard conditional logistic regression methods for matched case-control data are used to analyze the data with covariates computed at the age of death of the case. In general, one should choose as many controls as possible, while still being able to do an analysis in a reasonable amount of time. In fact, if it is possible to use all controls, this should be done. It is still computationally much faster to have the data organized in case-control sets and use conditional logistic regression than to analyze with the cohort data directly using Cox regression, although they are the same analysis. This is because, in the case-control arrangement, D(t) is calculated once for each set of latency parameters, whereas it is computed multiple times as a time-dependent covariate in Cox regression.

Estimating the Latency Model Parameters

Estimation of parameters from the bilinear and exponential decay latency models poses special technical problems because the parameters are not a function of some simple summary of the exposure history. The search for the maximum likelihood estimator involves computing a different effective dose D(t) for each parameter combination. Our approach to fitting these models was to estimate the dose–response (β) parameters over a grid of latency parameters and then to tabulate the likelihoods (or, equivalently, the deviances). (The deviance is -2 times the log likelihood. The likelihood ratio test is the difference between the deviances for nested models.) The values of the latency parameters which gave the lowest deviances are the maximum likelihood estimates. In order to estimate the standard errors and confidence intervals for the latency parameters and for the slope parameter in the excess relative risk model (2), which properly takes into account the estimation of the latency parameters, we calculated the covariance matrix as the inverse ‘‘expected information’’ (Thomas, 1981). This required the derivatives of D(t) with respect to the latency parameters and β (described in the Appendix). All computations were implemented using the statistical package Epicure (Hirosoft International Inc., Seattle, WA). The ‘‘scripts’’ used for the uranium miners cohort data analyses are available from the authors. We note, however, that aside from the restriction to a log-linear dose response model, these methods are relatively easy to implement in other packages such as SAS or S-Plus.

APPLICATION: LATENCY ANALYSIS OF RADON EXPOSURE IN THE COLORADO PLATEAU URANIUM MINERS

The Cohort Study

The Colorado Plateau uranium miners cohort was assembled to study the effects of radon exposure and smoking on mortality rates and has been described in detail in earlier publications (e.g., Lundin et al., 1971; Hornung and Meinhardt, 1987). The cohort consists of 3,347 Caucasian male miners recruited between 1950 and 1960 and was traced for mortality outcomes through December 31, 1990 (Roscoe, 1997). For reasons discussed below, we restricted the cohort to 2,704 miners whose first employment as uranium miners occurred after 1950. In this group, there were 263 lung cancer deaths and 716 deaths from other causes at the end of follow-up. Latency effects of radon exposure have been explored in previous publications using the piecewise constant latency model (Thomas et al., 1994; Lubin et al., 1994). In particular, the latter publication gives the results of a joint analysis of 11 uranium miner cohorts undertaken by researchers at the United States National Cancer Institute. Using the piecewise constant model described in the preceding section, it was found that the excess relative risk increases for 5–10 years after exposure, after which it slowly decreases. Independent of this latency effect, there was a decrease in effect with attained age. This means, for example, an identical exposure experienced 10 years in the past increases the relative risk of lung cancer in 50-year-olds to a much greater degree than in 65-year-olds.

Radon Exposure Histories

Annual radon exposures, in working level months (WLMs) [NAS/NRC, 1988, p. 27], were estimated in an exposure reconstruction undertaken by members of our group by linking radon level measurements taken (or estimated) in the mines to miner work histories. Because there were no measurements recorded prior to 1950, we felt that we could not reliably estimate the exposures during this time. Thus, we restricted our analysis to those miners who began working in the mines after 1950. Details about the exposure reconstruction process are available in a technical report (Stram et al., 1998). In computing any of the exposure history summaries, we ‘‘lagged’’ exposure by two years, i.e., exposures were accumulated only up to two years prior to the reference age, to crudely approximate exposure up to diagnosis of lung cancer. This was done for two reasons. First, exposure accrued between diagnosis and death could not have contributed to the occurrence of the lung cancer and is, thus, irrelevant. Second, and more importantly, miners would quit work after diagnosis and thus stop accruing exposure because of the occurrence of disease. Meanwhile, healthy workers would continue to work and accrue exposure. To include exposures during this period would thus result in a spurious ‘‘protective effect’’ of recent exposure.

Smoking Histories

Upon being enrolled into the study, miners were asked about their smoking histories, including their age at start of smoking and smoking levels (in packs per day). During medical examinations of the miners, undertaken periodically from 1952 into the 1960s, current smoking information was recorded and changes in level were included in the dataset. This information was summarized as cumulative smoking over 5-year intervals. It was assumed that the last reported smoking level was maintained thereafter.

The Nested Case-Control Sample

As discussed above, we drew a nested case-control dataset with 40 controls who were ‘‘on study’’ at the age of the case (alive at, and enrolled into the cohort by, the age of death of the case). To account for calendar trends in lung cancer mortality, we also restricted controls to subjects who had attained the age of the case during the same 5-year calendar period. For cases that had fewer than 40 eligible controls, all controls were used. As is required of the method, subjects are allowed to serve as controls for more than one case. Thus, the analysis dataset had 10,322 ‘‘case-control records’’ from 2,239 distinct cohort subjects, with 239 cases matched to 40 controls each while the remaining 24 cases had fewer than 40 potential controls each and all were used. This is the same case-control set used by Stram et al. (1999) to study the effects of measurement error on time-related effects.

RESULTS

Table I gives a comparison of pooled cases and controls from the nested case-control dataset by age of death of the case <60 or ≥ 60. This case-control comparison gives a descriptive look at data as it contributes in the estimation of these time-related effects. Aside from the observation that the cases were more highly exposed than the controls, these data show that there is good variation in timing of exposure in both age groups. For instance, ‘‘Timing of exposure’’ gives the proportion of subjects who experienced at least 90% of their total cumulative exposure during the past 20 years (the recent past), at least 90% of their total cumulative exposure during the period over 20 years in the past (the distant past), and those who had less than 90% in either (intermediate). There is a good distribution across these categories over case and controls for each age group.

TABLE I. Descriptive Statistics Comparing the Distribution of Radon and Smoking Exposure in Cases and Controls From the Colorado Plateau Uranium Miners Data

	Age <60		Age ≥60	
	Controls	Cases	Controls	Cases
Number of records[a]	5,360	134	4,699	129
Total cumulative radon exposure (WLM)				
1st quartile	155	509	161	321
2nd quartile	381	946	385	679
3rd quartile	816	1,894	806	1,150
Cumulative radon exposure during 0–9 years of latency (WLM)				
No exposure	66%	57%	80%	77%
Median among exposed	186	227	119	178
Cumulative radon exposure during 10–19 years of latency (WLM)				
No exposure	38%	28%	47%	39%
Median among exposed	172	499	182	235
Cumulative radon exposure during 20+ years of latency (WLM)				
No exposure	45%	43%	30%	28%
Median among exposed	284	747	304	521
Timing of exposure[b]				
90% in 0–19 years of latency	48%	46%	32%	29%
90% in 20+ years of latency	34%	31%	51%	47%
<90% in either	18%	22%	17%	24%
Exposure rate during previous 30 years (WL)[c]				
1st quartile	2.7	5.7	2.6	3.8
2nd quartile	5.3	8.8	5.1	6.2
3rd quartile	9.4	13.3	8.9	10.5
Smoking				
Nonsmokers	22%	7%	23%	11%
Amount smoked among smokers (100s of packs)				
1st quartile	81	82	105	134
2nd quartile	116	113	155	165
3rd quartile	141	146	190	208

All variables are computed up to 2 years prior to the case's age of death.
[a]Subjects may be controls in multiple case-control sets.
[b]Based on the percentage of total exposure within the latency period.
[c]Computed as the total exposure during the past 30 years divided by the time exposed.

Figure 7 shows the estimated excess relative risk per 100 WLM as a function of latency (formula (3)) as predicted by the bilinear model and a six-step piecewise constant model with intervals over 0–4, 5–9, 10–14, 15–19, 20–29, and 30+ years of latency. The plot indicates that the bilinear model fits the data well, an observation confirmed by the likelihood ratio test comparing the fit (deviance) of the bilinear model in the second row of Table II to that of the step function in the first row ($\chi^2_2 = 1803.7–1802.5 = 1.2$, $P > 0.5$). The maximum likelihood estimates and 95% confidence intervals for the bilinear model parameters are: $\hat{\alpha}_0 = -3$ (-13.2–7.2), so there is no evidence of a ''no-excess risk period,'' that the peak occurs at $\hat{\alpha}_1 = 7.5$ (0–17.8) years after exposure, and that risk returns to background after $\hat{\alpha}_2 = 33.5$ (29.6–37.4) years after exposure. The estimated increase in excess relative risk experienced near the peak of the latency curve (7.5 years in the past) is estimated to the $\hat{\beta} = 0.80$ (0.21–1.38) per 100 WLM. Table II summarizes the results of the analysis in terms of the deviances for the fitted models. The difference in the deviances for nested models is a chi-square test with number of degrees of freedom equal to the difference in the number of parameters used. Fixing $\alpha_0 = 0$ results in very little reduction in fit ($\chi^2_1 = 1803.8 - 1803.7 = 0.1$, $P > .5$) so, to simplify the analyses, we fixed $\alpha_0 = 0$ in all subsequent models. The remaining parameter estimates (and 95% confidence intervals) are virtually unchanged; $\hat{\alpha}_1 = 8.5$ (1.5–15.5), $\hat{\alpha}_2 = 34$ (29.9–38.1), and $\hat{\beta} = 0.74$ (0.25–1.23). The ''no decrease model,'' which tests whether latency curve shapes as in Figure 5b are consistent with the data, is clearly rejected ($P < 0.0001$) and so, of course, is the total cumulative exposure model. For the categorical dose–response model, the categories were determined by the quartiles of the case distribution of the latency weighted exposures. The resulting five degree of freedom model fits somewhat better than the linear dose–response form ($P = 0.052$). This is because the dose–response curve is less than linear at high exposures. In any case, the latency parameters remain virtually unchanged in the categorical dose–response model with $\hat{\alpha}_1 = 8.5$, $\hat{\alpha}_2 = 33$.

Finally, we fitted the exponential decay model with α_0 set to zero. The estimated peak was $\hat{\alpha}_1 = 7.5$ years after exposure with excess relative risk slope at the peak $\hat{\beta} = 0.54/100$ WLM. Thereafter, the excess relative risk is estimated to decrease by half every 8 years.

Effect Modification

Next, we explored the variation in latency over attained age, level of exposure, rate of exposure, and smoking. This was done by dichotomizing each factor (according to the median value in the case distribution of the variable) and comparing the bilinear model parameters across the two levels of the variable. The nested case-control dataset was split into two disjoint sets which are homogeneous for the dichotomous variable. Controls who do not match the case on the variable are dropped from the analysis. Thus, with the exception of attained age, for which all the controls match the case, the cases in the resulting ''post-stratified'' dataset have a subset of the 40 controls from the unstratified dataset. Three models were then fitted to the two strata, the deviance from each model is given in Table III. The Null model deviance is from the three parameter bilinear model (with $\alpha_0 = 0$) with parameters set to that of the unstratified

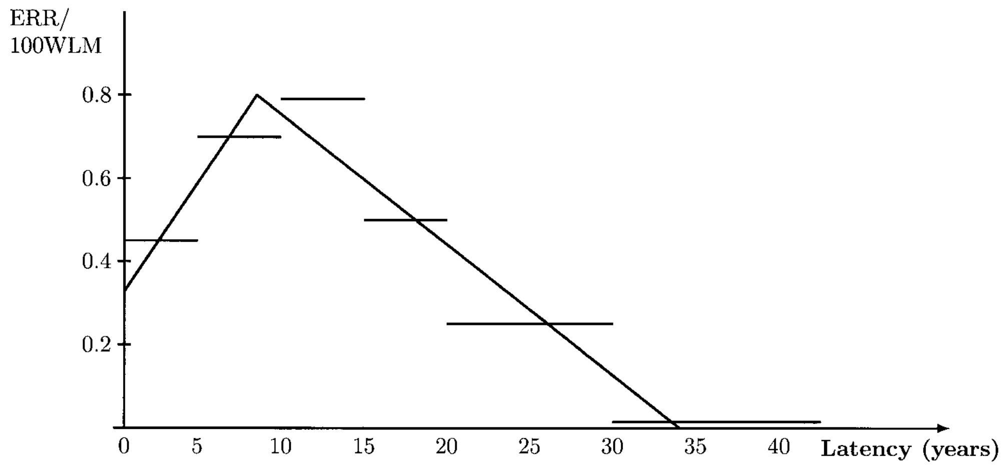

FIGURE 7. Excess relative risk as a function of latency from fitted piecewise constant and bilinear latency models. Estimated from the Colorado Plateau uranium miners data.

TABLE II. Analysis of Deviance for Comparison of Latency Models Fitted to the Colorado Plateau Uranium Miners Data

Latency model	Model degrees of freedom	Model deviance[a]	Likelihood ratio statistic	Degrees of freedom	P-value
Piecewise constant	6	1,802.5			
Bilinear (full)	4	1,803.7	1.2[b]	2	>0.5
Bilinear ($\alpha_0 = 0$)	3	1,803.8	0.1[c]	1	>0.5
No decrease in effect	2	1,815.5	11.7[d]	1	<0.001
Cumulative dose	1	1,815.7	11.9[d]	2	<0.001
Bilinear with categorical dose-response[e]	5	1,797.8	6.0[d]	2	0.05
Exponential decay	3	1,806.1	3.6[b]	3	0.31

The dose–response model is the excess relative risk form, except for the bilinear with categorical dose–response.
[a]The difference in deviance between nested models is a chi-square test with degrees of freedom equal to the difference in the model degrees of freedom.
[b]Compared to the piecewise constant model.
[c]Compared to the full bilinear model.
[d]Compared to the bilinear ($\alpha_0 = 0$) model.
[e]See equation (4).

analysis. Note that, except for age, the deviance will be smaller than that given for the unstratified analysis in the "Bilinear model $\alpha_0 = 0$" line of Table II because of the smaller number of controls per case. The second model (slope) holds α_1 and α_2 at the unstratified analysis estimates but the slope (β) parameter is estimated for each of the two strata separately. In the final model (all), all three parameters are estimated for each stratum. Likelihood ratio tests are given by the difference in deviances with degrees of freedom equal to the number of additional parameters in the respective models. However, we are primarily interested in whether there is evidence of a difference in the latency parameters after allowing for different slopes, as given by the two degree of freedom test comparing the second and third models. As seen in Table III, there is evidence of interaction across age, exposure, and exposure rate, but none across smoking level. For both exposure level and exposure rate, the interaction can be attributed to a difference in slopes, but not to a difference in latency parameters, across strata. For age, there is some evidence of interaction in the latency parameters after accounting for differences in slope ($1800.4 - 1791.9 = 8.5$, $P = 0.014$). Table IV gives the model parameter estimates within four age groups. Although the estimates are rather unstable because of the relatively small number of cases in each age group, it is clear that there is a decrease in α_2 and in β with increasing age. Recognizing that the under-60 age group was constrained with respect to the amount of time that could elapse between exposure and current age, we were concerned about the possibility that the decrease observed in the 60+ group was attributable solely to the lack of association with exposures far in the past. Thus, we fitted the latency model using exposures accrued

TABLE III. Deviance for Interactions With Bilinear Latency Parameters

Model	Degrees of freedom	Deviance[a]	Likelihood ratio statistic	Degrees of freedom	P-value
Attained age stratified (60 years)					
Null[b]	—	1,803.8			
Age*Bilinear (slope[c])	1	1,800.4	3.4	1	0.065
Age*Bilinear (all[d])	3	1,791.9	8.5	2	0.014
Exposure stratified (700 WLM)					
Null	—	1,484.3			
Exposure*Bilinear (slope)	1	1,477.1	7.2	1	0.007
Exposure*Bilinear (all)	3	1,473.5	3.6	2	0.17
Exposure rate stratified (7.5 WL)					
Null	—	1,475.6			
Rate*Bilinear (slope)	1	1,464.9	10.7	1	0.001
Rate*Bilinear (all)	3	1,461.4	3.5	2	0.17
Smoking stratified (13,500 packs)					
Null	—	1,491.3			
Smoking*Bilinear (slope)	1	1,490.1	1.2	1	0.27
Smoking*Bilinear (all)	3	1,488.8	1.3	2	0.52

The interaction variable is dichotomized at the level given in parentheses. The analysis is from the ‘‘post-stratified’’ Colorado Plateau uranium miners nested case-control dataset.
[a]Because of the different numbers of subjects involved, these deviances cannot be compared to those in Table II.
[b]Bilinear latency model and excess relative risk parameters fixed to those from the entire dataset ($\alpha_0 = 0$, $\alpha_1 = 8.5$, $\alpha_2 = 34$, $\beta = 0.74$).
[c]Bilinear latency model parameters fixed to the entire dataset values, a separate excess relative risk slope is estimated for each stratum.
[d]Bilinear latency model parameters and excess relative risk slope estimated for each stratum.

TABLE IV. Bilinear Latency Model and Linear Excess Relative Risk Parameter Estimates Over Age Groups From the Colorado Plateau Uranium Miners Data

	Attained age group			
Parameter	<50	50–59	60–69	70+
Bilinear model				
α_1	7	8	21	18
α_2	50	46	29	28
β	1.02	0.74	0.14	0.10
Exponential decay model				
α_1	11	8	14	10
Half-life	10	9	6	5
β	0.93	1.16	0.09	0.25

only in the past 30 years. We found that the decrease in excess relative risk with latency persisted.

DISCUSSION

After establishing that an association between an exposure and disease exists, characterizing the evolution of the increased risk with timing of exposure is important in understanding the health implications to exposed individuals. The bilinear model provides a flexible framework for exploring latency-related questions. It accommodates the latency patterns one would expect to encounter and the parameters are directly interpretable in terms of important public health issues. Other latency models may be more appropriate, depending on the goals of the analysis. In particular, latency curves predicted by a proposed mechanistic model would be of interest in the context of an investigation of etiology. For instance, the multistage model of carcinogenesis predicts particular age-dependent latency curves, depending on the stage of the process affected (Thomas, 1982).

In order to estimate the parameters of the latency curve with much precision, there must be variation in the rate and duration of exposure in the study group. As seen in Table I, there is a fair amount of detail and variation in the exposure histories for the Colorado Plateau uranium miners data. In many cases, intervals of employment in a job that had probable exposure may be known, but with no measure of level of exposure. Such a history may still be used to investigate latency effects with ‘‘dose-rate’’ d(u) one or zero, depending on whether the worker was employed or not at age u. The parameter β in the bilinear model is then interpreted as the maximum excess relative risk associated with employment at α_1 years in the past. We would predict that at the time of peak relative effectiveness, α_1 will be especially sensitive to the imprecision of the exposure histories, with α_0 and α_2 less so, since they depend more heavily on times of first and last exposures, which are usually measured reasonably well.

In our example, we took a nested case-control sample from a cohort study in order to reduce the computational effort. Of course, the methods for the analysis of latency apply without change to matched case-control studies.

In our analysis of the Colorado Plateau uranium miners cohort, one novel finding is that the strong decline in excess relative risk with latency reported in the NCI monograph (Lubin et al., 1994) appears to be restricted to risk of lung cancer at older ages (over 60). The latency curves decline slowly for ages less than 60, indicating that all past exposure contributes to current risk of lung cancer. Over age 60, we find a much steeper decline; exposures experienced more than about 30 years in the past are no longer relevant to lung cancer risk.

Although simple stratification by potential modifying factors allows for some exploration of the subtleties involved when there are complex exposure–time relationships, there is potential for misinterpretation of latency findings. First, because there is some correlation between age at exposure and latency, in general it is possible that age at exposure is the more important aspect of timing. For the Colorado Plateau, this did not seem to be the case (Lubin et al., 1994). However, age at exposure may modify latency effects in more subtle ways and descriptive tools are needed to investigate this possibility. One assumption that is implicit in our description of latency effects is that the total effective exposure is the sum of weighted component exposures. This dose additivity assumption implies that the increase in risk due to an exposure, say, 5 years in past, is the same whether this is the only exposure experienced or it is the 25th year of exposure. For instance, the incremental increase in the risk of lung cancer from recent cigarette smoking is much higher in ‘‘long-term smokers’’ than those who just started. This is because smoking acts at both an early and late stage of the carcinogenic process so that long-term smokers have a larger pool of cells that have undergone the multiple transformations needed to become cancer cells than short-term smokers (Doll, 1978; Brown and Chu, 1987). Further work is needed to develop simple tests of the dose-additivity assumption and models that quantify the degree to which it is violated.

The interpretation of latency effects analyses may be complicated by the trends in the imprecision of measurements in time. For example, the measurements of radon in uranium mines became more frequent and systematic with time so that, generally, we would expect a miner’s exposure history to be less accurate farther into the past. The implications of such a trend in the precision of the measurements on estimates of latency parameters are not obvious. For the Colorado Plateau uranium miners, Stram et al. (1999) generated measurement error-adjusted exposure histories using published estimates of the degree of imprecision with time. In order to get a sense of the impact of measurement error correction on our results, we fitted bilinear model using these adjusted exposure histories. We found that, while the slope estimate $\hat{\beta}$ increased by about 40%, the latency curve parameters did not change by more than 1 year. This indicates that, for the Colorado Plateau miners cohort, measurement error affects the estimated magnitude of the association with exposure, but not the latency characteristics.

ACKNOWLEDGMENT

Dr. Mark Huberman is acknowledged for his technical and programming assistance.

REFERENCES

Armenian H. 1987. Incubation periods in cancer epidemiology. J Chron Dis 40:9–16.

Armitage P, Doll R. 1961. Stochastic models for carcinogenesis. In: Neyman J, editor. Proceedings of the 4th Berkeley symposium on mathematical statistics and probability. Berkeley: University of California Press, pp 19–38.

Breslow NE, Day NE. 1987. Statistical methods in cancer research. Volume II—The design and analysis of cohort studies. IARC Scientific Publications, Vol. 82. Lyon: International Agency for Research on Cancer.

Brown CC, Chu KC. 1987. Use of multistage models to infer stage affected by carcinogenic exposure: Example of lung cancer and cigarette smoking. J Chron Dis 40 (Suppl 2):171S–179S.

Cox DR. 1972. Regression models and life-tables (with discussion). J R Stat Soc B 34:187–220.

Curtis B, Thomas D. 1992. Dose-time-response models for radiation carcinogenesis. Adv Radiat Biol 16:45–76.

Doll R. 1978. An epidemiologic perspective of the biology of cancer. Cancer Res 38:3573–3583.

Enterline P, Henderson V. 1973. Type of asbestos and respiratory cancer in the asbestos industry. Arch Environ Health 27:312–317.

Finkelstein M. 1991. Use of ‘‘time windows’’ to investigate lung cancer latency intervals at an Ontario steel plant. Am J Ind Med 19:229–235.

Hill A. 1965. The environment and disease: Association or causation? Proc R Soc Med 58:295–300.

Hornung R, Meinhardt T. 1987. Quantitative risk assessment of lung cancer in U.S. uranium miners. Health Phys 52:417–430.

Langholz B, Goldstein L. 1996. Risk set sampling in epidemiologic cohort studies. Stat Sci 11:35–53.

Lubin JH, Gail M. 1984. Biased selection of controls for case-control analyses of cohort studies. Biometrics 40:63–75.

Lubin J, Boice J, Edling C, Hornung R, Howe G, Kunz E, Kusiak R, Morrison H, Radford E, Samet J, Tirmarche M, Woodward A, Xiang Y, Pierce D. 1994. Radon and lung cancer risk: A joint analysis of 11 underground miners studies. Bethesda, MD: U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health. NIH Publication 94-3644.

Lundin F, Wagoner J, Archer V. 1971. Radon daughter exposure and respiratory cancer, quantitative and temporal aspects, joint monograph 1. Washington, DC: U.S. Public Health Service.

Moolgavkar S, Venzon D. 1979. Two-event models for carcinogenesis: Incidence curves for childhood and adult tumors. Math Biosci 47:55–77.

NAS/NRC. 1988. Health risks of radon and other internally deposited alpha-emitters, BEIR IV. Committee on the Biological Effects of Ionizing Radiation, Washington, DC: National Academy Press.

NAS/NRC. 1990. Health effects of exposure to low levels of ionizing radiation, BEIR V. Committee on the Biological Effects of Ionizing Radiation, Washington, DC: National Academy Press.

NAS/NRC. 1996. Veterans and Agent Orange: Health effects of herbicides used in Vietnam—1996 update. Committee to Review the Health Effects in Vietnam Veterans of Exposure to Herbicides, Washington, DC: National Academy Press.

Peto J. 1985. Some problems in dose-response estimation in cancer epidemiology. In: Voug V, Butler G, Hoel D, Peakall D, editors. Methods for estimating risk of chemical injury: Human and non-human biota and ecosystems. New York: John Wiley & Sons, pp 361–380.

Roscoe R. 1997. An update of mortality from all causes among white uranium miners from the Colorado Plateau study group. Am J Ind Med 31:211–222.

Stram D, Langholz B, Huberman M, Thomas D. 1999. Correcting for dosimetry error in a reanalysis of lung cancer mortality for the Colorado Plateau uranium miners cohort. Health Phys (in press).

Stram D, Langholz B, and Thomas D. 1998. Measurement error correction of lung cancer risk estimates in the Colorado Plateau cohort. I. Dosimetry analysis. Division of Biostatistics Technical Report 126, University of Southern California, School of Medicine, Department of Preventive Medicine.

Thomas D. 1981. General relative-risk models for survival time and matched case-control analysis. Biometrics 37:673–686.

Thomas D. 1982. Temporal effects and interactions in cancer: Implications of carcinogenic models. In: Prentice R, Whittemore A, editors. Environmental epidemiology: Risk assessment. Philadelphia: Society for Industrial and Applied Mathematics, pp 107–121.

Thomas D. 1983. Statistical methods for analyzing effects of temporal patterns of exposure on cancer risks. Scand J Work Environ Health 9:353–366.

Thomas D. 1987. Pitfalls in the analysis of exposure-time-response relationships. J Chron Dis 40(Suppl 2):71–78.

Thomas D, Pogoda J, Langholz B, Mack W. 1994. Temporal modifiers of the radon-smoking interaction. Health Phys 66:257–262.

APPENDIX

Mathematical Specifications of the Latency Curves

On the latency time scale v, for the bilinear model,

$$w(v) = \begin{cases} (v - \alpha_0)/(\alpha_1 - \alpha_0) & \text{if } \alpha_0 < v \le \alpha_1 \\ (\alpha_2 - v)/(\alpha_2 - \alpha_1) & \text{if } \alpha_1 < v \le \alpha_2. \\ 0 & \text{otherwise} \end{cases}$$

With α_2 the ‘‘half life’’ parameter, for the exponential decay model,

$$w(v) = \begin{cases} (v - \alpha_0)/(\alpha_1 - \alpha_0) & \text{if } \alpha_0 < v \le \alpha_1 \\ \exp\left[-(v - \alpha_1) \times \log(2)/\alpha_2\right] & \text{if } \alpha_1 < v \\ 0 & \text{otherwise} \end{cases}.$$

Expected Information Matrix

The formula for the expected information from general relative risk models is given in Thomas [1981], equation (6). For the linear excess relative risk model, equation (2), the relative risk function is $r(D(t);\beta,\alpha_0,\alpha_2,\alpha_2) = 1 + \beta D(t;\alpha_0,\alpha_1,\alpha_2)$ so that the derivative of r with respect to the latency parameters requires the derivatives of D(t). For the bilinear latency model these are:

$$\frac{dD(t)}{d\alpha_0} = \frac{1}{\alpha_1 - \alpha_0} \cdot \left[\int_{\max(0,\alpha_0)}^{\alpha_1} d(t - v)\frac{v - \alpha_0}{\alpha_1 - \alpha_0}\,dv - \int_{\max(0,\alpha_0)}^{\alpha_1} d(t - v)\,dv\right],$$

$$\frac{dD(t)}{d\alpha_1} = \frac{-1}{\alpha_1 - \alpha_0}\int_{\max(0,\alpha_0)}^{\alpha_1} d(t - v)\frac{v - \alpha_0}{\alpha_1 - \alpha_0}\,dv + \frac{1}{\alpha_2 - \alpha_1}\int_{\alpha_1}^{\alpha_2} d(t - v)\frac{\alpha_2 - v}{\alpha_2 - \alpha_1}\,dv,$$

and

$$\frac{dD(t)}{d\alpha_2} = \frac{1}{\alpha_2 - \alpha_1}\left[\int_{\alpha_1}^{\alpha_2} d(t - v)\,dv - \int_{\alpha_1}^{\alpha_2} d(t - v)\frac{\alpha_2 - v}{\alpha_2 - \alpha_1}\,dv\right].$$

Given the computer code that calculates the bilinear weighted total exposure and that calculates the total exposure over specified latency intervals, the integrals in these expressions are easily computed. Two of the integrals in these expressions are the latency-weighted exposures for the two parts of the bilinear ‘‘triangle.’’ Thus, they may be calculated by saving the integrated weighted exposure up to α_1. The other two integrals are the total exposures over the latency intervals from max $(\alpha_0, 0)$ to α_1 and from α_1 to α_2.

With α_2 the half-life parameter, for the exponential decay latency model the derivatives are:

$$\frac{dD(t)}{d\alpha_0} = \frac{1}{\alpha_1 - \alpha_0} \cdot \left[\int_{\max(0,\alpha_0)}^{\alpha_1} d(t - v)\frac{v - \alpha_0}{\alpha_1 - \alpha_0}\,dv - \int_{\max(0,\alpha_0)}^{\alpha_1} d(t - v)dv\right],$$

$$\frac{dD(t)}{d\alpha_1} = \frac{-1}{\alpha_1 - \alpha_0}\int_{\max(0,\alpha_0)}^{\alpha_1} d(t - v)\frac{v - \alpha_0}{\alpha_1 - \alpha_0}\,dv - \frac{\log(2)}{\alpha_2}\int_{\alpha_1}^{t} d(t - v)\exp\{-(v - \alpha_1)\log(2)/\alpha_2\}\,dv,$$

and

$$\frac{dD(t)}{d\alpha_2} = \frac{\log(2)}{\alpha_2^2}\int_{\alpha_1}^{t} d(t - v)(v - \alpha_1)\exp\{-(v - \alpha_1)\log(2)/\alpha_2\}\,dv.$$

As with the bilinear model, the needed integrals are computed using the exponential decay weighted exposure code with small alterations.

The covariance matrix for the four parameters is then inverse of the information matrix. For the three parameter model, with α_0 set equal to zero, the covariance matrix is the inverse of the β, α_1, α_2 components of the information matrix. Epicure scripts for computing these covariance matrices are available from the authors.