Original article
Scand J Work Environ Health 1983;9(4):353-366

doi:10.5271/sjweh.2401

Statistical methods for analyzing effects of temporal patterns of exposure on cancer risks.
by Thomas DC

This article in PubMed: www.ncbi.nlm.nih.gov/pubmed/6635614
Statistical methods for analyzing effects of temporal patterns of exposure on cancer risks

by Duncan C Thomas, PhD

THOMAS DC. Statistical methods for analyzing effects of temporal patterns of exposure on cancer risks. Scand j work environ health 9 (1983) 353–366. Two approaches to the inclusion of temporal factors in the analysis of exposure-response relations are compared: (i) multivariate logistic analysis incorporating simple exposure indices, together with additional temporal variables and their interactions, and (ii) direct fitting to models of the disease process. These approaches are illustrated on case-referent data relating lung cancer to asbestos and smoking. From approach i average age at exposure, age at risk, and duration of exposure were found to contribute direct or interaction effects. Two models of carcinogenesis were fit. In a two-stage initiation-latency model the median latency period was estimated at 11.0 years overall, but was somewhat longer for asbestos-induced and shorter for smoking-induced tumors. When this difference was allowed for, latency periods seemed to be inversely related to the average intensities of both asbestos and smoking. In the multistage model, the best fit was obtained with asbestos acting at the fourth and smoking at the fifth of six stages. Lifetime risks of lung cancer are projected under the different models.

Key terms: carcinogenesis, latency, temporal effects, statistical methods.

Both temporal effects and interactions are known to play important roles in cancer epidemiology. For all cancers there is a lag between exposure and diagnosis, a lag which is usually long and highly variable for most solid tumors (1). In addition age at first exposure, time since exposure ceased, duration of exposure, and age at risk (the age at which incidence or mortality rates are computed) have all been shown to modify exposure effects in various situations (4, 5, 8, 11, 12, 13, 19, 20).

Unfortunately the effects of temporal factors are often ignored in analyses of epidemiologic data. Though investigators sometimes go to considerable lengths to obtain detailed histories of exposure to the agent(s) under study, as well as to any potentially important confounding or modifying variables, this detailed information is often reduced in the analysis to one or two simple indices of exposure, such as duration, average intensity, or total lifetime exposure. In some cases multivariate analyses are carried out to see if temporal factors show confounding or modifying effects, but again such analyses are usually restricted to simple temporal variables, such as age at first exposure or duration of exposure, and the statistical models used are rarely related to any theoretical concept of the underlying carcinogenic process.

This communication describes how standard multivariate methods for risk analysis and more sophisticated methods for fitting stochastic models of carcinogenesis can be used to incorporate the effects of temporal variables in the analysis.
of epidemiologic data. After some general definitions are given, a set of case-referent data on lung cancer is described in relation to chrysotile asbestos exposure and smoking habits, and these data are used to illustrate various methods of analysis.

The next section describes multivariate analyses of simple exposure and temporal indices and their interactions with the use of a variant of logistic regression. Then the data are fitted to two particular models of carcinogenesis: (i) a two-stage initiation-latency model which addresses questions about the latency period and its relation to exposure and (ii) the multistage model which addresses the question of whether exposure factors act at early or late stages in the carcinogenic process. Finally, the various fitted models are compared in terms of projections of the lifetime risk of lung cancer and expected life shortening resulting from a particular exposure history.

Definitions

The term “temporal factors” includes all those potential determinants or modifiers of disease risks which vary as a person ages. Though exposure itself might vary with age, the term “exposure” is confined in the present communication to some overall summary of past exposure history, while all other aspects of the variation in exposure over time which might modify the effect of this summary variable are grouped under the heading of “temporal factors.” Those which have been considered the most frequently include age at exposure, duration of exposure, age at risk, and the interval from first exposure to onset or death (often loosely called “latency period”). For instantaneous exposures most of these terms present no particular definitional problems, but when exposure occurs over an extended period of time with varying intensity, their definitions become less obvious. These problems are best illustrated by a consideration of “latency periods.”

In common parlance, the term “latency period” refers to the distribution of intervals from first exposure to onset among those developing the disease of interest within the study period (henceforth called “epidemiologic latency period.” In mortality studies the term is often applied to the interval as measured up to death rather than to onset, and this terminology is followed in the present communication. A number of variants of this definition have been discussed recently by Peto (12). Any of these definitions of the epidemiologic latency period suffer from the following difficulties. First, because it is measured from first exposure, it overestimates the interval from the etiologically most important increment of exposure. Second, because deaths from other causes, losses, and termination of the study will preclude the cases with longer latency periods from being observed, the epidemiologic latency period will underestimate the lifetime distribution of latency periods, unless a suitable method of analysis for censored survival times is used. Finally, the epidemiologic latency period distribution may include many cases which were not attributable in any way to the exposure factor under study, and for these the interval is meaningless.

One way to avoid these difficulties is to define a “biological latency period” as the interval from each increment of exposure to later increases in risk which are attributable to them (8, 14, 17, 19). Even for an individual this is clearly not a single number but a probability distribution. For instantaneous exposures this distribution is estimable from computations of excess risk curves (e.g., the relative or attributable risks comparing the exposed against the nonexposed or the excess risks per unit of exposure) as functions of time since exposure. Thus the shape of the distribution of the biological latency period will depend on whether a relative or attributable risk scale is used. For extended exposures, whether constant or varying with intensity \(d(t)\) at age \(t\), the distribution of the biological latency period can only be estimated if some functional form \(f(h)\) is assumed and its parameters are fit to the excess relative or absolute risks \(ER(T)\) at age \(T\), eg, with the use of the linear model

\[
ER(T) = b \int_0^T f(t) dt - d(t) dt. \tag{eq 1}
\]

In other words \(f(h)\) describes the expected contribution of a unit of exposure at time \(t\) to the risk at time \(t + h\), the function \(f\) being assumed (at this stage) to be independent of \(t\) and \(d(t)\). Rothman (14) has
proposed taking $f(t)$ to be constant over some “window” of time and identically zero outside it, the optimal window being chosen by trial and error. On the other hand, in fitting data on lung cancer among the Colorado plateau uranium miners, Lundin et al. (8) assumed a lognormal distribution for $f(t)$ and estimated its median and dispersion by trial and error. In a similar way Brown & Chu (3) derived the distribution $f(t)$ from the multistage theory of carcinogenesis and compared the fits obtained by assuming that arsenic acted at either the first or penultimate stages. The section on stochastic models of carcinogenesis explores extensions of these approaches, which allow the parameters of any particular distribution of the biological latency period to be estimated by maximum likelihood.

Once $f(t)$ is estimated, “age at exposure” for extended exposures can be defined naturally. The problem with using ages at first or last exposure is that neither may be very closely related to the effective exposure. The “etiologically effective age at exposure” can be defined simply as an average of ages at exposure $t$, each weighted by their expected contributions to the risk at age $T$, namely $d(t)f(T-t)$. In a similar way one might define an “effective duration” of exposure as the standard deviation of ages at exposure, again weighted by the contribution of each exposure to the risk at age $T$. For an exposure to a constant intensity with a uniform distribution of the biological latency period the effective duration would be proportional to the total duration of exposure, but, otherwise, periods of relatively low exposure and periods outside the peak of the biological latency period distribution would contribute less to the effective duration.

Example — The Quebec asbestos mortality study

The data used for the following illustration come from a historical cohort study of 10,939 males, comprising all who had ever worked for at least a month in the Quebec chrysotile asbestos mining and milling industry and who had been born 1891 through 1920 (7, 10). Employers’ payroll records were combined with estimates and measurements of past dust concentrations to produce year-by-year exposure estimates for each man. [Recent efforts to convert these data to fibers (9) were not used in the presented analyses because the accuracy of these conversions varies with time and thus could have confounded the effects of temporal factors; nevertheless, as the quality of the original particle measurements may also have varied with time, some confounding no doubt remains.] Smoking information was obtained by questionnaire from most of those subjects (or their next of kin) who had survived beyond 1950. This information is limited to the age at starting and stopping smoking and the average number of cigarettes per day while a smoker. Other questions on pipe and cigar smoking and the maximum cigarette consumption were not found to be helpful. (The completeness and quality of these data may also have varied with time.)

By the end of 1975 about 44 % of the cohort had died, 245 of them from lung cancer, the only cause of death considered in the present communication. Each of these lung cancer cases was then matched with three referents drawn at random from those in the cohort who were born in the same year, had outlived the case, and had worked in the same mining area (Thetford Mines or Asbestos). In all the analyses, the subjects for whom no smoking information was available were excluded (together with their matched referents if the subject was a case). Exposures of the cases were compared with those of their matched referents as evaluated at the same point in time (7), here the death of the case, and exposure-response models were fitted with the use of a matched technique of analysis which allows for multiple referents per case, continuous multiple exposure factors, and arbitrary risk models (16). Different indices of exposure were used in the different sections of the paper, the goal being to explore various ways of incorporating temporal effects.

Multivariate methods

This section describes analyses in which the relative risk of lung cancer is assumed to be a linear combination of simple indices of cumulative exposure to asbestos and smoking, the interaction between
them, and interactions of each with simple temporal variables. A linear model was used because the exposure-response relationships for both asbestos and smoking were much more nearly linear than exponential in shape (16) though this choice imposes certain practical limitations, described later.

To provide a basis of comparison with the more complex models discussed, only the simplest indices of exposure were entered in the first step, namely, total asbestos exposure and total smoking. Both are accumulations of intensity without regard (at this stage) for latency periods or any other temporal variables. In other words, \( f(h) \) in equation 1 was assumed to be constant so that risk is linearly related to simple indices of the form \( D(T) = \int_0^T d(t) \, dt \), where \( d(t) \) is the intensity (of smoking or asbestos) at age \( t \) and \( T \) is the age at the point of comparison of cases and their matched referents, ie, the age at death of the case. The results of incorporating these variables and their product in the linear model are described in table 1. The best fit was intermediate between additive and multiplicative; though much closer to multiplicative, the additive model was not significantly rejected.

The following four temporal factors, defined in the appendix, were considered next for potential direct or interaction effects: age at risk – \( T \), the age at death of the case; average age at exposure, the exposure-weighted mean age (for smoking, there is no information on the variation in \( d(t) \) over time, so this reduces simply to the midpoint of the smoking interval); average interval from exposure to risk, the difference between the previous two indices; and duration of exposure – the standard deviation of the intensity-weighted ages at exposure (for smoking, this is proportional to the simple duration of smoking).

Table 2 gives the score test chi squares and signs of association for the additional contributions of each of the aforementioned variables ["direct effects," \( E_i(T) \) in the appendix] and their products with their corresponding simple exposure indices ["interaction effects," \( D_j(T) \)]. Because the cases and referents had been age-matched no direct effect of age at risk could be assessed, but there was a significant interaction of age at risk with asbestos (score test \( \chi^2 = 7.67 \)). This effect is illustrated in table 3, in which the effect of asbestos increases smoothly up to the age group 60–64 and thereafter declines sharply. After adjustment for smoking, the estimate of the slope in the oldest age group was negative, though not significantly different from zero. For smoking a similar pattern was apparent, with a peak in the age group 70–74, but neither the linear nor quadratic trends were significant. These results are consistent with the action of a latency period, as explored further in the section on stochastic models of carcinogenesis. There was also a weak trend (score test \( \chi^2 = 3.17 \)) for longer asbestos exposures to have a greater effect than shorter exposures for the same total dose. This phenomenon may be simply a reflection of the fact that exposure intensities were estimated with much greater probable errors than exposure durations, so that the latter would tend to receive greater weight in statistical fitting.

For smoking, age at exposure and duration of exposure showed positive direct and interaction effects. The direct effect of duration of smoking was so strong (score test \( \chi^2 = 12.46 \)) that, when this variable was added to the equation, the contribution of total smoking became nonsignificant. When duration of smoking, rather than total smoking, was used as the primary exposure variable, the only smoking variables which made significant additional contributions were the direct effects of the average interval from exposure to risk (negative, score test \( \chi^2 = 4.14 \)) and the intensity of smoking if a smoker nine years before the age at risk (positive, score test \( \chi^2 = 3.95 \)). Duration of smoking also showed less departure from additivity than either total smoking or smoking intensity, and its exposure-response relation showed marked convexity [ie, excess risks per unit of exposure increased with increasing exposure, likelihood ratio (LR) \( \chi^2 = 7.00 \)], whereas the other variables showed concavity (LR \( \chi^2 = 4.38 \) and 4.93 for total smoking and lagged intensity, respectively). These findings are also consistent with ages at smoking being measured more accurately than intensities (13).

Stepwise model building, starting with the variables in table 2, could in principle
allow the independent effects of these various temporal factors to be isolated. In practice, however, such analyses have often proved difficult to interpret (7). In the present case the first variable to be entered would be the direct effect of duration of smoking, which reflects more on the inappropriateness of the choice of primary smoking index than on a true modifying effect of duration. An additional complication arises with the use of a linear relative risk model, because some effects can lead to negative risk estimates. For example the age-asbestos interaction is so strong that negative risk estimates are obtained for some of the oldest, heavily exposed but nonsmoking subjects. This problem does not arise with certain other risk models, such as the exponential (16), but additional terms might need to be added to allow for the curvilinearity of the dose-response model. Table 4 shows that, after the inclusion of total asbestos, total smoking, their squares, and their product in the exponential model, the only interaction which remained significant was the modification of total smoking by average age at smoking \( (\text{LR } \chi^2 = 6.26) \). After the inclusion of this variable, the interactions of total asbestos with age at risk and age at risk squared were still apparent, but nonsignificant \( (\text{LR } \chi^2 = 2.29) \). When duration of smoking was used as the simple index, the interaction with average age at smoking was again present \( (\text{LR } \chi^2 = 7.39) \), but the interaction of age at risk with asbestos, which was initially apparent, disappeared after adjustment for the smoking interaction.

Table 1. Fits of simple indices of cumulative asbestos and smoking to linear additive and multiplicative relative risk models for lung cancer.

| Model      | Asbestos | Smoking | Mixture | Chi square | Degrees of freedom |
|------------|----------|---------|---------|------------|--------------------|
| Additive   | 0.00369  | 0.00372 | 0.0c    | 62.15      | 2                  |
| Multiplicative | 0.00179  | 0.00254 | 1.0c   | 64.51      | 2                  |
| Mixture    | 0.00208  | 0.00277 | 0.710   | 64.54      | 3                  |

\( ^a \) Fitted model: \( \text{RR} = 1 + aA + bS + abcAS \).
\( ^b \) Relative to the global null hypothesis \( a=b=0 \).
\( ^c \) Constrained parameter value.

Table 2. Score test chi squares (one degree of freedom) and signs of association for effects (+ = positive, - negative) of adding temporal variables to the linear-mixture model of table 1.

| Temporal factor          | Dust effects | Smoking effects |
|--------------------------|--------------|----------------|
|                          | Direct       | Interaction    | Direct       | Interaction    |
| Age at risk              | - \(^a\)     | 7.67 (-)       | - \(^a\)     | 2.29 (+)       |
| Average age at exposure  | 0.04 (+)     | 0.71 (-)       | 7.80 (+)     | 8.02 (+)       |
| Average interval from exposure to risk | 0.04 (-) | 1.86 (-) | 1.22 (+) | 1.10 (-) |
| Duration of exposure     | 0.92 (+)     | 3.17 (+)       | 12.46 (+)    | 3.49 (+)       |

\( ^a \) No direct effect of age at risk can be assessed as cases and referents were matched on this factor.

Table 3. Interaction of asbestos and smoking effects with age at risk.

| Age at risk group | Number of cases | Slope of asbestos effect | Slope of smoking effect |
|------------------|----------------|--------------------------|-------------------------|
|                  |                | Crude                    | Adjusted for smoking    | Crude               | Adjusted for asbestos |
| 38–54            | 32             | 0.00106                  | 0.00431                 | 0.00163             | 0.00243               |
| 55–59            | 36             | 0.00274                  | 0.00049                 | 0.00082             | 0.00133               |
| 60–64            | 42             | 0.00386                  | 0.01232                 | 0.00153             | 0.00168               |
| 65–69            | 52             | 0.00028                  | 0.00062                 | 0.00372             | 0.00793               |
| 70–74            | 37             | 0.00114                  | 0.00054                 | 0.00916             | 0.01341               |
| 75–85            | 24             | 0.00004                  | -0.00068                | 0.00363             | 0.00805               |
Table 4. Time interactions included in stepwise fitting to the exponential model.

| Variables included | Number of time interactions included for primary smoking variable used | Total smoking | Duration of smoking |
|--------------------|------------------------------------------------------------------------|---------------|---------------------|
|                    |                                                                        | 0             | 1                   | 3                   | 0             | 1                   |
| Dust               | + 0.121 x 10^-2                                                       | + 0.115 x 10^-2 | - 0.233 x 10^-2     | + 0.124 x 10^-2     | + 0.124 x 10^-2 |
| Dust^2             | - 0.161 x 10^-6                                                       | - 0.140 x 10^-6 | - 0.141 x 10^-6     | - 0.127 x 10^-6     | - 0.997 x 10^-7 |
| Smoking            | + 0.176 x 10^-2                                                       | - 0.140 x 10^-6 | - 0.557 x 10^-3     | + 0.210 x 10^-1     | - 0.507 x 10^-1 |
| Smoking^2          | - 0.147 x 10^-6                                                       | + 0.460 x 10^-6 | + 0.460 x 10^-6     | + 0.274 x 10^-6     | + 0.142 x 10^-3 |
| Dust x smoking     | - 0.135 x 10^-6                                                       | - 0.150 x 10^-6 | - 0.202 x 10^-6     | - 0.595 x 10^-5     | - 0.787 x 10^-5 |
| Smoking x average age at exposure | ... | + 0.597 x 10^-4 | + 0.587 x 10^-4 | ... | + 0.195 x 10^-2 |
| Dust x age at risk | ... | ... | + 0.145 x 10^-3 | ... | ... |
| Dust x (age at risk)^2 | ... | ... | - 0.139 x 10^-5 | ... | ... |

Likelihood ratio chi square*: Fitted model 66.80 (5) 73.06 (6) 75.35 (8) 75.76 (5) 83.15 (6)
Additional contributions 6.26 (1) 2.29 (2) 7.39 (1)

* Degrees of freedom in parentheses.

Table 5. Parameter estimates and likelihood ratio tests for the two-stage model.

| Dependence of relative initiation rates on current levels | Biological latency period (BLP) | Likelihood ratio |
|----------------------------------------------------------|--------------------------------|------------------|
|                                                          | Intercept for initiated tumors | Dependence of BLP on average intensity | |
| Asbestos (a)                                             | Asbestos (exp [a]) | Smoking (exp [b]) | Asbestos (exp [c]) | Smoking (exp [d]) | Chi square | Degrees of freedom |
| Smoking (az)                                             | 0.311 | 0.207 | 11.0 | ... | ... | 67.93 | 3 |
|                                                         | 0.279 | 0.257 | 13.6 | 1.005 | 0.965 | 71.55 | 5 |
|                                                         | 0.247 | 0.206 | 16.3 | 6.9 | ... | 69.57 | 4 |
|                                                         | 0.309 | 0.199 | 70.0 | 18.6 | 0.988 | 0.969 | 76.67 | 6 |

Stochastic models of carcinogenesis

This section illustrates how stochastic models of carcinogenesis can be used to derive indices of “effective” exposure for fitting exposure-response relations. The objective is to illustrate the general approach to data analysis rather than to distinguish between alternative theories of carcinogenesis, so the discussion is confined to two simple but rather differently formulated models. However, the basic approach could be applied to other models.

Two-stage initiation-latency model

The simplest model of carcinogenesis which incorporates a latency period assumes that tumors are “initiated” at times in response to the current tissue doses of carcinogens and that each initiated cell remains “latent” for some interval having a distribution . Then it can be shown (2, 18) that the total risk at age can be written as a linear combination of “effective exposure indices” , which are weighted accumulations of past exposures in the form of equation 1, ie,

\[ R(T) = \sum \alpha_j D_j(T) \]  

(eq 2)

where

\[ D_j(T) = \int_0^T f_j(t) f(T-t | b_j) dt \]

and \( \alpha_j \) and \( b_j \) are parameters to be estimated. Armenian & Lilienfeld (1) have reviewed many epidemiologic studies of cancer which have reported latency periods and have concluded that most can be well described by a lognormal distribution, with parameters median \( \mu \) and dispersion factor \( S \) (the antilog of the logarithmic standard deviation 0). This is a special case of a general two-stage model considered by Moolgavkar & Knudson (11), in which the rate of initiation events depends

358
on the rate of division of normal stem cells and the latency period depends on the rate of the proliferation of intermediate stage cells.

In the fitting of these data, the exposure intensities were taken to be the current levels of exposure to asbestos, smoking, and "background" causes, the last being assumed to be constant over time. (In preliminary analyses, not reported in this communication, the product of asbestos and smoking was included as a fourth carcinogen but was found to contribute little beyond the simple additive model; to avoid overparameterization, this term has therefore been omitted from subsequent analyses.) The latency periods for tumors initiated at time $t$ by carcinogen $j$ were taken to be lognormally distributed with median $\mu_j(t)$. This median might be either the same or separate for each carcinogen and might be either constant over time or depend on summary indices $P_k(t)$ representing variables which potentially alter the latency periods, eg, "promoters." The dispersion factor $S$ was arbitrarily fixed to the value 2, independent of exposure and identical for all carcinogens (see the appendix).

The results of fitting the model are shown in table 5. When a common distribution was assumed for the biological latency period of all carcinogens, the estimated median was 11.0 years. However, the addition of average intensities of asbestos and smoking over their respective periods of exposure as the "promoters" $P_k(t)$ seemed to indicate that asbestos lengthened the latency period, while smoking shortened it. This observation suggested that perhaps it was the assumption of a common latency period that was at fault. When separate biological latency periods were allowed for and the background biological latency period was assumed to be short, the median for asbestos-induced tumors was estimated at 16.3 years and that for smoking-induced tumors at 6.9 years (Wald's $\chi^2 = 2.15$ comparing the two medians). This finding is consistent with the results of simple comparisons of epidemiologic latency periods along the lines described in the section Multivariate Methods; for asbestos the mean interval between "effective" exposure and risk was 31.4 years, whereas for smoking it was 18.6 years. Now both asbestos and smoking showed shortening effects on latency period ($LR \chi^2 = 7.10$).

However this test of exposure-biological latency period relationships seems to be sensitive to a number of extraneous effects. For example, if the latency period is shortened by an increase in exposure, this effect will tend to induce a concave tendency for the overall exposure-response relationship (6, 17). Furthermore, if long durations increase the risk per unit of total exposure, then high exposures will tend to be associated with long biological latency periods. Finally, if total exposure has no shortening effect but exposure after initiation does, then exposure before initiation will appear to have a lengthening effect, and conversely. Thus an observed exposure-biological latency period relationship could reflect any of these phenomena (or perhaps others) and should therefore not be overinterpreted. In an effort to understand these findings better, a variety of other variables were considered (eg, cumulative versus current exposures as initiators, average versus cumulative exposures, before and after initiation, as promoters), as well as additional parameters (eg, retention, nonlinearity, dispersion factor, and interaction terms), but the interpretation of the "promoting" effect in the last line of table 5 remains somewhat uncertain.

An additional problem with the two-stage model is its prediction that the increase in cancer rates with age should begin to taper off beyond the median latency period. As the cohort age-specific incidence rates for lung cancer and most solid tumors continue to rise with age at an ever accelerating rate, this prediction would imply that the median must be longer than the normal lifespan. This problem could be reconciled with the relatively short latency periods estimated in this instance and in other occupational cohorts only if (i) the latency period were substantially shortened by an increase in exposure or (ii) the rate of initiation events or the distribution of the biological latency period depended on age. Just such an age dependence follows naturally from the multistage model (2, 18, 20) described in the following section or the general two-stage model of Moolgavkar & Kundson (11).

A final difficulty with the two-stage
model is that, because of its highly non-linear form, iterative fitting can require rather complex programming, and problems of convergence can arise. No problems were encountered in estimating the slopes or medians, but for some choices of promoting variables, convergence problems did arise, for example, in the last line of table 5 when the smoking variable was added – the value reported, corresponding to the highest likelihood obtained by Newton-Raphson iteration, may therefore not in fact be the global maximum likelihood estimate.

The Armitage-Doll multistage model

In the multistage model, it is assumed that at least one cell must pass through \( k \) distinct, heritable changes in a particular sequence to form a cancerous clone, and that the probabilities of any of these transitions might be affected by exposure to carcinogens (2). It can then be shown (20) that, if a single carcinogen acts at a single stage \( i \) of a \( k \)-stage process, then the model can be represented as a simple variant of equation 1, in which \( f(T-t) \) is replaced by

\[
\int_a(T,t) = \int_a(t) - (T-t)^k - t^k - 1,
\]

(eq 3)

which depends on both \( t \) and \( T \) and not just on \( T-t \) as in the previous model. This form includes a number of general shapes, from continuously increasing (if \( i = 1 \)) to continuously decreasing (if \( i = k-1 \)), a number of skewed distributions having different peaks for intermediate stage carcinogens (18). For two carcinogens \( d_i(t) \) and \( d_j(t) \) acting at stages \( i \) and \( j \), respectively, the corresponding generalization (18) is

\[
RR(T) = 1 + a_1D_1(T) + a_2D_2(T) + a_3P(T) + a_4C(T),
\]

(eq 4)

where \( D_i(T) \), \( P(T) \), and \( C(T) \), are single and double integrals of the \( d_i(t) \), as defined in the appendix. Thus each carcinogen makes an independent contribution to the risk \( [D_1(T) \text{ and } D_2(T)] \), and in combination the two can also make a joint contribution – if they act at different stages, the term \( P(T) \) is approximately equal to \( D_1(T)D_2(T) \), depending on the temporal relation between the two exposures (18), so the model is basically multiplicative in form; if they act at the same stage, \( P(T) \) is zero and the model is basically additive in form, unless there is a "cocarcinogenic" effect, represented by \( C(T) \). Fitting this model is much simpler than fitting the two-stage model. It suffices to evaluate the summary indices \( D_i(T), P(T), \) and \( C(T) \) for a variety of choices of the discrete parameters \( i, j, \) and \( k \), and then to estimate the continuous parameters \( a_1, a_2, \) and \( a_3 \) with the use of packaged programs for fitting general relative risk models (16).

The results of fitting this model are described in table 6. Like the parameter \( b_0 \) in the two-stage model, the parameter \( k \) could not be estimated directly because it primarily affects the dependence of rates on age at risk (2), and in these data the cases

| Stage at which carcinogens act | Asbestos | Smoking | Parameter estimates | Likelihood ratio |
|-------------------------------|----------|---------|---------------------|------------------|
| Asbestos Smoking              | Parameter estimates | Chi square | Degrees of freedom |
| 3 3                           | 0.1516   | 0.1843  | 0.00161             | 59.54            | 2 |
| 3 4                           | 0.0824   | 0.1458  | 0.1278              | 65.93            | 2 |
| 3 5                           | 0.0704   | 0.1298  | 0.1417              | 68.09            | 2 |
| 4 3                           | 0.0773   | 0.1217  | 0.1546              | 60.61            | 2 |
| 4 4                           | 0.1850   | 0.1676  | 0.1546              | 64.76            | 2 |
| 4 5                           | 0.1161   | 0.1546  | 0.00155             | 68.66            | 3 |
| 5 3                           | 0.0798   | 0.1298  | 0.1740              | 55.79            | 2 |
| 5 4                           | 0.1324   | 0.1231  | 0.1740              | 60.41            | 2 |
| 5 5                           | 0.1949   | 0.1740  | 0.00221             | 64.76            | 2 |
| 5 5                           | 0.2376   | 0.1740  |                     | 65.81            | 3 |
and referents had been age-matched. It was therefore arbitrarily fixed at \( k = 6 \), a value which provides a good fit to the population age-specific incidence curves for most solid tumors (2, 13). Preliminary analyses (17) indicated that asbestos and smoking both appeared to be relatively late-stage carcinogens, so, for the analyses of the two in combination, \( i \) and \( j \) were restricted to 3, 4 and 5. (The value \( i, j = 6 \) could be excluded because it would imply no lag between exposure and increase in risk.) When each carcinogen was considered separately, asbestos appeared to fit best at the fourth and smoking at the fifth of six stages, in support of the relatively short latency periods (somewhat longer for asbestos than for smoking) found in the two-stage model; indeed the weight functions \( f(T, t) \) in equation 3 are very close in shape to the lognormal distribution (18). [In a similar analysis, Brown & Chu (3) found that arsenic appeared to act at the penultimate rather than the first stage.] In combination, the fits with asbestos and smoking acting at the same and at different stages were not substantially different, an occurrence consistent with the pattern of interaction intermediate between additive and multiplicative apparent in table 1. The addition of the "cocarcinogenic" term \( C(T) \) made only a slight contribution (LR \( \chi^2 \) = 1.78 at stage 4 and 1.03 at stage 5). Again somewhat better fits could be obtained with different definitions of the exposure variables \( d(t) \) (eg, cumulative rather than current asbestos exposure, square root rather than linear smoking intensities, etc), but the overall patterns remained unchanged.

These results on smoking are somewhat inconsistent with the theory that smoking must act at both an early and a late stage of the carcinogenic process, which is supported by the increased risk in those starting earlier and the rapid decline in risk following cessation (5, 13). The model was therefore assuming that either smoking intensity or the fact of being a smoker acted at two stages. For all possible combinations of pairs of stages and smoking variables, the sign of the earlier stage dependence on smoking was negative, so this model does not appear to fit the data. This inconsistency may occur in part because the two-stage action would require a much more convex dose-response relation than seen for either smoking duration or total smoking (the latter being concave, as already noted) or simply because much more detailed data on changes in smoking intensities would be needed to distinguish various stages of action. As there is relatively little variability in ages at the start of smoking, the present analysis is probably influenced mainly by how recently the exsmokers stopped smoking, and this small group may be highly selected by factors related to health status.

Predictions of lifetime risk

Environmental standards are often set in terms of predictions of lifetime cancer risk or expected life shortening. In order to illustrate how different models of the effects of temporal variables lead to quite different projections of lifetime risk, standard life-table calculations (19) were carried out. These calculations combined the age-specific death rates for lung cancer and all other causes for Canadian males in 1978 with the relative risk models for asbestos and smoking effects already described. For illustrative purposes the rates for other causes were assumed to be independent of both smoking and asbestos, and the population lung cancer rates were multiplied by constants chosen for each model in such a way that, if 60% of the male population were nonsmokers and 40% smoked an average of 20 cigarettes per day, the predicted number of lung cancers in the combined, nonasbestos-exposed groups would equal the number expected in the Canadian life-table population. The numbers of excess lung cancer deaths are shown in table 7, and the age distributions of the excess deaths are shown in fig 1 for selected models. To summarize these distributions, table 7 also gives the loss of life expectancy per excess lung cancer death; the loss of life expectancy in the group as a whole can be calculated by multiplying this value by the lifetime excess probability of lung cancer.

If the two exponential models are compared using simple exposure indices, the addition of the temporal modifiers predicts fewer lung cancer deaths due to asbestos in nonsmokers (but at earlier ages) and more deaths due to smoking (but at...
Table 7. Life-table projections for selected exposure-response models.

| Model | Excess lung cancer deaths per 100,000 at risk | Loss of life expectancy (years) per excess lung cancer death |
|-------|---------------------------------------------|----------------------------------------------------------|
|       | Smokers : nonsmokers | Exposed : nonexposed | Smokers : nonsmokers | Exposed : nonexposed |
|       | Nonexposed | Nonsmokers | Smokers | Nonexposed | Nonsmokers | Smokers |
| Multivariate models with simple exposure indices |
| 1. Linear-additive (table 1, line 1) | 7,846 | 378 | 261 | 12.4 | 12.9 | 12.5 |
| 2. Linear-multiplicative (table 1, line 2) | 6,669 | 183 | 598 | 12.3 | 12.9 | 12.4 |
| 3. Exponential-no modifiers (table 4, column 1) | 6,892 | 120 | 365 | 12.2 | 12.9 | 12.2 |
| 4. Exponential-three modifiers (table 4, column 3) | 7,946 | 77 | 181 | 10.2 | 15.3 | 10.4 |
| Two-stage models with separate median latency periods |
| 5. Constant median (table 5, line 3) | 7,856 | 291 | 253 | 13.7 | 15.8 | 13.8 |
| 6. Dose-dependent median (table 5, line 4) | 7,873 | 180 | 153 | 13.6 | 11.6 | 13.6 |
| Multistage model with smoking acting at the fifth stage \(k = 6\) |
| 7. Asbestos stage 4 (table 6, line 8) | 6,565 | 171 | 555 | 13.6 | 14.2 | 13.7 |
| 8. Asbestos stage 5 (table 6, line 11) | 9,642 | 493 | 429 | 13.7 | 16.4 | 13.8 |

* 20 cigarettes per day from age 15 to death.
* 1 million particles per cubic foot per year from age 20 to age 60.

Discussion

These results demonstrate the feasibility of using carcinogenic models to construct meaningful indices of cumulative exposure for the analysis of epidemiologic studies. Biological mechanisms cannot be inferred simply by comparing the fits of several statistical models (15). Nevertheless, statistical models, such as those discussed in the present communication, do provide useful ways of exploring different kinds of hypotheses about temporal effects and interactions. Consideration of a variety of models allows one to produce a range of estimates of projected lifetime risk for setting hygienic standards.

Thus the choice of a model essentially boils down to the nature of the hypothesis one wishes to test. Many can be simply cast in the form of a combination of a basic exposure index and one or more products with modifying factors, as was done in the section on multivariate analysis. This approach may be adequate if one is content merely to test for associations, but is less satisfactory for estimation, particularly if one wishes to use such estimates for extrapolation, say for life-table projection of risk estimates to an entire lifetime. For such purposes, the statistical model ought to be grounded on some theoretical notions about the underlying disease process.

In this paper, two such statistical representations of the carcinogenic process were considered. As the fit of the two was fairly similar, a choice between them can-
Fig. 1. Projected age distributions of excess lung cancers resulting from occupational exposure to 1 million asbestos particles per cubic foot per year from age 20 to age 64, under various risk models.
not be made on purely statistical grounds. Indeed, though the two models appear to be formulated quite differently, both indicated that recent exposure was the strongest determinant of lung cancer risks in these data and both indicated that the interaction between the initiation rates was intermediate between additive and multiplicative. Thus the choice between them again reduces to the nature of the questions being asked. If the interest is primarily in the distribution of latency periods and their relationship to exposure, then the two-stage model is particularly useful because latency periods enter the model as explicit terms. On the other hand, if the interest is in identifying the role of a carcinogen as an "initiator," "promoter," "cocarcinogen," etc, then the multistage model is more attractive because these concepts can, to some extent, be associated with the stage at which the carcinogens act. Other hypotheses might better be cast in terms of different models of carcinogenesis, and future work will consider a wider range of models, such as the general two-stage model of Mooivandkar & Knudson (11).

There is, of course, a risk of overparameterizing such models. For example, the convergence problems encountered with the two-stage model when the effects of shortening the latency period were added may be a reflection of such overparameterization. The limitations of the data must always be borne in mind when inferences are drawn about each additional parameter, and any effects would need careful replication in other data. Thus the purpose of these approaches is not to allow strong inferences about biological mechanisms but to provide biologically plausible methods to allow temporal factors to be taken into account, where they are known to play a strong role.

The present case-referent data are somewhat limited in their usefulness in this context because two of the strongest variables for discriminating between models, age and calendar time, have been used for matching. On the other hand, variability in exposure histories does allow some discrimination between early and late effects at much lower cost than a full cohort analysis. As these data arose from a cohort study (5), future work will attempt to incorporate the known sampling fractions in order to take advantage of age and secular effects, while maintaining the economy of case-referent analyses.

Temporal effects can have a public health importance far beyond any light they might shed on underlying mechanisms. When predicting the impact of an intervention, such as a control limit for an "acceptable" level of exposure, one is generally extrapolating not only to low doses but also in time beyond the period of follow-up. It is not uncommon for uncertainties in temporal extrapolation to be much more important than uncertainties in low-dose extrapolation (4, 19). Thus, in the same way that ranges of risk estimates based on different exposure-response models provide the policymaker with an idea of the reliability of extrapolation, so ranges of estimates based on different models of the role of temporal factors can be quite important. Since the kinds of models discussed in the present communication can only be fitted by those with access to raw data, it is important for epidemiologists to consider a variety of temporal effects and ways of characterizing them when reporting their findings.

Acknowledgments

The author is grateful to Dr C McDonald for permission to use the data for the illustrated applications and to Mr R Mensah for help with the computer programming.

This research was supported by the National Cancer Institute of Canada.

References

1. Armenian HK, Lilienfeld A. The distribution of incubation periods of neoplastic diseases. Am J Epidemiol 99 (1974) 92–100.
2. Armitage P, Doll R. Stochastic models of carcinogenesis. In: Neyman J, ed. Proceedings of the 4th Berkeley symposium on mathematical statistics and probability. Volume 4 University of California Press, Berkeley, CA 1961, pp 19–38.
3. Brown CC, Chu KC. Approaches to epidemiologic analysis of prospective and retrospective studies: Example of lung cancer and exposure to arsenic. In Prentice RL, Whittemore AS, ed. Environmental epidemiology: Risk assessment. Society for Industrial and Applied Mathematics, Philadelphia, PA 1982, pp 94–106.
4. Committee on the Biological Effects of Ionizing Radiation. The effects on popula-
ations of exposure to low levels of ionizing radiation. National Academy of Sciences, Washington, DC 1980.

5. Day NE, Brown CC. Multistage models and primary prevention of cancer. J natl cancer inst 64 (1980) 977–989.

6. Enterline PE. Epidemiological methods for the establishment of dose-response relationships in asbestos-dust-related diseases. In: National Institute for Metallurgy, Proceedings asbestos symposium, Johannesburg 1977, Ranburg, South Africa, pp 39–42.

7. Liddell FDK, McDonald JC, Thomas DC. Methods of cohort analysis: Appraisal by application to asbestos mining (with discussion). J r stat soc A 140 (1977) 60–95.

8. Lundin FE, Archer VE, Wagoner JK. An exposure-time-response model for lung cancer mortality in uranium miners: Effects of radiation exposure, age, and cigarette smoking. In: Breslow NE, Whittemore AS, ed. Energy and health. Society for Industrial and Applied Mathematics, Philadelphia, PA 1978, pp 243–264.

9. McDonald JC, Gibbs GW, Liddell FDK. Chrysotile fibre concentration and lung cancer mortality: A preliminary report. In: Wagner JC, ed. Biological effects of mineral fibres. Volume 2. International Agency for Research on Cancer, Lyon 1980, pp 811–817. (IARC scientific publication no 30).

10. McDonald JC, Liddell FDK, Gibbs GW, Eysen GE, McDonald AD. Dust exposure and mortality in chrysotile mining, 1910–1975. Br j ind med 37 (1980) 11–24.

11. Moolgavkar SH, Knudson AG Jr. Mutation and cancer: A model for human carcinogenesis. J natl cancer inst 66 (1981) 1037–1052.

12. Peto J. Some problems in dose-response estimation in cancer epidemiology. In: Voug V, ed. Quantitative risk estimation. SGOMSEC, Rome (in press).

13. Peto R. Epidemiology, multistage models, and short-term mutagenicity tests. In: Hiatt HH, Watson JD, Winster JA, ed. Origins of human cancer. Volume 4. Cold Springs Harbor Laboratory, New York, NY 1977, pp 1403–1428.

14. Rothman K. Induction and latent periods. Am j epidemiol 114 (1981) 253–259.

15. Siemiatycki J, Thomas DC. Biological models and statistical interactions: An example from multistage carcinogenesis. Int j epidemiol 10 (1981) 383–387.

16. Thomas DC. General relative-risk models for survival time and matched case-control analysis. Biometrics 37 (1981) 673–686.

17. Thomas DC. Are dose-response, synergy, and latency confounded? Presented at Joint Statistical Meetings, Detroit, MI, August 1981.

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

19. Thomas DC, McNeill KG. Risk estimates for the health effects of alpha radiation. Atomic Energy Control Board, Ottawa 1982.

20. Whittemore AS. The age distribution of human cancer for carcinogenic exposures of varying intensity. Am j epidemiol 106 (1977) 418–432.

Appendix

Mathematical formulation of the exposure indices

All of the models described in the body of the paper can be written in the form

$$RR(T) = r(a, D(T, d(t), t \leq T); b),$$  \hspace{1cm} (eq A)

where $RR(T)$ denotes the risk at age $T$ relative to that associated with no exposure, $d(t)$ is the observed exposure history, $D(T)$ is a set of summary exposure indices, which may depend on parameters $b$ to be estimated, and $r(a, D)$ describes the overall exposure-response relationship, with parameters $a$. The summary indices $D(T)$ can all be expressed as time-weighted accumulations of the past exposure intensities of the form

$$D(T) = \int_0^T f(T, t; b) d(t) dt,$$  \hspace{1cm} (eq B)

where the form of the weight functions $f(T,t)$ varies from model to model or even from index to index.

For the simple indices described in the section on multivariate methods, the weight functions for the interactions effects are

$$f_0(T, t) = 1 \quad \text{simple indices}$$

$$f_1(T, t) = T \quad \text{age at risk}$$
$f_2(T,t) = t$ \hspace{1cm} \text{age at exposure}

$f_3(T,t) = T-t$ \hspace{1cm} \text{interval from exposure to risk}

$f_4(T,t) = (t-\bar{\tau}_{et})^2$ \hspace{1cm} \text{duration of exposure;}

and the indices for the direct effects are $E_i(T) = D_i(T) / D_0(T)$. Thus the "etiologically age at exposure" $\bar{\tau}_{et}$ is simply $E_g(T)$. The indices used for testing duration effects were transformed from variance to standard deviation scales, i.e., $E_4(T) = \sqrt{E_4(T)}$ and $D_4(T) = E_4(T) D_0(T)$. The relative risk functions $r(b, D)$ were assumed to be either linear or exponential, as described in the text.

For the two-stage model absolute risks were taken to depend linearly on exposure indices of the form $E_B$ for asbestos, smoking, and "background causes" $D_0(T) = 1$. The $f_j(T, t)$ were taken to be the lognormal distribution

$$f(T,t) = \frac{1}{(2\pi)^{1/2} \sigma} \exp \left\{ - \frac{[\ln(T-t) - \ln\mu_j(t)]^2}{2\sigma^2} \right\}$$

(eq C)

where $\sigma$ is the logarithmic standard deviation and $\mu_j(t)$ is the median associated with carcinogen $j$. In some analyses $\mu_j(t)$ was assumed to be a constant over time, either common for all carcinogens or separate. In other analyses the $\mu_j(t)$ was allowed to depend on summary exposure indices $P_j(t)$; for this purpose, an exponential dependence

$$\mu_j(t) = \exp \left[ b_j + \sum_k \xi_k P_k(t) \right]$$

(eq D)

was adopted to avoid negative medians and improve convergence. The dispersion factor $S$ was assumed to be a constant, independent of exposure and identical for all carcinogens. In the fitting of the model, it developed that there was not enough information in the data to allow the parameter to be estimated, so it was arbitrarily fixed at $S = 2.0$. Different choices were considered but did not substantially alter the findings. The relative risk could then be written in the form

$$RR(T) = 1 + \sum \rho_j D_j(T) / D_0(T).$$

(eq E)

Unless the median biological latency period for background causes, $\exp(b_0)$, is large (say, comparable to the ages at death of the cases), the denominator in equation $E$ will be essentially equal to unity and can be omitted; in any event, its parameter $b_0$ can only be reliably estimated from case-referent data if a common median is assumed for all carcinogens.

In the multistage model, the weight function for the relative risk due to a single carcinogen acting at a single stage $i$ of a $k$-stage process would be

$$f_{ik}(T,t) = \frac{t^{i-1} (T-t)^{k-i-1}}{T^{k-1}}.$$  

(eq F)

For two carcinogens acting at stages $i$ and $j$, the relative risk is

$$RR(T) = 1 + b_1 D_1(T) + b_2 D_2(T) + b_1 b_2 P(T) + b_3 C(T),$$  

(eq G)

where

$$D_1(T) = \int_0^T d_1(t) f_{ik}(T,t) dt,$$

$$D_2(T) = \int_0^T d_2(t) f_{ik}(T,t) dt,$$

$$P(T) = \int_0^T d_1(t) d_2(t) f_{ik}(T,t) dt,$$

$$C(T) = \int_0^T d_1(t) d_2(t) f_{ik}(T,t) dt.$$  

For a carcinogen acting at two stages, the relative risk is of the same form as equation $G$ except that $C(T) = 0$ and $d_3(t) = d_2(t)$.

Received for publication: 29 November 1982