OEM Online First, published on November 11, 2010 as 10.1136/oem.2009.053983

The two-stage clonal expansion model in occupational cancer epidemiology: results from three cohort studies

Ariana Zeka,[1,2] Rebecca Gore,[1] David Kriebel[1]

[1]Department of Work Environment, University of Massachusetts, Lowell, Massachusetts, USA
[2]Institute for the Environment, Brunel University, Uxbridge, UK

Correspondence to
Dr Ariana Zeka, Institute for the Environment, Brunel University, Halsbury 149, Uxbridge UB8 3PH, UK;
ariana.zeka@brunel.ac.uk

Accepted 17 October 2010

ABSTRACT

Objectives The objective of this work was to apply the two-stage clonal expansion model, with the intention to expand the literature on epidemiological applications of the model and demonstrate the feasibility of incorporating biologically based modelling methods into the widely used retrospective cohort study.

Methods The authors fitted the two-stage clonal expansion model model to three occupational cohort studies: (1) a cohort of textile workers exposed to asbestos and followed for lung cancer mortality; (2) a cohort of diatomaceous earth workers exposed to silica and also followed for lung cancer mortality; and (3) a cohort of automotive manufacturing workers exposed to straight metalworking fluid (MWF) and followed for larynx cancer incidence. The model allowed the authors to estimate exposure effects in three stages: cancer initiation (early effects), promotion or malignant transformation (late effects).

Results In the first cohort, the authors found strong evidence for an early effect of asbestos on lung cancer risk. Findings from analyses of the second cohort suggested early and less evidently late effects of silica on lung cancer risk. In the MWF (third) cohort, there was only weak evidence of straight MWF exposure effects on both early and late stages. The authors also observed a late birth cohort effect on larynx cancer risk.

Conclusions The findings for asbestos and silica were essentially confirmatory, supporting evidence for their early effects on lung cancer from a large body of literature. The effect of straight MWF on larynx cancer was less clear.

What this paper adds

- The most successful biologically based models in epidemiology, the multistage and two-stage carcinogenesis models first proposed by Armitage and Doll and Moolgavkar and Knudson respectively, have been applied to only a limited number of studies with individual quantitative exposure data.
- The present study was an epidemiological application of the two-stage clonal expansion (TSCE) Moolgavkar and Knudson model in three occupational cohorts.
- The findings strengthened previous studies on the relationships between lung cancer and asbestos and silica exposures by providing maximum likelihood evidence on the stage of action of these carcinogens. This application also demonstrates the feasibility of incorporating biologically based modelling methods into the widely used cohort study to provide evidence on the stage of action of occupational and environmental carcinogens.

INTRODUCTION

Perhaps the most successful biologically based models in epidemiology are the multistage and two-stage carcinogenesis models first proposed by Armitage and Doll,[1] and Moolgavkar and Knudson,[2 3] respectively. The latter model, and subsequent variants, have been applied to both animal and human cancer data,[3–14] but in only a limited number of cases has such a model been applied in studies containing quantitative exposure data on individuals.[8 10 13–15] The two-stage model allows one to compare the strength and goodness of fit of the association between exposure and cancer risk, alternately assuming that the exposure acts at an early or a late stage in carcinogenesis. Stage information is potentially useful for understanding human risk, and for making predictions about the time trends of environmental cancers.

The two-stage clonal expansion (TSCE) model (also known as the Moolgavkar or Moolgavkar–Knudson two-stage model of carcinogenesis) formalises carcinogenesis as a three-phase process. In the first phase (initiation), a susceptible stem cell undergoes one or more events which transform this normal cell into an intermediate stage. This may be one or more mutations, but epigenetic changes could also be involved with no loss of relevance of the model form. In a second phase, the initiated cell may undergo clonal expansion, also called promotion. If there is clonal expansion of intermediate cells, then the probability is increased that, in a final phase (progression), one of the clones of initiated cells will undergo an additional genetic change leading to full malignant transformation and subsequently to clinically detectable cancer. This final event is called malignant conversion. The two rare events (initiation and malignant conversion) are the 'stages' of the TSCE, but the growth of a clone of intermediate cells is also recognised as a critical phase during which environmental chemicals and cancer-prevention interventions can act; hence, the choice of the name TSCE rather than simply a two-stage model.

There is good experimental and clinical evidence that many human cancers pass through these three phases, and there is also a small but growing body of epidemiological evidence consistent with the

hypothesis that environmental carcinogens act on one or more of these phases.[8 10 13 14 16 17] The TSCE model is the simplest stochastic mathematical model of the initiation, promotion and malignant conversion paradigm of carcinogenesis.[18]

Despite its theoretical and mathematical development, the TSCE has not been applied to more than a handful of occupational cohort studies. Our objective in this paper was to apply the method to three existing datasets in order to expand the literature on epidemiological applications of the model. This research project had the long-term goal of aiding cancer-prevention efforts by demonstrating the feasibility of incorporating biologically based modelling methods into the widely used retrospective cohort study.

The TSCE was used to study exposure–cancer risk associations in: (1) a cohort of textile workers exposed to asbestos and followed for lung cancer mortality; (2) a cohort of diatomaceous earth workers exposed to silica and also followed for lung cancer mortality; and (3) a cohort of automotive manufacturing workers exposed to straight metalworking fluid (MWF) and followed for larynx cancer incidence. Quantitative individual lifetime exposure data were available for all three cohorts, so that the model could make use of both inter- and intraindividual variations in exposure intensity. Asbestos and silica are well-recognised lung carcinogens, and the particular datasets employed here have been previously analysed by standard methods and the results published by others.[19–21] In contrast, the association between straight MWF and larynx cancer has only been reported in a single occupational cohort,[22] and it was these same data that we have reanalysed in this paper.

METHODS

Form of the TSCE model

The two-stage clonal expansion model can be schematised as shown in figure 1. In this model, four time-dependent parameters describe the stochastic process of carcinogenesis, starting from the pool of normal cells to the occurrence of the first malignant cell, and are: $\alpha_1(s)$ the first mutation rate leading to the creation of an initiated or premalignant cell, b(s) and d(s) the birth and death rates of premalignant cells, and $\alpha_2(s)$ the second mutation rate leading to the creation of a malignant cell.

The model assumes that a number of normal susceptible cells are initiated at time s, therefore becoming premalignant, by a non-homogenous Poisson process with intensity $\alpha_1(s) \times X(s)$; where X(s) is the expected number of normal cells (assuming X is a fixed number), and $\alpha_1(s)$ is the rate of initiation (or the first mutation rate) at time s. At any given time (s), the premalignant (initiated) cells can divide into two premalignant cells with birth rate b(s), die with death rate d(s) or divide asymmetrically into one initiated cell and one malignant cell with rate $\alpha_2(s)$.[18] More detailed descriptions of this model are given in earlier work by Heidenreich and colleagues[18 23] and in more recent applications of the model.[24 25]

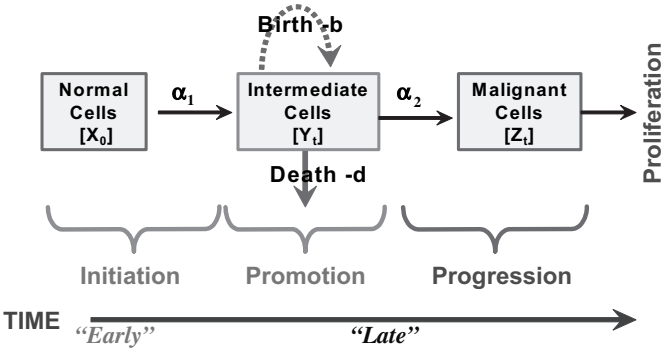

Figure 1 Two-stage clonal expansion model.

According to this model, exposure to a carcinogen could act in one of several ways: by increasing the transition or mutation rate from normal to intermediate cells (increasing α_1); increasing the transition rate from intermediate to malignant cells (increasing α_2); or increasing the proliferation of intermediate cells by altering either the birth (b) or death rate (d) of these cells. By studying the time course of exposure and its relationship to risk in an exposed population, it may be possible to determine which of these modes of action is occurring. Also, different carcinogenic exposures that occur in the same environment may act via different pathways, and these pathways may be distinguishable mathematically. In this model, a tumour initiator is postulated to act by increasing α_1, an early acting exposure agent, while a promoter should act by increasing the proliferation of intermediate stage cells through altering the balance of (b–d) so as to increase the pool of intermediate cells ready to receive the second 'hit' (α_2), or a late acting exposure.[4 14]

Datasets

Asbestos and lung cancer

The cohort includes all workers employed in asbestos textile operations in a factory in South Carolina for at least 1 month between 1 January 1940 and 31 December 1965, and followed for vital status through 31 December 1990 (table 1).[20] There were a total of 116 000 person years of observation and 1259 deaths. Of these, there were a total of 124 deaths from lung cancer. Final models (males only) included 74 lung cancer deaths. A detailed exposure reconstruction allowed Dement and colleagues to assign annual exposure estimates to each worker.[20]

Several authors have published analyses of these data, and all show strong associations between cumulative asbestos exposure and lung cancer risk.[19 20] Models with continuous exposure variables showed risk of lung cancer primarily associated with exposures occurring 20–24 years prior diagnosis (RR=4.6, 95% CI 1.3 to 16.3), and less evidently for exposures in the 15–20 years (RR=1.4; 95% CI 0.4 to 4.8).[26]

Silica and lung cancer

The Diatomaceous Earth (DE) cohort included 2342 white males (23% Hispanic) who were employed for at least 12 months including at least 1 day between 1 January 1942 and 31 December 1987 in the diatomaceous earth mining and processing industry.[21 27] Vital status was determined for 91% of the cohort, and cause of death was ascertained for 716 of 749 (96%) of identified deaths. Cumulative exposure estimates to respirable dust and respirable crystalline silica have been computed based on historical reconstruction of exposures for all subjects.[27] In the early study of this cohort, excess mortality based on 77 lung cancer deaths was found in association with cumulative crystalline silica equal or greater than 5 mg/m^3-years (RR=2.11, 95% CI 1.07 to 4.11, no exposure lag).[21]

Metalworking fluids and larynx cancer

Three automotive parts manufacturing facilities in Michigan were studied.[28] All hourly employees who had worked at least 3 years prior to 1 January 1985 were eligible for inclusion in the study. Hire dates ranged from 1917 to 1981, and a total of 46 384 employees met the study definition, of whom approximately 10% were female, and 21% African–American. By the end of initial follow-up in 1984, 10 159 (22%) had died. The cause of

Table 1 Description of datasets of the three occupational cohorts

	Asbestos—lung cancer	Silica—lung cancer	'Straight' (petroleum-based) metalworking fluid—larynx cancer
References	Dement *et al*[20]	Checkoway *et al*[21]; Seixas *et al*[27]	Zeka *et al*[22]
Definition of cohort	1247 white males employed between 1940 and 1965	2342 white males employed at least 1 year between 1942 and 1987	46 000 workers employed at least 3 years between 1938 and 1985
Period of follow-up	1940 to 1990	1942 to 1994	1941 to 1994; analyses only for males
Case definition	Mortality	Mortality	Incidence
No of cases	74	77	78
Mean (range) age of death/ diagnosis (years) of cases	64.0 (46.5 to 88.6)	63.4 (44.0 to 79.0)	63.2 (39.2 to 81.2)

death was ascertained for 92% of these subjects, based on death certificates. Past exposures to MWF and certain components were estimated based on air sampling measurements, review of plant records and interviews with personnel.[29]

This analysis included 78 incident cases of laryngeal cancer in the cohort, and 3093 subcohort members, males only.[22] A positive association was found for cumulative exposure to straight MWF more than 10 years before risk age and larynx cancer risk (RR=1.08, 95% CI 1.03 to 1.14, for each 5 mg/m^3-years).[22]

Fitting the TSCE model

To implement this model, we used software developed by Heidenreich, Luebeck and Moolgavkar.[18] Heidenreich and colleagues, and Hanin and Yakovlev[30] have found that there is a fundamental identifiability problem in the TSCE model which prohibits the estimation of the values of the fundamental model parameters directly from incidence or mortality data alone: the first and second mutation rates (α_1 and α_2); and the birth and death rates of intermediate cells (b and d). Hence, the parameter combinations A, B and C are estimated instead as functions of the four fundamental parameters, and also used in our version of the software.

The two sets of parameters can be related by the following approximations:

$$-A \cong b - d \quad B \cong \alpha_2 \quad C = \alpha_1 X/b$$

where b, d, α_1, α_2 and X were defined above. These approximations indicate that A is the negative of the net proliferation rate of intermediate cells. Parameter B is approximately $\alpha_2/(1-d/b)$, which may also be referred to an 'effective' malignant transformation rate, assuming d is negligible (if there is non-extinction of the intermediate cell clone). C is a function of the first mutation rate (α_1), birth rate of intermediate cells (b) and size of the normal cell pool (X). Because the latter is assumed to be large and essentially unchanging in size (the levels of exposure to carcinogens in a modern workplace are unlikely to be so high as to be directly cytotoxic, and the first mutation rates are likely to be so small that they will not deplete the pool of normal cells to any appreciable degree), C can be seen as a ratio of the first mutation rate to the birth rate of intermediate cells. That is, it represents the balance of these two pathways for the growth of the intermediate cell pool.

Fitting the model to epidemiological data

Exposure was treated as a continuous lifetime cumulative exposure variable and allowed to modify the baseline parameters of the model. The model parameters are time-dependent and were estimated as a function of exposure changing overtime (annually). The model was fitted in terms of parameters A, B and C, as noted above. Following the work of Heidenreich and colleagues,[18] we systematically evaluated five functional forms by which exposure was allowed to modify one or more of the three parameters. The following exposure functions were used (P represents one of the three parameters A, B or C):

$$\begin{aligned}
\text{Linear}: \quad & P = P1 \times (1 + P2 \times E) \\
\text{Power}: \quad & P = P1 \times \left(1 + P2 \times E^{P3}\right) \\
\text{Log}: \quad & P = P1 \times (1 + \ln(1 + P2 \times E)) \\
\text{Exponential}: \quad & P = P1 \times e^{P2 \times E} \\
\text{Negative exponential}: \quad & P = P1 \times e^{-P2 \times E}
\end{aligned}$$

where P1 is a baseline parameter, and P2 and P3 are additional parameters representing the unit change in P as a function of exposure (E). We fitted the model in the following way. First, the model was fitted to the cancer data in the cohort with no provision for exposure affecting any of the parameters of the model (baseline model). This yielded baseline estimates of the three unknown parameters. Then, cumulative exposure (E) effects on each of the parameters were allowed, investigating the five functional forms.

We also allowed year of birth (birth cohort, or BC) to affect model parameters. Moolgavkar and colleagues have found that for lung cancer, there is a strong birth cohort effect, which they interpret as primarily an effect of changing smoking prevalence over the second half of the 20th century. We subtracted the earliest birth year in the cohort from all others, so that the term was always positive. We investigated linear and exponential effects of birth cohort on all three parameters. Finally, we combined exposure and birth cohort effects in the same model.

Maximum likelihood methods are used to find the most likely values of the parameters. The fits of alternative models can be compared using the −2 log likelihood statistics (−2LL). Differences in −2LL are expected to have a χ^2 distribution. When two competing models are nested one within the other, one can test the null hypothesis of no improvement in fit by the larger model compared with the smaller using the difference in −2LL, with degrees of freedom equal to the number of additional parameters in the larger model. One can also examine 95% CIs for the estimated coefficients, to evaluate the strength of the evidence for an exposure (or birth cohort) effect on a model parameter.

We did not perform formal hypothesis testing, as it was not appropriate for this methods development research. However, p values do provide a way to judge of what is a 'large' improvement in goodness of fit for an exposure model, compared with a baseline model. We used the following criteria for the weight of evidence indicated by a reduction in −2LL (the table below assumes one degree of freedom difference between the exposure and baseline models):

Weight of evidence	Reduction in −2 LL	Associated p value
Weak	3.8–2.1	0.05–0.15
Modest	6.6–3.8	0.01–0.05
Strong	>6.6	<0.01

Table 2 Results of fitting a two-stage clonal expansion model to the study of lung cancer and exposure to asbestos (Dement *et al*[20])

Model	Parameter A	Parameter B	Parameter C	−2log likelihood
Baseline	−0.24 (−0.15 to −0.32)*	0.17e−6 (0.34e−8 to 0.82e−5)	0.041 (0.020 to 0.083)	937.1
Birth cohort affecting†				
A	$-0.12\times(1+BC^{0.14e-8})$	0.17e−6	0.41e−1	936.8
B	−	−	−	§
C	−0.24	0.17e−6	$0.19e-1\times(1+BC^{0.33e-1})$	936.9
Exposure affecting‡				
A	−	−	−	§
B	−	−	−	§
C	−0.24	0.18e−6	$0.018\times(1+\ln(1+1.6*E))$	928.1

Parameter $-A\cong b-d$ is the negative of the net proliferation rate of intermediate cells. Parameter $B\cong\alpha_2$ is approximately the malignant transformation rate. Parameter $C=\alpha_1\times X/b$ is the rate of growth of the intermediate cell pool.
*95% CIs shown for baseline model. See text for 95% intervals for final models.
†Birth cohort (BC) affecting one parameter at one time, while the other two parameters remain unchanged.
‡Exposure (E) affecting one parameter at one time, while the other two parameters remain unchanged.
§Poor fit—models did not converge.

RESULTS

Overview

A large number of different models were fitted to the data from each of the three cohorts. Not all models converged, and this was interpreted as meaning that the model that was being estimated did not fit the observed data well. In the detailed results presented below, only the best-fitting parameterisation for the effect of exposure, birth cohort or the combined effects (exposure and birth cohort) on one of the three parameters A, B or C is shown.

Asbestos and lung cancer

The baseline model produced reasonably precise estimates of parameters A and C with 95% CIs of less than an order of magnitude (table 2). Parameter B, equivalent to the second mutation rate (α_2), was estimated to be very small—about 10^{-6}, and imprecisely estimated—95% CI of about three orders of magnitude. The −2LL for the baseline model was 937.1. Models in which one of the parameters was allowed to vary with birth cohort either did not converge (parameter B) or had only very slightly better fits than the baseline model (parameters A and C).

When models were fitted in which exposure was allowed to modify one of the three parameters, convergence was achieved only when exposure was modifying C. All five parameterisations of the exposure effect on C fitted better than the baseline model. The log exposure model had the smallest values of −2LL, 928.1. With one degree of freedom difference from the baseline model, this reduction in −2LL of 9.0 was substantial.

Models in which exposure was allowed to modify either A, the net proliferation rate of intermediate cells, or B, the second mutation rate fitted poorly.

The best-fitting model in which exposure modified parameter C had this form:

$$C=0.018\times(1+\ln(1+1.6\times E))$$

The 95% CI for the first coefficient of the above equation was (0.0059 to 0.058), and for the second coefficient it was (0.065 to 39).

Silica and lung cancer

The baseline model, in which neither exposure nor birth cohort was allowed to influence model parameters was qualitatively similar to the baseline model for the asbestos data (table 3). Again, parameters A and C were estimated with fair precision, as evidenced by their 95% CIs, while parameter B was smaller and less precisely estimated. The baseline model had a goodness of fit (−2LL) of 1048.0. As in the asbestos data, adding birth cohort alone to the baseline model did not improve the fit. For this reason, we did not attempt to fit models combining exposure and birth cohort effects.

Models were fitted in which exposure was allowed to affect parameters A or C. In none of these models was the improvement in fit compared with the baseline model as impressive as it was in the asbestos dataset. The best fitting model was the linear form of an exposure effect on parameter C, with −2LL value of 1045.9 (table 3). This model had a reduction in −2LL of about 2 and can be expressed as follows:

Table 3 Results of fitting a two-stage clonal expansion model to a study of lung cancer and exposure to crystalline silica (Checkoway *et al*[21])

Model	Parameter A	Parameter B	Parameter C	−2log likelihood
Baseline	−0.26 (−0.37 to −0.14)*	0.78e−7 (0.41e−9 to 0.15e−4)	0.20e−1 (0.94e−2 to 0.44e−1)	1048.0
Birth cohort affecting:†				
A	$-0.26\times(1+0.43e-8\times BC)$	0.79e−7	0.20e−1	1048.0
B	−0.26	$0.79e-7\times e^{0.41e-7\times BC}$	0.20e−1	1048.0
C	−0.26	0.79e−7	$0.20e-1\times e^{0.50e-10\times BC}$	1048.0
Exposure affecting:‡				
A	$-0.25\times(1+(1-e^{-0.37\times E}))$	0.11e−6	0.22e−1	1046.8
B	−0.26	$0.78e-7\times(1+0.95e-9\times E)$	0.20e−1	1048.0
C	−0.26	0.78e−7	$0.19e-1\times(1+3.4\times E)$	1045.9

Parameter $-A\cong b-d$ is the negative of the net proliferation rate of intermediate cells. Parameter $B\cong\alpha_2$ is approximately the malignant transformation rate. Parameter $C=\alpha_1\times X/b$ is the rate of growth of the intermediate cell pool.
*95% CIs shown for baseline model. See text for 95% intervals for final models.
†Birth cohort (BC) affecting one parameter at one time, while the other two parameters remain unchanged.
‡Exposure (E) affecting one parameter at one time, while the other two parameters remain unchanged.

$$C = 0.019 \times (1 + 3.40 \times E)$$

The 95% CIs of the two coefficients were: (0.009 to 0.039) and (0.50 to 25), respectively.

Models in which exposure modified parameter A fitted slightly less well. Similar to the case of asbestos, the results for silica suggest that exposure may act at an early stage in carcinogenesis, however, with also a suggestion for a late stage effect (promotion).

Metalworking fluids and larynx cancer

The baseline model, in which neither exposure nor birth cohort was allowed to influence model parameters, was again qualitatively similar to those for asbestos and silica. Parameters A and C were estimated with fair precision, while parameter B was much smaller and less precisely estimated.

Unlike in the other two datasets, we found strong evidence for a birth cohort effect in these data (table 4). Large improvements in the −2LL were consistently observed when birth cohort was allowed to affect parameter A. The best fit was observed for the simple power function of birth cohort affecting A. A represents the net proliferation rate of intermediate cells, and the birth cohort effect we found could be interpreted to indicate a late stage effect. Birth cohort effects were also found on parameters B and C, although the model fits were not as good as those for models allowing effects on parameter A.

For investigation of exposure effects, we defined a new 'baseline' model which included an effect of birth cohort on parameter A, and compared all exposure models to this one. Models in which exposure was allowed to modify any of the three parameters showed only very modest improvements in fit over the baseline model. No evidence of an effect on parameter B was observed. Nearly identical fits were found with models that allowed an exposure effect on A or C, but the largest reduction in −2LL was only about 2. The best fitting model had a linear effect of exposure on parameter A:

$$A = -0.049 \times (1 + BC^{0.47}) \times (1 + 0.079E)$$

The 95% CIs for these three coefficients were: (−0.068 to −0.029), (0.40 to 0.54) and (0.016 to 0.40) respectively.

Because of the observation of both birth cohort and exposure effects, we investigated a variety of different model forms including both of these effects. We allowed birth cohort to affect A and C simultaneously with the best functional forms (results not shown), while exposure remained on either A or C at its best functional form. Also, exposure was allowed to affect A and C simultaneously, while birth cohort remained on A in its best functional form. Either there were no improvements in the model fit, or the fits deteriorated. We also examined if there was collinearity of birth cohort and exposure, when these two variables were used together in the models. Results suggested no important collinearity between birth cohort and exposure.

DISCUSSION

Biologically based cancer models

Occupational and environmental epidemiology and associated risk assessment research rely heavily on quantitative models to accurately assess the strength of exposure–disease associations. Most quantitative exposure–response models are largely empirical in structure, incorporating only a limited number of generic assumptions, such as the multiplicative relation between covariates included in common RR regression models. These models—logistic and Poisson regression, and the Cox model, for example—are flexible and seem to fit a variety of data sets reasonably well. Unfortunately, no direct validation is ever really possible, and so doubts remain about the accuracy of risk predictions from these models.[31–33] An additional challenge to the application of these models is that when the results of an incorrectly specified model are applied to a population with different distributions of exposure and other covariates than those in the study population, the risk predictions may be biased. This will be true, even if the model fits the data adequately. This creates another reason for caution in the interpretation of the results of standard empirical modelling methods.

Biologically based models may help to improve the accuracy of exposure-risk estimation in several ways.[2 34–37] A biologically based epidemiological model derives its structure, and possibly some of its parameters, from experimental studies and theoretical models of the pathophysiological mechanisms underlying the disease processes being studied. There are only a limited number of examples of the use of such models, partly because of their complexity, and partly because of the normal skepticism of researchers to adopt unproven methods. However, agreement between a biologically based model and empirical models fitted to the same data lends validity to both. In addition, the biological model tends to have parameters which are

Table 4 Results of fitting a two-stage clonal expansion model to a study of larynx cancer and exposure to straight metalworking fluid (Zeka *et al*[22])

Model	Parameter A	Parameter B	Parameter C	−2log likelihood
Baseline	−0.20 (−0.27 to −0.12)*	0.21e−5 (0.83e−7 to 0.51e−4)	0.15e−1 (0.78e−2 to 0.29e−1)	1156.1
Birth cohort affecting:†				
A	$-0.51e-1\times(1+BC^{0.46})$	0.11e−8	0.16e−1	1094.3
B	−0.40	$0.21e-14\times e^{(0.25\times BC)}$	0.10e−1	1105.2
C	−0.20	0.20e−6	$0.69e-6\times(1+BC^{0.33e+1})$	1103.8
Exposure affecting:‡				
A	$-0.49e-1\times(1+BC^{0.47})\times(1+0.079\times E)$	0.18e−8	0.18e−1	1092.0
B	–	–	–	§
C	$-0.51e-1\times(1+BC^{0.46})$	0.14e−8	$0.16e-1\times e^{(0.72\times E)}$	1092.9

Parameter $-A \cong b-d$ is the negative of the net proliferation rate of intermediate cells. Parameter $B \cong \alpha_2$ is approximately the malignant transformation rate. Parameter $C = \alpha_1 \times X/b$ is the rate of growth of the intermediate cell pool.
*95% CIs shown for baseline model. See text for 95% intervals for final models.
†Birth cohort (BC) affecting one parameter at one time, while the other two parameters remain unchanged.
‡Exposure (E) affecting one parameter at one time, while the other two parameters remain unchanged. BC affecting parameter A simultaneously.
§Poor fit—models did not converge.

interpretable as aspects of physiological processes, lending plausibility to study results.

Experimental evidence now makes it clear that for most types of cancer, there are more than two irreversible steps in the process of complete cell transformation.[38 39] Furthermore, if the number of steps or stages is variable, meaning that there are numerous pathways from an initial carcinogenic exposure to tumour initiation, the very concept of a model based on discrete stages may be incorrect. Despite this complexity, there may be public health utility in the distinction between agents that act early in the carcinogenic process (initiators) and those that act late (promoters). More detailed subdivision of the process, even if it were possible to achieve with epidemiological data, may not have much public health relevance (eg, determining whether an agent acts at the fifth or sixth of eight stages). For these reasons, a two-stage model may be quite useful, even when the true process is much more complex. In particular, late-stage carcinogens (promoters) have been poorly studied in epidemiology, partly because of the strong tendency to assume that lag periods of a decade or more are the rule in cancer epidemiology. Thus, a model that explicitly investigates the early/late distinction could be a useful addition to the epidemiologist's standard repertoire.

Both asbestos and silica appear to act early in the process of lung carcinogenesis. MWF had a less clear effect, but there was some evidence for a later stage of action on larynx cancer. The investigations of asbestos and silica were, in a sense, tests of 'positive controls'—compounds for which we have strong evidence that they act as cancer initiators. Thus, our results are reassuring evidence that the method is capable of finding these expected effects.

Asbestos is perhaps the best-known example of a human cancer initiator. In a previous reanalysis of the same data used here, Pearce calculated cumulative exposure in 5-year time windows, and found that the risk of lung cancer was strongly associated with exposures occurring 20–24 years before lung cancer death, with little or no risk associated with exposures at any other time (table 10.9).[26] In the present analyses, the evidence for an early effect of asbestos was strong, and there was no evidence for a late-stage effect.

Silica has also been found to increase lung cancer risk after a long latency in standard published epidemiological models.[21 40–42] We interpret our findings with the two-stage model to be consistent with this observation of an early effect of exposure on risk. The evidence was not strong, however, and models with a late-stage effect fitted only slightly worse.

Metalworking fluid has been found to increase larynx cancer in a single large cohort study.[22] Our TSCE model application in the same dataset found an important late-stage birth cohort effect. Moolgavkar and colleagues suggested that controlling for birth cohort might adjust for time trends in lifestyle risk factors, such as smoking and drinking. Tobacco smoking is thought to have both initiating and promoting effects on laryngeal cancer, while alcohol probably serves as a promoter.[43–46] The study by Moolgavkar on lung cancer risk among coke oven workers also found that birth cohort affected the net-proliferation rate of intermediate cells, a later effect.[13] In contrast, studies by Luebeck among Colorado uranium miners[14] and of Hazelton in a large cohort of Chinese tin miners[16] both suggested an effect of birth cohort on lung cancer initiation. Thus, the role of birth cohort in the two-stage model probably represents more than simply the time trends in smoking and drinking prevalence.

The weak evidence of MWF exposure effects on both early and late stages might be interpreted as essentially a negative finding. Overall, the more modest fits for MWF and larynx cancer than for either lung carcinogen are consistent with the results of standard analyses.[22]

Smoking information, the main risk factor for both cancer types, lung and laryngeal cancer, was not available for all three cohorts. Lack of information on non-occupational exposures, including smoking, is a common issue in retrospective occupational cohort studies. Because this is an important issue for epidemiological investigation and health policy making, we have previously shown that smoking (or drinking in the case of larynx cancer), even in the more extreme confounding scenarios, is unlikely to increase or decrease the relative risks associated with exposure by more than 20%.[47] The confounding effect, if present, would be of more concern in the case of small to medium relative risks, as may be the case of the association for larynx cancer and MWF. However, such an effect would likely have less impact on strong associations, such as those between asbestos and silica and lung cancer.

CONCLUSIONS

Application of the TSCE model to three occupational cohort studies found evidence that both asbestos and silica appear to act early in the process of lung carcinogenesis. The effect of straight MWF on larynx cancer was less clear, but there was some evidence for a later stage of action. The findings for asbestos and silica are essentially confirmatory, as the evidence for their early effects on lung cancer from a large body of literature is quite strong.

The TSCE model is a different approach to epidemiological modelling from that provided by standard methods like logistic or Poisson regression or the Cox proportional hazards model. It allows formal investigation of competing hypotheses about whether a chemical acts at early or late stages of carcinogenesis. Recent papers by Richardson[48] also show promising alternatives to fitting the TSCE model using SAS software. We hope that these examples will encourage researchers to fit biologically based models in their epidemiological datasets.

Acknowledgements We thank G Luebeck for his great help with initial programming and access to their two-stage cancer model, and also for his insight with the manuscript. We also thank H Checkoway for providing the silica cohort data for this work.

Funding This work was supported by a grant from the US National Institute for Occupational Safety and Health, Centers for Disease Control and Prevention, R01-OH03575, and a grant from the Centers for Disease Control and Prevention/Association of Teachers of Preventive Medicine number TS 0699.

Competing interests None.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. **Armitage P,** Doll R. The age distribution of cancer and a multi-stage theory of carcinogenesis. *Br J Cancer* 1954;**8**:1–12.
2. **Moolgavkar SH,** Knudson AG Jr. Mutation and cancer: a model for human carcinogenesis. *J Natl Cancer Inst* 1981;**66**:1037–52.
3. **Morrison PF.** Effects of time-variant exposure on toxic substance response. *Environ Health Perspect* 1987;**76**:133–9.
4. **Tan WY,** Singh KP. Assessing the effects of metabolism of environmental agents on cancer tumor development by a two-stage model of carcinogenesis. *Environ Health Perspect* 1987;**74**:203–10.
5. **Clewell HJ 3rd,** Andersen ME. Biologically motivated models for chemical risk assessment. *Health Phys* 1989;**57**(Suppl 1):129–37.
6. **Moolgavkar SH.** Cancer models. *Epidemiology* 1990;**1**:419–20.
7. **Moolgavkar SH,** Cross FT, Luebeck G, *et al*. A two-mutation model for radon-induced lung tumors in rats. *Radiat Res* 1990;**121**:28–37.
8. **Moolgavkar SH,** Luebeck EG, Krewski D, *et al*. Radon, cigarette smoke, and lung cancer: a re-analysis of the Colorado Plateau uranium miners' data. *Epidemiology* 1993;**4**:204–17.

9. **Crump KS.** Use of mechanistic models to estimate low-dose cancer risks. *Risk Anal* 1994;**14**:1033—8.
10. **Stayner L,** Smith R, Bailer AJ, *et al*. Modeling epidemiologic studies of occupational cohorts for the quantitative assessment of carcinogenic hazards. *Am J Ind Med* 1995;**27**:155—70.
11. **Portier CJ,** Kopp-Schneider A, Sherman CD. Calculating tumor incidence rates in stochastic models of carcinogenesis. *Math Biosci* 1996;**135**:129—46.
12. **Sherman CD,** Portier CJ. The two-stage model of carcinogenesis: overcoming the nonidentifiability dilemma. *Risk Anal* 1997;**17**:367—74.
13. **Moolgavkar SH,** Luebeck EG, Anderson EL. Estimation of unit risk for coke oven emissions. *Risk Anal* 1998;**18**:813—25.
14. **Luebeck EG,** Heidenreich WF, Hazelton WD, *et al*. Biologically based analysis of the data for the Colorado uranium miners cohort: age, dose and dose-rate effects. *Radiat Res* 1999;**152**:339—51.
15. **Hazelton WD,** Clements MS, Moolgavkar SH. Multistage carcinogenesis and lung cancer mortality in three cohorts. *Cancer Epidemiol Biomarkers Prev* 2005;**14**:1171—81.
16. **Hazelton WD,** Luebeck EG, Heidenreich WF, *et al*. Analysis of a historical cohort of Chinese tin miners with arsenic, radon, cigarette smoke, and pipe smoke exposures using the biologically based two-stage clonal expansion model. *Radiat Res* 2001;**156**:78—94.
17. **Jacob P,** Meckbach R, Sokolnikov M, *et al*. Lung cancer risk of Mayak workers: modelling of carcinogenesis and bystander effect. *Radiat Environ Biophys* 2007;**46**:383—94.
18. **Heidenreich WF,** Luebeck EG, Moolgavkar SH. Some properties of the hazard function of the two-mutation clonal expansion model. *Risk Anal* 1997;**17**:391—9.
19. **Dement JM,** Harris RL Jr, Symons MJ, *et al*. Exposures and mortality among chrysotile asbestos workers. Part II: mortality. *Am J Ind Med* 1983;**4**:421—33.
20. **Dement JM,** Brown DP, Okun A. Follow-up study of chrysotile asbestos textile workers: cohort mortality and case—control analyses. *Am J Ind Med* 1994;**26**:431—47.
21. **Checkoway H,** Heyer NJ, Seixas NS, *et al*. Dose—response associations of silica with nonmalignant respiratory disease and lung cancer mortality in the diatomaceous earth industry. *Am J Epidemiol* 1997;**145**:680—8.
22. **Zeka A,** Eisen EA, Kriebel D, *et al*. Risk of upper aerodigestive tract cancers in a case-cohort study of autoworkers exposed to metalworking fluids. *Occup Environ Med* 2004;**61**:426—31.
23. **Moolgavkar SH,** Luebeck EG, Turim J, *et al*. Quantitative assessment of the risk of lung cancer associated with occupational exposure to refractory ceramic fibers. *Risk Anal* 1999;**19**:599—611.
24. **Crump KS,** Subramaniam RP, Van Landingham CB. A numerical solution to the nonhomogeneous two-stage MVK model of cancer. *Risk Anal* 2005;**25**:921—6.
25. **Meza R,** Hazelton WD, Colditz GA, *et al*. Analysis of lung cancer incidence in the Nurses' Health and the Health Professionals' Follow-Up Studies using a multistage carcinogenesis model. *Cancer Causes Control* 2008;**19**:317—28.
26. **Checkoway H,** Pearce N, Kriebel D. *Research Methods in Occupationa Epidemiology*. Oxford: Oxford University Press, 2004.
27. **Seixas NS,** Heyer NJ, Welp EA, *et al*. Quantification of historical dust exposures in the diatomaceous earth industry. *Ann Occup Hyg* 1997;**41**:591—604.
28. **Eisen EA,** Tolbert PE, Monson RR, *et al*. Mortality studies of machining fluid exposure in the automobile industry I: A standardized mortality ratio analysis. *Am J Ind Med* 1992;**22**:809—24.
29. **Hallock MF,** Smith TJ, Woskie SR, *et al*. Estimation of historical exposures to machining fluids in the automotive industry. *Am J Ind Med* 1994;**26**:621—34.
30. **Hanin L,** Yakovlev A. Identifiability of the joint distribution of age and tumor size at detection in the presence of screening. *Math Biosci* 2007;**208**:644—57.
31. **Louis TA,** Robins J, Dockery DW, *et al*. Explaining discrepancies between longitudinal and cross-sectional models. *J Chronic Dis* 1986;**39**:831—9.
32. **Ware JH,** Dockery DW, Louis TA, *et al*. Longitudinal and cross-sectional estimates of pulmonary function decline in never-smoking adults. *Am J Epidemiol* 1990;**132**:685—700.
33. **Maldonado G,** Greenland S. Impact of model-form selection on the accuracy of rate estimation. *Epidemiology* 1996;**7**:46—54.
34. **Thomas DC.** Pitfalls in the analysis of exposure-time-response relationships. *J Chronic Dis* 1987;**40**(Suppl 2):71S—8S.
35. **Smith TJ.** Occupational exposure and dose over time: limitations of cumulative exposure. *Am J Ind Med* 1992;**21**:35—51.
36. **Kriebel D.** The dosimetric model in occupational and environmental epidemiology. *Occupational Hygiene* 1994;**1**:55—68.
37. **Kriebel D,** Checkoway H, Pearce N. Exposure and dose modelling in occupational epidemiology. *Occup Environ Med* 2007;**64**:492—8.
38. **Weiss KM.** Cancer models and cancer genetics. *Epidemiology* 1990;**1**:486—90.
39. **Boland CR,** Ricciardiello L. How many mutations does it take to make a tumor? *Proc Natl Acad Sci U S A* 1999;**96**:14675—7.
40. **Rice FL,** Park R, Stayner L, *et al*. Crystalline silica exposure and lung cancer mortality in diatomaceous earth industry workers: a quantitative risk assessment. *Occup Environ Med* 2001;**58**:38—45.
41. **Steenland K,** Mannetje A, Boffetta P, *et al*. Pooled exposure—response analyses and risk assessment for lung cancer in 10 cohorts of silica-exposed workers: an IARC multicentre study. *Cancer Causes Control* 2001;**12**:773—84.
42. **Steenland K,** Sanderson W. Lung cancer among industrial sand workers exposed to crystalline silica. *Am J Epidemiol* 2001;**153**:695—703.
43. **International Agency for Research on Cancer (IARC).** Tobacco smoking. In: *IARC Monographs of the Evaluation of the Carcinogenic Risk of Chemicals to Humans*, **Vol 38**. Lyon: WHO, IARC, 1986.
44. **IARC.** *International Agency for Research on Cancer*. Alcohol drinking. In: *IARC Monographs of the Evaluation of the Carcinogenic Risk of Chemicals to Humans*, **Vol 44**. Lyon: WHO, IARC, 1988.
45. **International Agency for Research on Cancer (IARC).** Tobacco smoke and involuntary smoking. In: *IARC Monographs of the Evaluation of the Carcinogenic Risk of Chemicals to Humans*, **Vol 83**. Lyon: WHO, IARC, 2002.
46. **Zeka A,** Gore R, Kriebel D. Effects of alcohol and tobacco on aerodigestive cancer risks: a meta-regression analysis. *Cancer Causes Control* 2003;**14**:897—906.
47. **Kriebel D,** Zeka A, Eisen EA, *et al*. Quantitative evaluation of the effects of uncontrolled confounding by alcohol and tobacco in occupational cancer studies. *Int J Epidemiol* 2004;**33**:1040—5.
48. **Richardson DB.** Multistage modeling of leukemia in benzene workers: a simple approach to fitting the 2-stage clonal expansion model. *Am J Epidemiol* 2009;**169**:78—85.

The two-stage clonal expansion model in occupational cancer epidemiology: results from three cohort studies

Ariana Zeka, Rebecca Gore and David Kriebel

Occup Environ Med published online November 11, 2010
doi: 10.1136/oem.2009.053983