Risk assessment using exposure intensity: an application to vermiculite mining

Pamela M Vacek, J Corbett McDonald

Abstract
Estimation of exposure-response relations from epidemiological data is complicated by the fact that exposures usually vary in intensity over time. Cumulative exposure indices, which do not separate the effects of intensity and duration, are commonly used to circumvent this problem. In this paper the estimation of relative risk for specific ranges of exposure intensity from such data is considered using existing statistical methods for fitting multivariate relative risk models. This has the advantage that it does not assume that exposure intensity and duration have equivalent effects on risk. It also throws light on the possible existence of a threshold. The procedure was applied to data from a cohort of 406 vermiculite miners to examine the lung cancer risk associated with exposure to fibrous tremolite, which contaminated the vermiculite. The pattern of exposure-response differed substantially from that obtained using a cumulative exposure index to assess risk.

Once the health effects of exposure of man to a toxic substance have been shown it is important to quantify risk in terms of exposure so that control limits may be established. In many studies only the duration of exposure is known and no information exists about intensity. Even where intensities have been estimated with some accuracy they usually vary over time. This is certainly true in occupational studies in which workers may hold a variety of jobs entailing differing amounts of exposure and where the industrial process or work environment may change radically.

The usual approach is to compute a cumulative exposure index as the product of duration and intensity summed over all the intensities to which a subject has been exposed. This is attractive for several reasons. Useful methods of analysis, such as those based on the standardised mortality ratio, require that person-years be grouped into categories based on exposure. With more than one or two exposure variables this becomes unworkable. Also, examination of interactions between exposure and other risk factors, smoking for example, is greatly simplified when only one exposure variable is present. Multiple variables lead to multiple interaction terms and to models that may be difficult to fit and interpret.

The use of cumulative exposure, however, has widely recognised limitations. Analyses based on this index implicitly assume that duration and intensity have equivalent effects on risk. Thus for a given cumulative exposure, risk is assumed to be the same regardless as to whether the person was exposed briefly at high intensity or for a longer period at lower intensity. Whereas this assumption may be reasonable in some circumstances, it has little justification in others. Animal and human studies have suggested that the physiological response to some carcinogens may depend on the concentration at which a fixed total exposure is administered.1,2 In such cases, the use of cumulative exposure to estimate the exposure-response relation can be misleading, particularly as results may well be used to estimate risk and set safety standards in terms of intensity, not cumulative exposure.

The existence of statistical methods for fitting multivariate risk models makes it unnecessary to rely only on an overall index such as cumulative exposure to assess risk when intensity data are available. In this paper we describe a general strategy for directly estimating the relative risk associated with specified ranges of exposure intensity using data from subjects who have been exposed at a variety of intensities. Our approach is applied to data from vermiculite miners to assess the exposure-response relation between amphibole fibres in the tremolite series and lung cancer.

Statistical methods
The approach described here is applicable to any cohort for which data reflecting intensity are available over time and the outcome of interest is dichotomous, such as mortality from a specific disease. As a first step we group exposure intensities into categories. The
selection of cutpoints for the categories should take into account the nature of the data as well as statistical considerations. For example, if exposures were recorded as ranges of values for a particular area, it would be desirable to use the same ranges in defining the categories. It is unnecessary for the intensity intervals to be of equal duration and in fact they do not have to be based on quantitative measures. They could equally well be descriptive measures such as job or department within the plant. Although this would not define the exposure-response relation in quantitative terms it could be useful in estimating risk for specific types of work. Although we need enough categories to permit characterisation of the exposure-response relation, this will be restricted by the requirement that they be broad enough to include adequate numbers for meaningful risk estimation. If interruptions in exposure are believed to have a mitigating effect, a zero category should be included to represent these gaps.

Once exposure categories have been selected we define the variates $X_{ik}(u), \ldots, X_{im}(u)$ to be the durations that the $i^{th}$ subject has spent at each of the $m$ categories up to time $u$. These durations may be computed to exclude a lag period representing the minimum time necessary for a disease to develop and cause death. In that case $X_{ik}(u)$ would be the duration up to time $u - L_i$, where $L_i$ is the lag period. Absolute or relative risk can then be modelled as a multivariate function of the $X_{ik}(u)$ as well as any other variables that affect risk or might modify the effects of exposure. The last should be evaluated at either time $u$ or time $u - L_i$ as appropriate.

We chose the following multiplicative model for risk, as it is widely used in epidemiology and is tractable in terms of estimation:

$$\lambda_i(u) = \exp \left\{ \sum_{j=1}^{k} \beta_j X_{ij}(u) \right\} \lambda_i^*(u)$$

Using the notation of Breslow et al., $\lambda_i(u)$ is the probability that the $i^{th}$ subject will die during year $u$, conditional on exposure history and survival to year $u$; $\lambda_i^*(u)$ is the probability of death for an unexposed subject reaching the same age in the same year. The $p$ independent variables include durations at the $m$ exposure levels and any other variables. This model is based on the proportional hazards model of Cox, which assumes that the relative risk (the ratio of $\lambda_i(u)$ to $\lambda_i^*(u)$) is constant over all time points $u$. Exceptions to this can be accommodated by including independent variables that represent the interaction between time and exposure. The regression coefficients ($\beta_j$s) for the exposure variables may be interpreted as the increase in the logarithm of relative risk for each year of exposure at a particular level after adjusting for other risk variables and covariates including time spent working at other intensities.

The regression coefficients can be estimated by fitting the model using either person-years or case-control (referent) methods depending on whether an external or internal comparison is preferred. For an external comparison age, sex, and year specific death rates from an appropriate reference population are used as values for $\lambda_i^*(u)$. Each year $u$ that a member of the cohort is at risk, typically from the time of first exposure until death or end of follow up (or loss from view), contributes one observation to the analysis and values of the $X_{ik}(u)$ must be calculated for each of these years; $\lambda_i(u)$ takes a value of 0 or 1 depending on whether the $i^{th}$ person survived or died during year $u$. The log likelihood of the data can then be numerically maximised to obtain estimates of $\beta_j$.

Berry describes how this can be accomplished with standard statistical software, such as GLIM and GENSTAT, because the log likelihood is identical to that for a Poisson variable.

For an internal comparison, case-control methods are used to fit the model in terms of relative risk, $\lambda_i(u)/\lambda_i^*(u)$. Cases and controls are stratified by year of birth and any other risk factors that are not to be included as variables in the model. Controls can then be selected for each case by identifying cohort members in the same stratum who were still alive at the time of the case's death. All eligible controls or a randomly selected subset may be used. To avoid bias subjects who die of the disease under investigation at a later date should be included as potential controls.

Values of the $X_{ik}(u)$ for both cases and controls correspond to the same year $u$, which can be expressed as either the age or year of death for the case. (Both age and date are equivalent measures for the time variable $u$, because we have matched on year of birth to control for both age and calendar year.) The model can be fitted by iteratively maximising the conditional likelihood as described by Breslow and Day.

**Application to vermiculite miners**

We used both the person-years and case-control approaches described above to analyse data from a cohort of 406 men employed at a vermiculite mine and mill in Montana. All subjects in the cohort were employed before 1963 for at least one year and their mortality was ascertained to July 1983. Our aim was to assess the relation between exposure to an amphibole fibre in the tremolite series, which contaminated the vermiculite ore, and death from lung cancer. For each worker exposure intensities in terms of respirable fibres/ml had been estimated from available air measurements. These data were expressed as an average intensity for each year of employment.

Exposure intensities were grouped into four levels (table 1) and years worked at each level were com-
Risk assessment using exposure intensity: an application to vermiculite mining

Table 1  Summary of exposure data

<table>
<thead>
<tr>
<th>Variable</th>
<th>Range of fibre concentrations (f/ml)</th>
<th>Average fibre concentration (f/ml)</th>
<th>Years spent at each exposure level</th>
<th>Person-years analysis</th>
<th>Case-control analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lung cancer (Total (%))</td>
<td>Alive (Total (%))</td>
</tr>
<tr>
<td>No exposure (gaps)</td>
<td>0-0</td>
<td>0-0</td>
<td></td>
<td>7 (3)</td>
<td>2291 (5)</td>
</tr>
<tr>
<td>Level 1</td>
<td>0-1-4.9</td>
<td>2-1</td>
<td></td>
<td>47 (22)</td>
<td>1806 (42)</td>
</tr>
<tr>
<td>Level 2</td>
<td>5-0-9.9</td>
<td>7-0</td>
<td></td>
<td>65 (31)</td>
<td>12339 (28)</td>
</tr>
<tr>
<td>Level 3</td>
<td>10-0-49.9</td>
<td>24-1</td>
<td></td>
<td>50 (24)</td>
<td>4956 (11)</td>
</tr>
<tr>
<td>Level 4</td>
<td>&gt; 50-0</td>
<td>82-0</td>
<td></td>
<td>43 (20)</td>
<td>5860 (13)</td>
</tr>
<tr>
<td>Average duration of exposure (y)</td>
<td></td>
<td></td>
<td></td>
<td>10-1</td>
<td>5-4</td>
</tr>
<tr>
<td>Average time since first exposure (y)</td>
<td></td>
<td></td>
<td></td>
<td>25-7</td>
<td>21-2</td>
</tr>
</tbody>
</table>

Deaths as well as for the remaining 8003 person-years. Table 1 also presents similar data for the controls and the case-control analysis, that for the cases being the same as in the person-years analysis. Durations are expressed as the number of years spent at a level and as the proportion of all years at that level. The table also shows the mean duration of total exposure and the mean time since first exposure.

Table 2 presents the estimates of the model parameters from the person-years and case-control analyses. As would be expected, results obtained by the two approaches are similar, apart from a substantial difference between the estimated coefficients for level 3. Both procedures yielded negative non-significant coefficients of risk at the lowest exposure level (less than five fibres/ml). In both analyses the risk coefficient for the highest exposure (level 4) is about the same as that for level 2 even though the average fibre concentration for this level was 12 times that for level 2. In neither analysis was the coefficient for years since first exposure significant. The standard errors of all coefficients were smaller in the person-years analysis.

Discussion
As these analyses are based on a small data set, the parameter estimates defining the relation between exposure to tremolite and risk of lung cancer lack precision. Also, information about smoking was not available for the cohort; this would be an important

Table 2  Estimates of regression coefficients (b_j)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Person-years analysis</th>
<th>Case-control analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>b_j</td>
<td>SE b_j</td>
</tr>
<tr>
<td>Years at level 1</td>
<td>-0.0157</td>
<td>0.0556</td>
</tr>
<tr>
<td>Years at level 2</td>
<td>0.1252</td>
<td>0.0622</td>
</tr>
<tr>
<td>Years at level 3</td>
<td>0.1319</td>
<td>0.0602</td>
</tr>
<tr>
<td>Years at level 4</td>
<td>0.1039</td>
<td>0.0575</td>
</tr>
<tr>
<td>Years since first exposure</td>
<td>-0.0234</td>
<td>0.0331</td>
</tr>
<tr>
<td>Constant</td>
<td>1.043</td>
<td>0.7755</td>
</tr>
</tbody>
</table>
variable if cases and non-cases differed substantially in their smoking habits. The results, however, are consistent with a relation in which risk is absent at low concentrations, increases rapidly as concentrations increase, and levels off at high concentrations. This is a biologically feasible relation that was not evident when risk was modelled as a function of cumulative exposure.\textsuperscript{8}

Unless duration and intensity contribute equally, the use of cumulative exposure in risk analysis provides no information about the exposure-response relation in terms of intensity. Despite this, results based on cumulative exposure are commonly used to estimate risk at specific intensities. Even if a good fit is obtained with a cumulative exposure risk model it cannot be confidently assumed that the effects of intensity and duration are equivalent. When risk was modelled as a function of cumulative exposure and time since first exposure using the case-control procedure the fit was nearly as good as with the model using durations at specific exposures (the log likelihood was $-42.5$, as compared with $-36.2$). Risk estimates from the two analyses were quite different, however, when compared in terms of intensity. The estimated coefficient for cumulative exposure ($0.00202$) was used to extrapolate risk coefficients for the four intensity levels in our original analysis and table 3 shows the results. The coefficients for levels 2 and 3 obtained by direct estimation (using the case-control results for comparability) are well above the range of risk extrapolated from cumulative exposure. The estimates from the two analyses show better agreement at level 4, which is expected as cumulative exposure is strongly influenced by high intensity. Even so, the extrapolated coefficient for level 4 is high compared with the direct estimate, the last being closer to the extrapolated coefficient for the low end of the intensity range than its midpoint.

A consequence of inappropriate use of cumulative exposure is that analyses based on cohorts exposed to differing levels of intensity may give conflicting results. For example, consider a situation in which risk is linearly related to cumulative exposure. If the effect of intensity is not simply proportional to its contribution to cumulative exposure, then the slope of the line will depend upon the distributions of intensities and durations that generated the cumulative exposures in the data. Data from a cohort with differing exposure intensities would give different slope estimates and extrapolation of results to other intensities would produce erroneous results.

The validity of any risk estimate depends on the adequacy of the model used and the assumptions on which it is based. The model described above assumes that the exposure variables have multiplicative effects on risk and that relative risk increases exponentially with duration at a specified intensity. To determine whether an alternative model might be more appropriate we also fitted an additive relative risk model and obtained very similar results. The two models differ substantially only at longer durations, when the multiplicative model predicts higher risk. Because exposures were short in our cohort, both models fit the data equally well and it was not possible to determine which might be more appropriate for longer durations.

Although the methods we employed for estimating risk are well established in statistical publications, their usefulness for separating the effects of intensity and duration have seldom been exploited in occupational epidemiology. Methods of analysis based on grouped data, which compare observed and expected deaths for specified exposure categories, are less useful for addressing this issue. Unless intensities are reasonably constant over time, grouping based on both intensity and duration will produce an unworkable number of exposure categories. If we had grouped durations at our four exposure levels into only three categories there would have been $4^3 = 64$ combinations. With only 21 deaths in the cohort few of these exposure groups would contain adequate data for estimating risk. If this difficulty were circumvented by using average intensity, reducing the number of categories to 12, variations in intensity would be ignored in much the same way as when cumulative exposure is used.

Although any method for assessing exposure-response from epidemiological data has shortcomings, some attempt should be made to separate the effects of intensity and duration whenever such data are available. This is particularly important when the ultimate objective is to set safety standards in terms of intensity. Extrapolations based on cumulative exposure may be misleading unless intensity and duration impact equally on risk. The analyses in this paper illustrate a straightforward method for dealing with the problem.

Requests for reprints to: Pamela M Vacek, Medical Biostatistics, College of Medicine, The University of Vermont, Burlington, VT 05405, USA.

Risk assessment using exposure intensity: an application to vermiculite mining


6 Berry G. The analysis of mortality by the subject-years method.


Accepted 17 December 1990

---

Destruction of manuscripts

From 1 July 1985 articles submitted for publication will not be returned. Authors whose papers are rejected will be advised of the decision and the manuscripts will be kept under security for three months to deal with any inquiries and then destroyed.
Risk assessment using exposure intensity: an application to vermiculite mining.
P M Vacek and J C McDonald

doi: 10.1136/oem.48.8.543

Updated information and services can be found at:
http://oem.bmj.com/content/48/8/543

These include:

Email alerting service
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/