The limitations of case-control studies in the detection of environmental carcinogens

IAIN K. CROMBIE

From the Cancer Epidemiology Research Unit, University of Birmingham

SUMMARY The ability of the case-control study to detect human carcinogens has been investigated theoretically for varying fractions of the population exposed to hazards carrying different relative risks. The method is shown to be useful for the investigation of factors to which exposure is widespread (for example, common foods or beverages) but it is of limited use for the study of uncommon types of exposure, such as those associated with occupation. The case-control study is unable to detect very small relative risks (<1.5) even where exposure is widespread and large numbers of cases of cancer are occurring in the population. The principal limitation of the method is the maximum number of cases which can be recruited and analysed. It will only be through large-scale collaborative multicentre or international studies that important risk factors will be detected.

Recent advances in our knowledge of cancer and its aetiology have led to suggestions that most if not all cancer may be preventable.1 International comparisons2,3 and migrant studies4 suggest that environmental factors are responsible for up to 90% of human cancer. Epidemiological studies have had an encouraging degree of success in detecting some of these factors: the carcinogenic properties of tobacco smoke, ionising radiation, asbestos, and a wide range of other factors have been clearly established.1 The IARC programme on the evaluation of the carcinogenicity of chemicals has examined the available evidence on 368 chemicals and concluded that for 26 of them 'a positive association between exposure and the occurrence of cancer in man was observed'.4

At present considerable effort is directed towards the detection of additional risk factors in the hope that reducing exposure will reduce cancer incidence.5 It would thus seem important to ask what kind of risk factors can be detected and where the limits of detection by epidemiological methods lie. This paper investigates the effects of various factors on the study size required to detect an increased risk of cancer after exposure to a carcinogen. It does not present any methodological developments but draws attention to several important factors which could affect both the planning and the interpretation of case-control studies. The range of values of the relative risk, the incidence of the disease and the fraction of the population exposed which are investigated have been selected to take realistic values for a cancer hazard.

THE CASE-CONTROL STUDY

The principal features of the case-control study have been fully discussed by Mantel and Haenszel,5 McMahon and Pugh,6 and also in two recent monographs.7,8 In brief, samples of cases and controls are classified by exposure to a certain substance:

<table>
<thead>
<tr>
<th></th>
<th>Exposed</th>
<th>Not exposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cases</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>Controls</td>
<td>c</td>
<td>d</td>
</tr>
<tr>
<td></td>
<td>M1</td>
<td>M2</td>
</tr>
</tbody>
</table>

The bias in the recruitment of cases or controls and in the measurement of their exposure to study factors has been reviewed in detail recently by Sackett9 but for the sake of simplicity, the following discussion considers only the situation in which bias has not occurred and where potential confounding factors do not complicate the situation.

The factors which will determine the distribution of cases (and controls) between exposed and non-exposed can be defined as follows:

I—the incidence of the disease among the non-exposed
r—the relative risk, that is, the risk of disease among the exposed relative to the non-exposed

\( \frac{1}{f} \)—the fraction of the population exposed

\( N_1 \)—the number of cases of the disease

\( a \)—the number of controls per case, that is, \( N_2 = aN_1 \)

From these definitions it is shown in the Appendix that the proportions of exposed cases

\( \frac{a}{a+b} \) and exposed controls \( \frac{c}{c+d} \) are:

\[ \frac{a}{a+b} \propto \frac{r}{f+r-1} \quad \frac{c}{c+d} \propto \frac{1-rI}{f-1(f+r-1)} \]

As was shown by Mantel and Haenszel\(^8\) the minimum study size necessary to detect a relative risk as being significantly greater than 1·0 can be calculated by the method to differentiate between two binomial proportions. The equation for the calculation of sample size without correction for continuity is described in the Appendix.

The problem of the likelihood of detecting a risk, the power of the study, has been explored by Schlesselman,\(^9\) Walter,\(^10\) and Mack et al\(^11\); for this paper a power of 80% and a significance level of 5% were used throughout. The number of controls per case and its effect on the required number of cases has been discussed by Gail et al\(^12\); a ratio of one control per case has been used throughout.

The size of the study
In practice the size is limited by the number of ascertainable cases and by the amount of work entailed. To obtain an indication of the number of cases which are commonly analysed, we studied 115 published case-control studies. These were drawn from a computer search of papers with ‘case-control’ in the title, and a detailed personal search of specialised epidemiological and general cancer journals for studies on breast and bladder cancers and malignant melanoma. Although neither exhaustive nor random, this sample should serve to place realistic limits on the range of study sizes.

The size of the studies investigated ranged from 9 to just over 1000 cases, with two exceptions (breast cancer 4323 cases and all sites 7518) (Table 1). The ‘all sites’ studies were usually subdivided into individual sites for analysis, so that the effective number of cases would be considerably smaller. Apart from the ‘all sites’ group, only seven studies had over 600 cases, a further 13 had over 400 cases, whereas just over half had less than 200 cases.

The difficulty of obtaining larger numbers of cases can be illustrated by considering the occurrence of cases of bladder cancer in a medium-sized city of, say, 500 000 people. The annual incidence of this moderately common tumour in western countries is about 20 per 100 000 for men and 6 per 100 000 for women,\(^14\) so that about 65 cases would occur each year. Some of these will be post-mortem diagnoses, others will die before interview, and some will live in outlying areas and be difficult to contact. It would require several years of continuous effort to interview even 200 cases.

The incidence of the disease
The incidence of many cancers increases rapidly with age, so to investigate the effects of incidence (rather than the confounding factor age) we have assumed that all the groups have the same age structure, and can thus treat the incidence used as an appropriately weighted average. Site-specific incidences for whole populations fall within the range from less than one to 250 per 100 000 people, although the rates for certain sites among older age groups can be as high as 700 per 100 000, as, for example, lung cancer among men aged 70–74 in Liverpool.\(^14\)

The principal effect of incidence is on the availability of cases: there are about 20 times as many male cases of bladder cancer as of liver cancer. But the proportion of controls exposed to a risk

\[ \frac{c}{c+d} \]

can also be affected by the incidence, as the expression derived in the Appendix shows:

\[ \frac{c}{c+d} \propto \frac{1-rI}{f-1(f+r-1)} \]

This follows because the combination of a very common disease and a large relative risk associated with exposure would lead to a depletion of the
exposed controls. In practice, however, only when the incidence rose above 1000 (per 100 000) and the relative risk was above 6:0 would this effect become important (Fig. 1). Even with a large relative risk of 20:0, incidences below 300 (per 100 000) would have no effect. It is clear that when I is small the expression for
\[
\frac{c}{c + d}
\]
becomes
\[
\frac{1}{f}
\]
which is the frequency of exposure in the whole population. The following calculations in this paper all use the same incidence of 10 per 100 000.

The relative risk and the fraction of the population exposed

There are two important questions to be answered when assessing the ability of case-control studies to detect a carcinogen: what is the lowest relative risk we would wish to detect and what is the lowest which the case-control method can detect. An indication of the range of relative risks to be detected can be gained by considering those associated with proven carcinogens: Table 2 shows that the relative risks of some well-known carcinogens are in the range 1·2 to 400.15--19 Large relative risks will always be important; the problem is how small a relative risk should we be able to identify. A different approach is to ask how many cases of cancer a particular risk factor would cause. Table 3 shows the number of cases which would result if various relative risks were applied to different cancer sites in circumstances where either 2% or 50% of the population were

<table>
<thead>
<tr>
<th>Carcinogen</th>
<th>Site</th>
<th>Risk factor</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\beta)-Naphthalene</td>
<td>Bladder</td>
<td>8-7</td>
<td>Case et al(^{14})</td>
</tr>
<tr>
<td>X-ray therapy of</td>
<td>Leukaemia</td>
<td>9-5</td>
<td>Court-Brown and Doll(^{14})</td>
</tr>
<tr>
<td>ankylosing spondylitis</td>
<td>Lung</td>
<td>1-0</td>
<td>Doll and Peto(^{17})</td>
</tr>
<tr>
<td>Tobacco smoking</td>
<td>Lung</td>
<td>1-8</td>
<td>Gricite(^{18})</td>
</tr>
<tr>
<td>15--24 g/day</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vinyl chloride</td>
<td>Liver (angiosarcoma)</td>
<td>400-0</td>
<td></td>
</tr>
<tr>
<td>Asbestos</td>
<td>Lung</td>
<td>1-2-- 9-2</td>
<td>Enteline(^{19})</td>
</tr>
</tbody>
</table>

Fig. 1 The minimum number of cases required to detect a significant risk with different incidences of disease (2-tailed test, \(p \leq 0-05\), with power 80%: using one control per case) △——△ risk factor 6-0 with 1% of the population exposed; ○——○ risk factor 10-0 with 1% exposed; ■——■ risk factor 20-0 with 1% exposed

Fig. 2 The minimum number of cases required to detect a significant risk with different fractions of the population exposed (2-tailed test, \(p \leq 0-05\), with power 80%: using one control per case and disease incidence 10 per 100 000); ○——○ risk factor 1-5; ■——■ risk factor 2-0; △——△ risk factor 3-0
exposed. The level at which the tumour burden of a particular risk factor is deemed unacceptable to society is a matter for debate but Table 3 demonstrates that a small relative risk (1.2) acting at a common site (lung) would produce four times as many tumours as a moderate relative risk (5.0) acting at a less common site (liver). It is also apparent that for each site a small relative risk (1.2) to which 50% of the population are exposed causes more cancer than a moderate relative risk (5.0) if exposure is restricted to 2%. In general, small relative risks can present a major problem if associated with widespread exposure or a common site. A worthwhile objective of the case-control study would be to detect those risk factors which cause the most tumours. Thus the critical test would concern the ability to detect small relative risks with common exposure and moderate relative risks with rare exposure.

The types of potential carcinogens to which large fractions of the population might be exposed would include air pollution, and substances in common foods and beverages. Fig. 2 shows that medium-sized relative risks (greater than 3.0) would be detected easily when between 5% and 90% of the population were exposed. The minimum number of cases required is smallest when between 40% and 50% of the population is exposed. Around the minimum value similar numbers of cases are required for a wide range of frequencies, and this range is larger for a relative risk of 3.0 than for 1.5. It is interesting to note that it is as difficult to detect an increased cancer risk when exposure is very common (> 90%) as when it is rare.

Relative risks of 2.0 or less become increasingly difficult to detect (Fig. 3). Even with 50% of the population exposed, risks of less than 1.4 would be difficult to detect, but when only 5% are exposed risks of 1.8 would frequently escape detection. It is clear that very small relative risks are nearly impossible to detect. A relative risk of 1.1 with 50% of the population exposed, which is the most favourable situation, would require a study of 6917 cases.

Exposure to some carcinogens will affect only small fractions of the populations. Examples of uncommon exposures include the use of certain prescribed drugs, unusual dietary items, or exposure to chemicals associated with specific occupations. Many human carcinogens have been detected through the study of occupationally exposed groups, but such groups constitute only a very small
The limitations of case-control studies in the detection of environmental carcinogens

The proportion of the population. For example, in the census of 1971, coal miners underground comprised 1.1% of all men aged 15 or over, leather workers 0.3%, plastic workers 0.2%, and radiographers 0.01%.

If 1% of the population is exposed then relative risks of 4.0 or greater will be detected easily. But if the fraction exposed is only 0.2% then only relative risks above 10.0 will be detected (Fig. 4). At smaller fractions exposed the number of cases required increases very rapidly: thus to detect a relative risk of 20.0 would require 917 cases if 0.5% were exposed and 4569 cases with 0.01% exposed.

Discussion

The problems of sample size and the sizes of detectable relative risks in case-control studies have been considered in several papers. The present paper has considered the consequences for the detection of risk factors of the variables taking values likely to be encountered in practice.

The principal strength of the case-control method lies in the detection of risk factors to which exposure is widespread in the environment. To detect a given relative risk, the required study size reaches a minimum when between 40% and 50% of the population is exposed. Thus this method could be suitable for the investigation of hazards associated with additives or contaminants of common foods or beverages or even cosmetics. But it becomes progressively more difficult to detect relative risks when exposure is rare or very common. Outside the range of 5% to 90% of the population exposed, it is very difficult to detect the effect of moderate or small relative risks.

The difficulty with rare exposures (an analogous but converse argument applies to very common ones) occurs because the proportions of exposed cases and controls will be based on small numbers and will be potentially very variable. Thus unless the relative risk is large, case-control studies will be unlikely to detect a cancer hazard. Occupational exposures fall into this category: even numerous groups such as underground mine workers constitute only 1.1% of the male population, and most occupations contain fewer persons. A further reduction in the size of the exposed group may occur because not all the individuals in an occupation may be exposed; for example, among American rubber workers and in the British rubber and cable-making industries, carcinogenic exposure was restricted to those engaged in specific sectors of the industry. In general, because of the very low proportion of exposed persons, case-control studies are unsuitable for the detection of occupational carcinogens. However, in some local areas the proportion of the population employed in a particular industry may be much higher than the national average: for example, mine workers in South Wales or pottery workers in Stoke-on-Trent. Thus the hazards in specific industries may be investigated by selecting an area for study where frequency of a particular occupation is high.

An advantage of the case-control study is that it can be used not only to test the role of a specific factor in the aetiology of a particular cancer, but also to investigate the possible role of a wide variety of other factors. This is in contrast to the cohort study, where a group of individuals is investigated because its members have been exposed to the single factor of interest (although its effect on a wide variety of diseases can be observed). But care must be taken in the interpretation of analyses where a large number of associations are tested for statistical significance, because the likelihood of false-positives, of spuriously significant results, is greatly increased. For example, Cochran and Cox have calculated that if several independent t-tests are performed, the probability of at least one spuriously significant result is 0.23 for five tests, 0.40 for 10 tests, and 0.64 for 20 tests. The investigator should first test his principal hypothesis or hypotheses for statistical significance, and then any further associations which emerge should be treated as interesting observations which would require further study for their significance to
be evaluated. The problems involved in multiple comparisons and some solutions to them have been discussed in depth by Miller.23

Case-control studies are also susceptible to the problem of false-negatives, of failing to declare as significant a real increased cancer risk. The examples discussed in the present paper have illustrated the difficulty in small studies of detecting increased relative risks with the comparatively low power of 80%. This degree of certainty would lead to an average of one study in five failing to detect a given relative risk as being significantly greater than one. In view of the large number of relatively small studies which are published each year, it would not be surprising if many failed to detect real carcinogenic hazards. Two recent papers24,25 have reviewed several series of case-control studies in which conflicting results have been reported. Although the authors discussed these conflicts in terms of methodological differences between the studies, it is possible that some of the discrepancies may have been due to the adverse play of chance. It is clear that reporting a failure to detect a significantly increased relative risk can be misleading and is also wasteful of the available information. Instead a confidence interval for the relative risk should be calculated (for a review of methods see Fleiss26) to indicate the possible size of the true relative risk which could be associated with the exposure.

The effects of bias are not considered in this paper, although these are likely to exert a major influence on the detection sensitivity of the case-control study. A total of 35 sources of bias have been catalogued to date, and nine of them recently discussed in detail.27 Copeland et al.27 have shown that 'misclassification bias' (incorrect separation of exposed and non-exposed) will bias the estimated relative risk toward unity if both cases and controls are equally affected. But when one group is more greatly affected (for example, selective recall) then the bias can be in either direction. Schlesselman28 has discussed the effects of confounding variables and indicated the basis of the widely held view that 'large values for relative risks are unlikely to be completely explained by some uncontrolled variable'. In general bias will make it more difficult to detect human carcinogens, particularly when exposure carries a small relative risk.

A major limitation of the case-control method is its inability to detect small relative risks even when associated with widespread exposure. Small relative risks might be considered of low importance, but if they are experienced by a large section of the population, such risks could cause many hundreds of cancers each year. This limitation is likely to be compounded by the effects of bias and may represent one of the major problems in the detection of environmental carcinogens: it is a disturbing thought that a cancer hazard in the tea or coffee we drink or the food we eat could be beyond the powers of detection of standard epidemiological methods.

The above limitations of the case-control study derive principally from the practical considerations which limit the number of cases which can be collected and analysed. Thus the incidence of the disease under study will be important: except for high-incidence sites, such as male lung or female breast, it will be difficult to recruit more than a few hundred cases. But once cases are collected, the method of analysis is essentially independent of the incidence of the disease, so that risk factors affecting low- as well as high-incidence sites can be studied.

Some increase in the detection sensitivity of the case-control study can be achieved by increasing the number of controls. Gail et al.29 have presented a methodology whereby the optimum number of controls in any situation can be calculated. But an increase in the number of controls increases the total study size, and there is a limit to the number of subjects who can be interviewed and analysed by an individual or a small research team. Only when studies are carried out on a multicentre or international basis with large collaborating teams will sufficient cases be analysed for definite statements to be made about all types of carcinogenic risk factors. In such multicentre studies, however, particular care will have to be taken in the methods of case and control recruitment and in the subsequent analyses to ensure that systematic differences between the areas do not confound the results.

I thank Mr. A. Girling for statistical advice and Ms. A. Peters, Mr. T. Marshall, and the editor of this journal for advice in the preparation of the manuscript.

Reprints from Dr. Iain K. Crombie, ARC Epidemiology Research Unit, Stopford Building (University of Manchester), Oxford Road, Manchester M13 9PT.

Appendix

From the definitions given in the methods section, it follows that the probability of being both exposed and a case is

\[ r \times \frac{1}{f} \]

and that of being non-exposed and a case is

\[ (1 - \frac{1}{f}) \times 1 \]
The limitations of case-control studies in the detection of environmental carcinogens

Thus among cases the ratio of exposed to non-exposed is
\[ r_f : (1 - r_f) \]
and thus the proportion of the cases exposed, \( p \), is
\[ \frac{r_f}{1 - r_f} \]
which becomes
\[ \frac{r}{f + r - 1} \]
Similarly the probability of being both exposed and a control is
\[ \frac{1 - r_f}{f} \]
and of a control and non-exposed is
\[(1 - \frac{1}{f})(1 - 1)\]
Thus the ratio of exposed: non-exposed is
\[ \frac{1 - r_f}{f} : (1 - \frac{1}{f})(1 - 1) \]
and thus the proportion of controls exposed, \( p_2 \), is
\[ \frac{1 - r_f}{f - 1}(f + r - 1) \]
The well-known equation to calculate the sample size required to differentiate between two proportions (based on equal-sized samples) is given by Schlesselman:
\[ N = \left\{ \frac{Z_\alpha^2 pq + Z_\beta^2 (p_1 - 1)p_1 + p_2 - 1)p_2}{(p_1 - p_2)^2} \right\}^2 \]
where \( p_1 \) and \( p_2 \) are the frequencies of exposure among cases and controls derived above; \( \bar{p} = \frac{(p_1 + p_2)}{2} \) and \( \bar{q} = 1 - \bar{p} \); and \( Z_\alpha \) and \( Z_\beta \) are the points in the standard normal distribution defined by the type I and II error rates, \( \alpha \) and \( \beta \).
It can be easily shown that with \( N \) cases and \( aN \) controls the formula becomes:
\[ N = \left\{ \frac{Z_\alpha^2 pq + Z_\beta^2 (p_1 - 1)p_1 + p_2 - 1)p_2}{(p_1 - p_2)^2} \right\}^2 \]
where
\[ p = \frac{P_1 + ap_2}{1 + a} \]
and \( q = 1 - p \)

References