

Assessing the possible extent of confounding in occupational case-referent studies

by Harvey Checkoway, PhD,¹ Gwen T Waldman, PhD^{1,2}

CHECKOWAY H, WALDMAN GT. Assessing the possible extent of confounding in occupational case-referent studies. *Scand J Work Environ Health* 11 (1985) 131–133. Case-referent studies nested within occupational cohort studies are efficient designs for estimating relative risks associated with exposures and for obtaining data on nonoccupational confounding factors. Frequently, however, only incomplete data on confounders can be obtained from sources such as next-of-kin or co-workers. A method for assessing the potential extent of confounding in cohort studies described previously by Axelson [*Scand J Work Environ Health* 4 (1978) 98–102] can be adapted for use in nested case-referent studies. An example is shown to illustrate that the adapted Axelson method is preferable to the usual confounder adjustment techniques when data on the confounder are incomplete.

Key terms: epidemiology, statistical methods.

Case-referent studies nested within retrospective studies of occupational cohorts have been used effectively to specify high-risk work areas and to quantify exposure-response associations (3, 6). Furthermore, attempts to obtain data on potentially confounding variables are facilitated in nested case-referent studies because study subjects typically represent a fraction of the total source cohort.

Axelson (2) has described a method for evaluating the extent of distortion in the crude rate ratio obtained from a cohort study that could result from asymmetrical distributions of a confounding factor in the study and reference populations. Steenland et al (10) recently reviewed existing methods, including Axelson's approach, for assessing confounding in occupational cohort studies. The present communication describes an application of Axelson's method to nested case-referent studies in which data on the potential confounders are either incomplete or of questionable reliability.

Control of confounding

When reliable data on a confounding factor are available in a case-referent study, the evaluation and control of confounding can be accomplished effectively by means of stratified analysis, standardization, or mathematical modeling techniques (4). Incomplete ascertainment of information on a confounder will necessitate a controlled analysis for the

subset of study subjects for whom data have been obtained, and comparison of crude and adjusted results will reveal the extent of confounding. However, indiscriminate use of statistical adjustments for confounding factors can produce erroneous results when the extraneous factor is also an effect modifier, ie, interacts with the effects of the study factor (4).

Frequently, in occupational epidemiologic research, data on nonoccupational confounding factors, such as cigarette smoking or alcohol consumption, are less reliable than occupational exposure information because the sources for data on confounders (eg, responses to questionnaires by next-of-kin or fellow workers) are not as objective as the sources of occupational exposure data. Kupper (5) recently demonstrated that the unreliability of a surrogate for a confounding factor is more likely to produce biased results than unreliability in surrogate measures of either the exposure or disease outcome variables.

In situations where there is considerable doubt about the reliability of the data on a confounder, a sensitivity analysis showing the possible extent of confounding on the estimate of effect is therefore preferable to statistical adjustment.

Axelson (2) showed that the total incidence rate of disease in the i^{th} population or exposure group can be expressed as:

$$I_i = RI_oP_i + I_o(1-P_i), \quad (\text{ex } 1)$$

where R is the independent incidence rate ratio associated with the confounding factor, usually estimated from previous studies, I_o is the baseline incidence rate of disease among persons not exposed either to the confounder or to the exposure under study, and P_i is the proportion of the i^{th} population exposed to the confounder. Asp (1) recently applied expression 1 to adjust for smoking habits in a study

¹ Department of Epidemiology, University of North Carolina, School of Public Health, Chapel Hill, North Carolina, the United States.

² Current address: Epidemiology Branch, National Institute for Environmental Health Sciences, Research Triangle Park, North Carolina 27709, the United States.

Reprint requests to: Dr H Checkoway, Department of Epidemiology, University of North Carolina, School of Public Health, Chapel Hill, NC 27514, USA.

of lung cancer in various occupational groups with differing distributions of smokers. In the case where there are two groups ($i=1, 2$) of exposure to the study factor, the incidence rate ratio due to confounding will be:

$$\frac{I_2}{I_1} = R_C = \frac{(R-1)P_2 + 1}{(R-1)P_1 + 1}, \quad (\text{ex } 2)$$

where group 1 is the reference category.

It should be noted that the assumption of a common baseline incidence rate, I_0 , for all exposure groups applies only when the exposure groups are derived from the same source population. In an occupational cohort study, this assumption is met provided that the occupational cohort is a subset of the reference population (eg, national population) or the cohort and reference population are subsets of a common source (ie, exposure subcohorts from the overall study cohort).

In a case-referent study, nested within an occupational cohort, the potential extent of confounding can be assessed with expression 2, even though the odds ratio estimate for exposure is computationally and, in some investigators' view, conceptually different from the rate ratio formulation given in expression 2. [Miettinen (9) has shown that the odds ratio is an unbiased estimator of the rate ratio, and it is convenient to consider the odds ratio as an estimate of the rate ratio parameter.]

In a nested occupational case-referent study, the proportions of subjects exposed to the confounder are determined from the distributions of the confounder among referents. Similarly, this type of information on a confounding factor can be obtained

in case-referent studies which are not nested within occupational cohorts. Thus, in expression 2, P_2 represents the proportion of referents known to be exposed to the confounding factor among the referents who are exposed to the study factor, and whose confounding factor exposure status is known, and P_1 represents the proportion of referents not exposed to the study factor but known to be exposed to the confounder among those whose confounding status is known. As in a cohort study, R is the rate ratio associated with the confounder, as indicated from previous research, and R_C is computed as an estimate of the rate ratio associated with the confounder. R is estimable from data on the confounder obtained in the occupational case-referent study; however, a more reliable estimate of R generally can be obtained from previous investigations specifically designed to quantify the rate ratio associated with that factor.

Example

The hypothetical data given in table 1 illustrate the application of Axelson's method for evaluating the extent of a potential confounding factor. These data are for a study of 100 cases and 200 referents for whom exposure to the study factor (eg, occupational carcinogen) can be determined for all subjects, but exposure information for some confounding factor is incompletely ascertained. In this example, a rate ratio of 3.0 is assumed for the confounding factor, based on previous knowledge. The crude odds ratio for exposure to the study factor (1.86) is obtained simply from the marginal totals of the exposed and unexposed among the cases and referents. Among persons not exposed to the study factor, the odds ratio associated with exposure to the confounder (2.50) can be estimated from the distribution of ascertained data on exposure to the confounder among the cases and referents.

Application of expression 2 yields a rate ratio estimate of 1.16 due to confounding. This estimate is based only on data from persons with known values of exposure to the confounding factor. It is also possible to compute a plausible range for R_C from extreme assumptions regarding the distributions of the exposure to the confounder among persons with missing data. The two most extreme situations would be (i) where all of the unknowns among the referents exposed to the study factor are assumed also to be exposed to the confounder and all unknowns among the referents unexposed to the study factor are assumed to be unexposed to the confounder and (ii) where the unknowns among the referents exposed to the study factor are assumed to be unexposed to the confounder and all the unknowns among the referents unexposed to the study factor are assumed to be exposed to the confounder. In the hypothetical example of table 1 the range for R_C resulting from these two conditions is 1.50 and 0.83.

Table 1. Hypothetical example of case-referent distributions according to exposures to a study factor and a confounding factor. A rate ratio of 3.0 is assumed for the confounding factor, as determined from previous investigations.

Exposure to study factor	Exposure to confounding factor			
	Yes	No	Unknown	Total
<i>Cases</i>				
Yes	20	20	10	50
No	20	10	20	50
<i>Referents</i>				
Yes	30	20	20	70
No	40	50	40	130

Crude odds ratio for exposure to study factor = $(50)(130) \div (50)(70) = 1.86$.

Estimated odds ratio for exposure to confounder^a = $(20)(50) \div (10)(40) = 2.50$.

Estimated rate ratio due to confounding associated with exposure to study factor: $R_C = [(3-1) \left(\frac{30}{30+20} \right) + 1] \div$

$$[(3-1) \left(\frac{40}{40+50} \right) + 1] = 1.16.$$

^a Among persons not exposed to study factor with known values for exposure to the confounder.

It can also be noted that the adjusted odds ratio computed for the study factor is 2.10 when adjustment is made for the distribution of the confounder (8), but only among persons with known values of the confounder. This value suggests an R_C of 0.89, which is obtained by dividing the crude by the adjusted odds ratio (7).

Discussion

In their review of options for controlling confounding effects of nonoccupational exposures, Steenland et al (10) described the desirability of conducting nested case-referent studies for examining both the extent of confounding and the interaction between the study factor and the extraneous covariate. For reasons mentioned previously, we recommend a sensitivity analysis using Axelson's approach (2) rather than statistical adjustment procedures for situations where data on the confounder are incomplete or likely to be unreliable. In addition unreliability of the data on a nonoccupational covariate is likely to introduce substantial errors in the estimates of interactive effects between the study factor and the covariate on the disease rate, in much the same way that adjustment for a poorly measured confounder yields biased summary estimates of the rate ratio (5).

References

1. Asp S. Confounding by variable smoking habits in different occupational groups. *Scand J Work Environ Health* 10 (1984) 325—326.
2. Axelson O. Aspects on confounding in occupational health epidemiology. *Scand J Work Environ Health* 4 (1978) 98—102.
3. Breslow NE, Lubin JH, Marek P, Langholz B. Multiplicative models and cohort analysis. *J Am Stat Assoc* 78 (1983) 1—12.
4. Kleinbaum DG, Kupper LL, Morgenstern H. *Epidemiologic research: Principles and quantitative methods*. Lifetime Learning Publications, Belmont, CA 1982, pp 312—319.
5. Kupper LL. Effects of the use of unreliable surrogate variables on the validity of epidemiologic research studies. *Am J Epidemiol* 120 (1984) 643—648.
6. Liddell FDK, McDonald JC, Thomas DC. Methods of cohort analysis: Appraisal by application to asbestos mining. *J R Stat Soc A* 140 (1977) 469—491.
7. Miettinen OS. Components of the crude risk ratio. *Am J Epidemiol* 96 (1972) 168—172.
8. Miettinen OS. Confounding and effect modification. *Am J Epidemiol* 100 (1974) 350—353.
9. Miettinen OS. Estimability and estimation in case-referent studies. *Am J Epidemiol* 103 (1976) 226—235.
10. Steenland K, Beaumont J, Halperin W. Methods of control for smoking in occupational cohort mortality studies. *Scand J Work Environ Health* 10 (1984) 143—149.

Received for publication: 4 December 1984