American Journal of Industrial Medicine 13:105-118 (1988)

Indirect Methods of Assessing the Effects of Tobacco Use in Occupational Studies

Olav Axelson, MD, and Kyle Steenland, PhD

For various reasons, data on smoking are frequently missing, or only partially available, in retrospective epidemiologic studies of occupational risk factors. In such situations, indirect methods may be used to evaluate the magnitude and direction of the potentially confounding effects of smoking. Such an evaluation can be made quantitatively or qualitatively. Here we describe both approaches. A specific problem relates to case-referent studies, where sampling variation in referent selection may limit the possibility of controlling for confounding by smoking, even when smoking data are available. We present data showing that estimates of risk from occupational exposures which are not controlled for smoking may be as accurate as estimates derived after controlling for smoking, when the number of referents is relatively small. The problem of interaction is also discussed. In the absence of smoking data, the investigator has no indication of how smoking and occupation jointly affect disease risk (eg, additively or multiplicatively). The multiplicative model is usually assumed. However, if exposure and smoking act independently (additively), rate ratios are diminished. In such situations, in the presence of negative confounding by smoking, rate ratios may actually even be less than one—also when exposure and disease are strongly related.

Key words: epidemiologic studies, methodology, risk model, case-control studies, smoking

INTRODUCTION

General Remarks

The influence of occupational and environmental factors has attracted little concern when smoking has been studied as a determinant of disease. In contrast, studies of occupational or environmental hazards are usually expected to control for smoking as soon as there is the slightest suspicion of a relationship between smoking and the disorder(s) under study.

It is often difficult and expensive to acquire data on smoking and other tobacco use in occupational health studies. Gathering information on smoking by interviews or questionnaires from dying patients or close relatives of newly deceased cases can also be a delicate matter. It may even be difficult to obtain approval for such data acquisition from an ethics committee. Studies undertaken without interest or support

Department of Occupational Medicine, Linkoping, Sweden (O.A.).
National Institute for Occupational Safety and Health, Cincinnati (K.S.).
Address reprint requests to Dr. Olav Axelson, Department of Occupational Medicine, University Hospital, S-581 85 Linkoping, Sweden.
Accepted for publication July 28, 1987.

from management and trade unions may also meet difficulties in efficient data acquisition. Despite such difficulties, most investigators have nevertheless been able to obtain relatively good information on smoking habits through interviews or self-administered questionnaires.

The main problem in collecting smoking data from live study subjects is that of obtaining a good response rate. Obtaining smoking data on decedents also involves questions of validity [Rogot and Reid, 1975; Lerchen and Samet, 1986]. Furthermore, the current (and highly justified) antismoking propaganda may induce live study subjects to be less frank in declaring their own smoking habits in the future (similar to the situation regarding drinking habits).

Lacking smoking data, the possible impact of smoking on epidemiologic results must be evaluated on a theoretical basis. Such an evaluation would be especially important if the study reports a positive association between the exposure and a disease strongly related to smoking (eg, lung cancer). In such cases, investigators may even find it necessary to acquire smoking data in a second phase of the study to improve the validity of the results.

Our intention here is to discuss the potential effects of smoking on the results of an occupational study, when smoking information is missing or uncertain.

Confounding and Modifying Effects From Smoking

We here define a confounding factor in the conventional sense—namely, as a determinant of the disease (independent of exposure) and also associated with the exposure [Kleinbaum et al, 1985]. Confounding occurs when such a factor causes a distortion in the measurement of the true association between exposure (eg, as occurring in an occupation) and disease. For example, when the smoking habits of an occupationally exposed group are different from those of the reference population, confounding may arise when smoking-related diseases are considered. Therefore, an association found between a disease and a particular exposure may be spurious if smoking is not taken into account, since differing smoking habits rather than the exposure might explain at least part of an observed increase of the morbidity or mortality. Usually, confounding by smoking is thought of as exaggerating the effect of the exposure under study, but masking effects also may occur. This would be the case if smoking were negatively associated with the exposure (or when smoking in some rare instances happens to be preventive, as perhaps for Mb Parkinson) [IARC 1986].

An issue of particular interest with regard to the control of confounding is the case-referent (case-control) type of study. In a situation without any (substantial) confounding in the population from which the cases are obtained (ie, no association between smoking and exposure in the base population), the referents drawn from that population might nevertheless, by random variation, misrepresent the distribution of smoking in relation to the exposure (ie, the degree of an association between smoking and exposure in the sample of referents may be different from that in the base population). Smoking would then act as a confounder in the data, although it did not confound the exposure-disease relationship in the base population. Alternatively, confounding in the base population might not be reflected in the case-referent data obtained from that population.

In addition, to the need for evaluating the potential confounding effects from smoking, it may also be of interest to consider the possibility of a modifying influence from smoking on the epidemiological risk estimates for an occupational exposure, ie, on the risk (rate) ratio or risk (rate) difference. By modification, we again adopt the conventional meaning, ie, a difference in the effect of exposure on disease depending on the level of another factor (eg, smoking) [Kleinbaum et al, 1985].

CONFOUNDING BY SMOKING

Magnitude of Effects

In general terms, the influence from smoking or another potential confounding factor might be thought of as contributing to the total incidence of the disorder in relation to its strength and prevalence in the population. This view may be expressed as

$$I = I_{CF}\, P_{CF} + I_0\,(1 - P_{CF})$$

where I is the overall incidence of the disease in the population, I_{CF} is the incidence related to the confounding factor, and P_{CF} is the proportion of the population with the factor. In this context the factor is smoking, and I_0 is the rate for nonsmokers.

If the effect of the confounder is known in terms of a risk ratio, R, the formula might be rewritten as

$$I = RI_0\, P_{CF} + I_0\,(1 - P_{CF})$$

Several terms could be included in this expression for various levels of the confounder or different confounders. Similar expressions were presented in a discussion of confounding by Cornfield et al [1959].

To estimate the magnitude of possible confounding by smoking, consider a reference population with 40% moderate smokers with a tenfold risk ratio for lung cancer and 10% heavy smokers with a 20-fold increased risk. Based on the relationship above, we have

$$I = 0.5I_0 + (0.4)(10)I_0 + (0.1)(20)I_0$$

Hence,

$$I_0 = 0.1538\ I$$

With the utilization of the above expression and the value obtained for I_0, and considering populations with different distributions of nonsmokers, moderate, and heavy smokers, Table I was constructed [Axelson, 1978]. This table illustrates the risk ratio for lung cancer which would be obtained due to smoking alone, for various hypothetical populations with differing smoking habits, compared to the reference group discussed above (40% moderate smokers, 10% heavy smokers, 50% nonsmokers).

TABLE I. Risk Ratios for Lung Cancer in Relation to the Fraction of Smokers in Various Hypothetical Populations [from Axelson, 1978]

% and type of smokers in the population			
Non-smokers (risk of 1)	Moderate smokers (risk of 10)	Heavy smokers (risk of 20)	Risk ratio
100	—	—	0.15
80	20	—	0.43
70	30	—	0.57
60	35	5	0.78
50*	40*	10*	1.00*
40	45	15	1.22
30	50	20	1.43
20	55	25	1.65
10	60	30	1.86
—	—	100	3.08

*Compared to reference population with 50% nonsmokers, 40% moderate smokers, and 10% heavy smokers.

It is clear from Table I that the potential confounding effect of smoking is rather weak in relation to the relatively high risk ratios that are often found in occupational health studies. Stated another way, smoking differences alone are unlikely to account for a large observed excess of disease in a population with a particular occupational exposure.

The degree of potential confounding depends on the strength of the association of the disease in question with smoking, and on the degree to which the smoking habits of the exposed differ from the smoking habits of the reference group.

Table I is based on lung cancer, a disease strongly related to smoking. Smoking is a much weaker determinant for most other diseases, and the confounding effects are therefore smaller. For example, the risk ratio for pancreatic cancer and bladder cancer from smoking seems to be about 2–4 with considerable variations from study to study [IARC, 1986]. When information from the US surgeon general [1979] with regard to cardiovascular deaths and smoking was used, it seemed reasonable to suggest about a 2.5-fold risk ratio for "average smokers" and 3.5 for "heavy smokers" [Edling and Axelson, 1984].

Regarding the degree to which smoking habits between exposed and reference groups may differ, the smoking habit of various worker groups have been studied by Sterling and Weinkam [1976] and Covey and Wynder [1981]. Further information is also obtainable from the US surgeon general [1985], Weinkam and Sterling [1987], Stellman et al [1988] and Brackbill et al [1988]. The data in the surgeon general's report indicate that the percentage of current smokers among US men in 1978–1980 ranged from 55.1% among painters, construction, and maintenance workers to 16.2% among electrical and electronic engineers. The average for blue-collar workers was 47.1%, while for white-collar workers it was 33.0%. Blair et al [1985] noted that the proportions of nonsmokers (never smoked) among US veterans in the 1950s varied from 51% for clergymen to 13% for bakers.

Asp [1984] reported on the smoking habits of Finnish working populations and evaluated the resulting potential for confounding effects vis-à-vis other risk factors.

She found between 7% and 49% to be nonsmokers in the various worker groups with 24% as the overall figure. The percentage of exsmokers varied from 11 to 45%, while the range for current smokers was 22–74%. Applying the expression discussed above (assuming a risk ratio of 10 for exsmokers and 20 for smokers for lung cancer), she derived a table similar to our Table I. Risk ratios for smoking alone varied from 0.67 to 1.31, comparing specific occupations to the total population composed of 24% nonsmokers, 28% exsmokers, and 48% smokers.

Let us assume that one of the specific populations in Table I, say the population composed of 40% nonsmokers, 45% moderate smokers, and 15% heavy smokers, was exposed to some occupational agent while the reference group (50% nonsmokers, 40% moderate smokers, 10% heavy smokers) was not. Let us also assume that the occupational agent in question in fact did not increase the risk of lung cancer, and that the investigators had no smoking data on either the exposed or reference group. Then the investigators might falsely conclude that exposure to the occupational agent resulted in a risk ratio of 1.22. This result would differ from the true risk ratio (1.00), due to confounding by smoking. For our purposes here, let us define the confounding risk ratio as the crude (unadjusted for smoking) risk ratio (for disease due to exposure) divided by the true (adjusted for smoking) risk ratio (for disease due to exposure) [Miettinen, 1972a]. For this example, the confounding risk ratio would be 1.22. From the data presented by Asp which was discussed above, the range of the confounding risk ratio would range from 0.67 to 1.31.

These theoretical considerations might also be compared with findings from the study by Blair et al [1985], where comparisons were made between crude and smoking-adjusted standardized mortality ratios (SMRs). As expected from the theoretical points made above, Blair et al found small differences between crude and smoking-adjusted SMRs. The most extreme case involved bakers and cooks, for whom the SMR for lung cancer shifted from 1.57 to 2.25 when adjusted for smoking. The confounding risk ratio was therefore 1.57/2.25 = 0.70, ie, close to one of the extremes calculated by Asp.

To the extent that relatively low excess risks are considered, the control of confounding by smoking is increasingly important and may change the result of a study. For example, Suta and Thompson [1983] found that the proportional mortality ratio for lung cancer among automotive workers declined from 1.3 to about 1.1 after an adjustment for smoking using the procedures discussed above. This new mortality ratio, in contrast to the old, was nonsignificant. In such situations investigators may want to go beyond theoretical adjustments and collect new data on smoking from the population in question.

A Note on Confounding in Case-Referent Studies

Confounding due to smoking (or other factors) in case-referent studies may deserve a special comment. The referents in such a study are supposed to represent the entire study population (or base population) over a period of time with regard to the various exposures to be evaluated, as well as with regard to potential confounders. However, the referents may not be representative because of the influence from random variation in their selection. Therefore, even if there is no confounding in the base population itself, as would occur if there was no association between the potential confounder (smoking) and exposure in the population, the obtained distribution of the smoking and exposure variables in the referent sample may nevertheless suggest confounding to be present (positive or negtaive).

The possibility of a poor representation of the base population over time by the referents is considerable, if the sample of referents is small [Axelson et al, 1984]. Such "data-based" confounding [Kleinbaum et al, 1985], which can occur due to random variation in sampling even though there is no confounding in the base population, can have an impact on the exposure risk ratio that might be even greater than if there had truly been confounding in the base population which was not controlled in the analysis. Tables IIa and IIb show the extreme results of simple computer simulations with 20 sets of 320 referents drawn from a base population of 160,000, under different assumptions about confounding in the base population (Fig. 1 and the notes for Table II explain the assumptions of these simulations). The referents were drawn from either an unconfounded (Table IIa) or a confounded (Table IIb) population with 50% (Table IIa) and 75% smokers (Table IIb) among the exposed, vs 50% among the unexposed (both IIa and IIb). The exposure of interest was assumed to either cause no increase in disease risk or a fivefold increase in disease risk. The risk ratio for disease due to smoking also varied, being 1.0, 5.0, 10.0, or 20.0 (as applicable to different disorders). The joint effect of the occupational exposure and smoking was assumed to be multiplicative. These tables show two types of risk ratios due to the occupational exposure—either the crude risk ratios (assuming no adjustment for smoking) or the adjusted risk ratio (controlling for smoking). The confounding risk ratio (CoRR) is the crude risk ratio divided by the adjusted risk ratio. In Table IIa (no confounding in the base population), the expected value of this confounding risk ratio is 1.0, while in Table IIb the expected value is greater than 1.0 (due to positive confounding in the base). One can observe that the range of values for the confounding risk ratio in Table IIb may overlap the corresponding range in Table IIa. Hence, due to sampling variation, it will be difficult to know whether an adjusted risk ratio has properly controlled for confounding in the base population or merely takes care of sampling variation in relation to an (essentially) unconfounded base population.

It is clear from Table II that a wide range of risk ratios is possible due to sampling variation and that formal control over smoking in some cases actually yielded a result further from the true risk ratio in the base population than the crude risk ratio. Therefore, in view of validity, it is tempting to suggest that the omission of control for smoking and other potential confounding factors should be considered relatively trivial in case-referent studies. This would especially be true when only a few hundred referents are involved, making the risk estimates unstable. Some improvement over theoretical adjustments may be achieved when actual smoking data are available through the traditional adjustment techniques (ie, stratification and weighted summary risk ratios), if there are a large number of referents and there is good information about smoking. When only partial information on smoking habits is available, theoretical adjustments may be preferable. This same point has been made recently by Checkoway and Waldman [1985].

If a cohort study (uncontrolled for smoking) shows an excess risk due to exposure for a smoking-related disease, and if there are some nonexposed or low-exposed individuals in the cohort, investigators may decide to collect data on smoking (and/or other factors) in order to do a nested case-control study to control for possible confounding by smoking (and/or other factors). The preceding discussion would

TABLE II. Sampling Variation in Case-Referent Studies and the Effects of Confounding*

Underlying risk ratio of exposure			Underlying risk ratios for smoking (or other determinants)							
			a. Absence of confounding				b. With confounding			
			(1.0)	5.0	10.0	20.0	(1.0)	5.0	10.0	20.0
1.0	Crude RR	min	0.813	0.813	0.813	0.813	0.802	1.07	1.13	1.17
		max	1.20	1.20	1.20	1.20	1.18	1.58	1.67	1.72
	Adjusted RR	min	0.816	0.766	0.755	0.748	0.74	0.705	0.700	0.700
		max	1.20	1.20	1.26	1.29	1.19	1.21	1.22	1.22
	CoRR	min	0.995	0.821	0.781	0.758	0.931	1.14	1.19	1.21
		max	1.03	1.10	1.12	1.13	1.08	1.51	1.62	1.67
	CoRR expected		1.00	1.00	1.00	1.00	1.00	1.33	1.41	1.45
5.0	Crude RR	min	3.81	3.81	3.81	3.81	4.27	5.69	6.02	6.21
		max	6.64	6.64	6.64	6.64	6.08	8.10	8.56	8.83
	Adjusted RR	min	3.76	3.65	3.52	3.44	3.75	3.53	3.49	3.47
		max	6.63	6.89	7.14	7.29	6.09	6.66	6.76	6.82
	CoRR	min	0.999	0.849	0.811	0.789	0.949	1.16	1.20	1.23
		max	1.03	1.26	1.31	1.34	1.14	1.62	1.72	1.79
	CoRR expected		1.00	1.00	1.00	1.00	1.00	1.33	1.41	1.45

*Range of crude and adjusted (of the SMR type) [Miettinen, 1972] risk ratios (odds ratios) in each sector of the table are based on 20 simulations with varying strength (taken as the risk ratio) for the exposure and for smoking (or another determinant) with regard to a multiplicative model. The confounding risk ratio, CoRR, is a measure of the confounding effect and obtained as the crude R R divided by the corresponding adjusted R R (SMR) and is expected to be unity when no confounding is present in the population used for the simulations. The same 20 sets of referents are used for obtaining the risk ratios within each sector of the table, i.e. 40 simulations were undertaken in total. Assumptions: For Table IIa, the base population was 160,000 individuals with 50% nonsmokers and 50% smokers and with 50% exposed and 50% nonexposed, with no association between exposure and smoking (or alternative, other determinant). For Table IIb, exposure and smoking were taken as associated (75% of the exposed smoking compared to 50% of the nonexposed). For both tables the risk among nonexposed nonsmokers was taken as 1/10,000, and for exposed nonsmokers either 1/10,000 (no effect) or 5/10,000. The combined risk from the exposure and smoking, for example, given a risk 5.0 for each, was 25/100,000 under a multiplicative model. Three hundred twenty randomly selected referents were drawn from the population in each simulation. Adjustment for smoking was done via stratification and indirect standardization (of the SMR type). For an illustration of these assumptions see Figure 1.

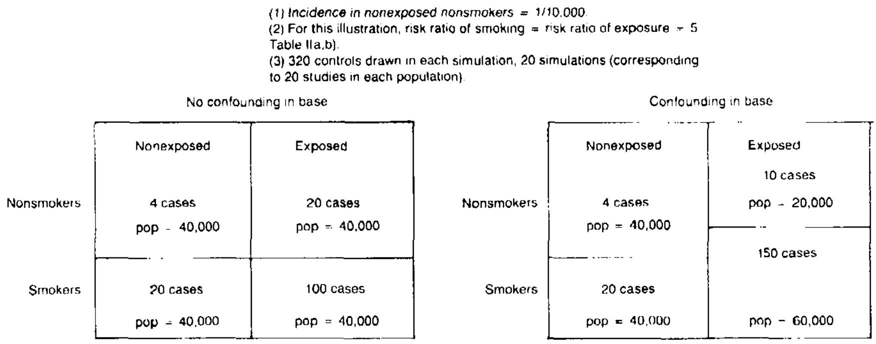

Fig. 1. Base populations underlying case-control simulations in Table II.

indicate that such an effort might not be worthwhile, unless the number of controls is large enough to ensure a fair representation of the cohort.

Judgments and Adjustments

As demonstrated in the previous section, the effects of not controlling for smoking tend to be rather modest, unless the smoking habits of the study population are quite extreme. Still, for studies with no smoking data, it is often necessary to make some estimates of the possible degree of confounding by smoking (or other factors). Such considerations about the effects of smoking have been discussed by Steenland et al [1984]. Methods of indirect control for smoking included (1) an analysis of other smoking-related diseases in the study population, (2) the use of an internal nonexposed cohort, (3) an adjustment based on hypothesized differences in smoking habits for the exposed and nonexposed populations, and (4) an analysis of the data for a dose-response relationship. We have already discussed method (3) above.

Regarding method (1), an analysis of other smoking-related diseases besides the disease of interest may indicate whether smoking or the occupational exposure caused an observed excess. If smoking were to blame, one would expect an excess of smoking-related cancers (lung, larynx, esophagus, bladder, kidney, and pancreas), and of smoking-related nonmalignant disease (emphysema, bronchitis, and coronary heart disease). A problem in the application of this principle is that industrial exposures (such as asbestos) as well as smoking might both cause an excess of multiple diseases. Furthermore, an excess of some smoking-related disease might also result from the joint action (or synergism) of smoking and exposure rather than from the weak influence of uncontrolled confounding. Finally, heart disease may not be a good indicator of excess smoking, given that in many occupational studies heart disease is decreased due to the healthy worker effect [McMichael, 1976].

In an internal reference group, method 2 is often used because such a referent group is thought to be more similar to the exposed group than other possible nonexposed groups or the national population, regarding potential confounders such as smoking. An example of the inclusion of such an internal reference group can be found in a study of lung cancer among welders by Beaumont and Weiss [1981]. However, smoking may vary considerably within the same socioeconomic categories [Covey and Wynder, 1981; Registrar General, 1978; Sterling and Weinkam, 1976;

Weinberg et al, 1982], which would decrease the value of this method for the indirect control of smoking.

A positive dose-response (method 4), showing that a disease increases with level of occupational exposure, might indicate a disease risk due to exposure which is independent of smoking (assuming those exposed to different levels of the occupational agent had similar smoking habits). In support of this hypothesis, evidence has recently been presented [Siemiatycki et al, 1986] that smoking prevalence may not vary much by degree of exposure, within a given occupational group. On the other hand, if years of employment are used as the measure of exposure, there will be a close correlation between exposure and age and therefore also with years smoked. Adequate age standardization is therefore required before any judgments can be made. Even in the presence of a true effect of the exposure, a clear dose-response pattern may not appear in terms of the risk ratio, due to the well-known problem of comparing indirectly standardized risk ratios (SMRs) for different exposure groups. This problem may occur even in the absence of confounding. In principle, a better alternative would be a direct standardization, but this is often not achievable or attractive due to small numbers of cases in the various age groups and exposure categories.

Another problem with method (4) is that short-term employees may have had intensive exposures, although of short duration. Years of employment or exposure therefore might not necessarily provide a basis for dose-response studies. Consequently, little help may be obtained from dose-response analyses for finding out about confounding in such situations; ie, the lack of a dose-response pattern does not necessarily imply uncontrolled confounding as explaining an increased risk but could rather depend on other circumstances. In general, it should be noted that the epidemiological dose concept is far from a simple and clear entity [Axelson, 1985; Checkoway, 1986], which should also be taken into account when an expected dose-response relationship does not appear.

Adjustment if Smoking Data Are Available for Part of the Cohort

Frequently, an investigator may have some smoking data for a portion of the exposed cohort. For example, company medical records may contain smoking information collected at one point in time, for a cross-section of the cohort. The investigator may then use such smoking data to compare smoking prevalence of the cohort to the referent population (eg, the US), and using the above-described techniques may adjust the expected number of deaths accordingly [for example, Edling and Axelson, 1984; Thun et al, 1985]. Such an adjustment may also take former smokers into account.

Some bias may result unless such cross-sectional data on smoking come from a point in calendar time which is relatively early in the follow-up period. It is differences in past smoking habits, at a point in time prior to disease incidence, which are most critical. Furthermore, if one compares recent smoking prevalence data in the exposed cohort with smoking prevalence in the nonexposed referent group, the comparison will exclude the majority of cohort members who have died during the follow-up period. By focusing only on cohort "survivors," one runs the risk of underestimating the cohort's smoking prevalence, given that decedents are likely to have smoked more than survivors [McLaughlin et al, 1985].

In estimating the prevalence of smoking in the exposed cohort vs the referent population, it is also useful to adjust this comparison for age differences, since smoking prevalence rates vary by age. Gail et al [1988] have discussed in more rigorous detail some of the assumptions which are commonly made in making adjustments of this type.

EFFECT MODIFICATION BY SMOKING (INTERACTION BETWEEN SMOKING AND EXPOSURE)

As indicated in the introductory remarks, not only confounding but also the question of interaction, additive or multiplicative, between smoking and the exposure of interest should be considered. In Table II, in assessing the effects of confounding, we have assumed that smoking and an occupational exposure operate jointly in a multiplicative model. However, the joint effect may be additive. If an additive model holds, rate ratios may show an artificially low measure of risk, and negative confounding in such a situation may in fact result in missing a strong effect of exposure altogether. Gail et al [1988] discuss in more detail the importance of the underlying (unknown) model (eg, multiplicative or additive) in assessing confounding.

These aspects may be illustrated by Table III, where we assume a fivefold risk from the exposure. In this table we have used the two occupational groups with the most extreme smoking habits as described by Asp [1984]. We have assumed the same risk ratios for various smoking categories as before, and we illustrate the effect of confounding assuming either an additive or multiplicative model of interaction. Under the additive model with no confounding, a fivefold risk from the exposure leads to risk ratios of only 1.47 (comparing exposed to nonexposed, both with low smoking prevalence like the civil servants) and 1.24 (comparing exposed to nonexposed, both with high smoking prevalence like the construction workers). When confounding from smoking is also present, then the respective risk ratios for the exposure are 0.99 and 1.62, comparing each group to a nonexposed group assumed to have the same smoking prevalence as the total population described by Asp. Thus, assuming the additive model and negative confounding by smoking, the observed risk ratio from the exposure no longer exceeds 1.0, although the true risk ratio is 5.0.

If a multiplicative rather than an additive effect were present between the exposure and smoking, then the risk estimates come out as 5.0 for both populations by unconfounded comparisons. When confounding is involved, the risk ratios would be 3.37 and 6.53, again in reference to a nonexposed population with the same smoking prevalence reported by Asp for her total population.

The conclusion from these examples seems to be that if the (unevaluated) joint effect between smoking and exposure is additive (and investigators report risk ratios), confounding is more likely to obscure an effect of the exposure than if the interaction is multiplicative.

INTERPRETATION OF STUDIES WITH LACK OF SMOKING DATA

As shown by the preceding discussion, the effect of not controlling for confounding by smoking tends to be relatively modest in most situations. Therefore, there is no reason to be overly critical and disregard the results obtained in studies without data on smoking, especially if the risk estimates are relatively high. The

TABLE III. Confounding in the Presence of Effect Modification

	Range of smoking prevalence in occupational groups (%) [Asp, 1984]			
	Nonsmokers	Exsmokers	Smokers	Groups
Low L	41	37	22	Civil servants
High, H	10	16	74	Construction workers
Average	24	28	48	All occupational groups

Risk assumptions: Nonsmokers = 1; exsmokers = 10; smokers = 20; exposure = 5; I_1 = disease rate in exposed, I_0 = disease rate in nonexposed

IIIa. Additive model, comparing populations similar with regard to smoking (no confounding)

Risk ratio, I_1/I_0, for two groups with low smoking frequency (similar to civil servants, L) and no confounding:

$I_{0L} = 0.41 \times 1 + 0.37 \times 10 + 0.22 \times 20 = 8.51$ (nonexposed)

$I_{1L} = 0.41 \times 5 + 0.37\,(10 + 5 - 1) + 0.22\,(20 + 5 - 1) = 12.51$ (exposed)

$I_{1L}/I_{0L} = 1.47$ = risk ratio due to exposure

Correspondingly, for two groups with high smoking frequency (similar to construction workers, H)

$I_{0H} = 0.1 \times 1 + 0.16 \times 10 + 0.74 \times 20 = 16.5$ (unexposed)

$I_{1H} = 0.1 \times 5 + 0.16\,(10 + 5 - 1) + 0.74\,(20 + 5 - 1) = 20.5$ (exposed)

$I_{1H}/I_{0H} = 1.24$ = risk ratio due to exposure

IIIb. Additive model, comparing populations dissimilar with regard to smoking (confounding), assuming nonexposed to have smoking prevalence identical to the total (average) population above

For the nonexposed

$I_0 = 0.24 \times 1 + 0.28 \times 10 + 0.48 \times 20 = 12.64$

Hence

$I_{1L}/I_0 = 12.51/12.64 = 0.99$ and $I_{1H}/I_0 = 20.5/12.64 = 1.62$

IIIc. Multiplicative model, comparing populations with and without confounding; the background risk is increased five-fold by the exposure

$I_{1L} = I_{0L} \times 5 = 8.51 \times 5$ and $I_{1H} = I_{0H} = 5 = 16.5 \times 5$ (no confounding)

Assuming the nonexposed have a smoking prevalence similar to the total (average) population above

$I_{1L}/I_0 = 8.51 \times 5/12.64 = 3.37$ (negative confounding)

$I_{1H}/I_0 = 16.5 \times 5/12.64 = 6.53$ (positive confounding)

interpretation of "relatively high" has to be in relation to the particular disease under consideration. For lung cancer and smoking as the confounder, risk ratios for the exposure of more than 1.5 or 2.0 are unlikely to depend on uncontrolled confounding from smoking. For other diseases, less related to smoking, even smaller excesses are unlikely to be the result of confounding by smoking.

Even if an observed excess is unlikely to be totally explained by confounding, adjusting a risk ratio to account for possible confounding by smoking may mean that the remaining excess risk would no longer be significant. Such a consequence obviously could change the overall interpretation of a study. To the extent that testing is based on the Poisson distribution, the crude expected number of cases may be found and then adjusted for potential confounding to derive a new expected [Edling et al, 1983].

For case-referent studies, there is no corresponding general suggestion at present for significance testing after making an adjustment for possible confounding by smoking. Nor may such testing be particularly meaningful with regard to the inherent problem of controlling confounding in case-referent studies, as indicated above. However, here we discuss one method for such testing.

For pair-matched case-referent studies, the risk ratio is taken as the quotient of the discordant pairs (p/q), assumed to be 1.0 under the null hypothesis. In a situation with uncontrolled confounding, the confounding risk ratio (CoRR), as defined above and in Table IIa, would be a component of this quotient. It is possible to derive an adjusted probability for the occurrence of exposure among cases vs referents rather than the equal probability (reflected by $p = q$, where $q = 1-p$) as assumed under the null hypothesis. This adjusted probability might be obtained by taking p/q = CoRR instead of unity.

A numerical example of such an adjustment may be taken from a study by Sjoberg et al [1977], which indicated an excess of lung cancer among shipyard workers, with no information on smoking. However, under the simplifying assumption of 80% smokers with a 20-fold risk of lung cancer among the shipyard employees, vs 50% smokers among the nonexposed, the theoretical confounding risk ratio was calculated (according to the principles discussed above) as 1.54. The ratio p/q or $p/(1-p)$ can then be assumed to equal 1.54, so that $p = 0.61$ and $q = 1-p = 0.39$.

Out of 11 discordant pairs, in ten the case was exposed and the referent nonexposed. The adjusted probability for the observed outcome (ten cases exposed) or a more extreme realization (11 cases exposed) is therefore obtained via the binomial distribution as

$$P = \sum_{X=10}^{11} (11!/(11-X)!X!)\ (0.61)^{x}(0.39)^{(11-x)}$$

resulting in $p = 0.035$.

Even if such formal significance testing is possible after an adjustment for uncontrolled confounding, many investigators may choose an alternative practical approach. Frequently investigators use some relatively crude calculations and make a general determination whether an elevated risk ratio is likely to reflect an effect of the exposure or merely could be an effect of uncontrolled confounding.

A particularly critical situation with regard to uncontrolled confounding may be found in studies linking cancer incidence or mortality data in a register with census data on occupation. Here occupation is taken as a proxy for some particular exposure. Such broad occupational categories tend to involve a number of individuals within exposure. Due to misclassification, risk ratios from such linkage studies are notoriously low, and may be low enough such that they could occur due to confounding by smoking. As an example, assume the risk ratio for smoking to be 12, eg, for lung cancer as an average over age and smoking categories. With 40% smokers in the particular job category of interest vs 50% as the national average, there would be negative confounding (confounding risk ratio of 0.83). This is sufficient to reduce a true risk ratio for the occupation of 1.20 to unity. A risk ratio of 1.20 for the occupation would occur in a situation where 20% within the occupation are exposed and the true risk ratio for the exposure is 2.0. Such a percentage of workers actually exposed was found, for example, among farmers and forestry workers with regard to phenoxyacid herbicides in Sweden [Hardell et al, 1981].

NEEDS AND RECOMMENDATIONS

There is a general need to consider more formally the potential impact of not controlling for smoking (or other strong risk factors) on study results, especially when the risk estimate for the exposure is relativley low. Quite rigorous and semi-quantitated estimations can be applied in such evaluations according to the methods delineated here. We recommend that reviewers of studies phrase critical comments about missing smoking data in such quantitative terms. To date, criticism of epidemiological studies has usually been only qualitative in character and therefore not always particularly valid.

REFERENCES

Asp S (1984): Confounding by variable smoking habits in different occupational groups. Scand J Work Environ Health 10:325-326.

Axelson O (1978): Aspects on confounding in occupational health epidemiology. Scand J Work Eviron Health 4:85-89.

Axelson O (1985): Dealing with the exposure variable in occupational and environmental epidemiology. Scand J Soc Med 13:147-152.

Axelson O, Johansson B, Axelson T (1984): On the problem of controlling confounding in case-referent studies. Ann Acad Med Singapore: [Suppl] 13:308-311.

Beaumont J, Weiss N (1981): Lung cancer among welders. J Occup Med 23:839-844.

Blair A, Hoar SK, Walrath J (1985): Comparison of crude and smoking-adjusted standardized mortality ratios. J Occup Med 27:881-884.

Brackbill R, Frazier T, Shilling S (1988): Smoking characteristics of U.S. workers, 1978-1980. Am J Ind Med 13:5-41.

Checkoway H (1986): Methods of treatment of exposure data in occupational epidemiology. Med Lav 77:48-73.

Checkoway H, Waldman GT (1985): Assessing the possible extent of confounding in occupational case-referent studies. Scand J Work Environ Health 11:131-133.

Cornfield J, Haenszel W, Hammond EC, Lilienfeld AM, Shimkin MB, Wynder EL (1959): Smoking and lung cancer: Recent evidence and discussion of some questions. J Natl Cancer Inst 22:173-203.

Covey LS, Wynder EL (1981): Smoking habits and occupational status. J Occup Med 23:537-542.

Edling C, Axelson O, Kling H (1983): Diesel exhaust exposure, cardiovascular disease and an approach to allow for smoking. Scand J Work Environ Health 9:68 (Abstract).

Edling C, Axelson O (1984): Risk factors of coronary heart disease among personnel in a bus company. Int Arch Occup Environ Health 54:181-183.

Gail M, Wacholder S, Lubin J (1988): Indirect corrections for confounding under multiplicative and additive risk models Am J Ind Med 13:119-130.

Hardell L, Eriksson M, Lenner P, Lundgren E (1981): Malignant lymphoma and exposure to chemicals, especially organic solvents, chlorophenols and phenoxyacids; a case-control study. Br J Cancer 43:169-176.

IARC (1986): "Tobacco Smoking." IARC Monogr Eval Carcinog Risk Chem Hum Vol 38.

Kleinbaum DG, Kupper LL, Morgenstern H (1985): "Epidemiologic Research. Principles and Quantitative Methods." London: Lifetime Learning Publications (Wadsworth).

Lerchen ML, Samet JM (1986): An assessment of the validity of questionnaire responses provided by a surviving spouse. Am J Epidemiol 123:481-489.

McMichael AJ (1976): Standardized mortality ratios and the "healthy worker effect": Scratching beneath the surface. J Occup Med 18:165-168.

McLaughlin J, Blot W, Mehl ES, Mandel JS (1985): Problems in the use of dead controls in case-control studies I. General results. Am J Epidemiol 121:131-139.

Miettinen OS (1972a): Components of the crude risk ratio, Am J Epidemiol 96:168-172.

Miettinen OS (1972b): Standardization of risk ratios. Am J Epidemiol 96:383-388.

Registrar General (1978): "Occupational Mortality, Series DS No 1: Decennial Supplement for England and Wales 1970-72." London: Her Majesty's Stationery Office.

Rogot E, Reid D (1975): The validity of data from next-of-kin in studies of mortality among migrants. Int J Epidemiol 4(1):51–54.

Siemiatycki J, Wacholder S, Dewar R, Wald L, Bégin D, Richardson L, Rosenman K, Gérin M (1988): Smoking and degree of occupational exposure: Are internal analyses in cohort studies likely to be confounded by smoking status? Am J Ind Med 13:59–69.

Sjoberg A, Brondum F, Lundgren KM, Tejler T, Axelson O (1977): Lungcancerns epidemiologi med hansyn till industriell verksamhet. Socialmed Tidskr 54:183–188.

Steenland K, Beaumont J, Halperin W (1984): Methods of control for smoking in occupational cohort mortality studies. Scand J Work Environ Health 10:143–149.

Stellman S, Boffetta P, Garfinkel L (1988): Smoking habits of 800,000 American men and women in relation to their occupations. Am J Ind Med 13:43–58.

Sterling T, Weinkam J (1976): Smoking characteristics by type of employment. J Occup Med 18:743–754.

Surgeon General of the U.S. (1979) "Smoking and Health." Rockville, Maryland: Department of Health, Education and Welfare.

Surgeon General of the U.S. (1985): "The Health Consequences of Smoking. Cancer and Chronic Lung Disease in the Workplace." Rockville, Maryland: Department of Health and Human Services.

Suta B, Thompson C (1983): Smoking pattern of motor vehicle industry workers and their impact on lung cancer mortality rates. J Occup Med 25:661-667.

Thun MJ, Schnorr TM, Smith AB, Halperin WE, Lemen RA (1985): Mortality among a cohort of US cadmium workers—an update. JNCI 74(2)325–333.

Weinberg G, Kuller L, Redmond C (1982): The relationship between the geographical distribution of lung cancer incidence and cigarette smoking in Allegheny County, Pennsylvania. Am J Epidemiol 115:40–58.

Weinkam J, Sterling T (1987): Changes in smoking characteristics by type of employment from 1970 to 1979/1980. Am J Ind Med 11:539–561.