Applied Occupational and Environmental Hygiene

ISSN: 1047-322X (Print) 1521-0898 (Online) Journal homepage: https://www.tandfonline.com/loi/uaoh20

False Positive Exposure Errors and Low Exposure Prevalence in Community-Based Case-Control Studies

Walter F. Stewart & Adolfo Correa-Villaseñor

To cite this article: Walter F. Stewart & Adolfo Correa-Villaseñor (1991) False Positive Exposure Errors and Low Exposure Prevalence in Community-Based Case-Control Studies, Applied Occupational and Environmental Hygiene, 6:6, 534-540, DOI: 10.1080/1047322X.1991.10387924

To link to this article: https://doi.org/10.1080/1047322X.1991.10387924

False Positive Exposure Errors and Low Exposure Prevalence in Community-Based Case–Control Studies

Walter F. Stewart and Adolfo Correa-Villaseñor
Department of Epidemiology, School of Hygiene and Public Health, The Johns Hopkins University, 615 North Wolfe Street, Baltimore, Maryland 21205

The community-based case–control study (CBCC) offers a potentially efficient strategy for identifying putative causes of disease and has special relevance to the surveillance of occupationally related diseases such as cancer. The primary limitation in using the CBCC design for surveillance purposes is that defining exposure is highly prone to error. This paper quantifies the bias to the odds ratio caused by nondifferential exposure misclassification under a range of conditions common to occupational studies.

When the exposure prevalence is low (< 10%), a condition common to community-based occupational and other case–control studies, and assuming nondifferential misclassification, false positive exposure errors (i.e., nonexposed are classified as exposed) are the principal source of bias to the odds ratio. In contrast, even a relatively large number of false negative exposure errors results in negligible bias.

Knowledge of the major source of bias to the odds ratio offers the means to develop specific and efficient strategies to correct for exposure misclassification errors. Three strategies are recommended to minimize the effect of false positive errors: making use of exposure groups rather than occupational title groups in the analysis; repeat interviews of subjects designated as exposed; and methodological research regarding the manner in which occupational histories are obtained and exposure decisions are made.
Stewart, W.F.; Correa-Villaseñor, A.: False Positive Exposure Errors and Low Exposure Prevalence in Community-Based Case–Control Studies. Appl. Occup. Environ. Hyg. 6:534–540; 1991.

Introduction

The community-based case–control study (CBCC), where details on exposures are obtained from incident cases and controls, is a potentially important population-based surveillance tool.[1,2] The primary weakness of this design is that defining exposure status for putative causes is highly prone to error.

Exposure misclassification errors often cause the exposure odds ratio to be biased toward the null, and they increase the likelihood of a false negative conclusion.[3–9] The magnitude of the bias depends, in part, on specific study conditions such as the observed prevalence of exposure. No single strategy for reducing bias is likely to be optimal and logistically efficient for all possible study conditions. Instead, the problem may be simplified by focusing on the conditions specific to a particular area of study (e.g., CBCC studies focusing on occupational factors) and the range and types of errors that are most likely to arise. In this way, developing realistic and efficient strategies either to remedy the errors or to correct for their effect is more straightforward. This paper focuses primarily on CBCC studies where the exposure prevalence is low, a condition common to occupational and other case–control studies. Given several simplifying assumptions regarding the nature and range of errors likely to occur, analyses suggest that, under nondifferential conditions, false positive errors cause relatively large biases to the odds ratio. The implications with regard to both respondent errors and exposure assessment errors are discussed.

Methods

In case–control studies, the exposure odds ratio for a binary variable is derived as a measure of association between an exposure and disease.[10] The effect of exposure errors in biasing the true odds ratio, ψ, is examined by first defining the true exposure distribution of controls. Let C_{00} designate the true proportion of controls not exposed, and C_{01}, the true proportion of controls exposed. The true distribution of cases by exposure status, C_{1j}, is derived by specifying C_{0j} and ψ (e.g., $C_{11} = \psi \cdot [C_{10} \cdot C_{01}]/C_{00}]$). Exposure misclassification alters the C_{ij}s in Table I.[5,6]

Exposure misclassification can be expressed by a two by two table of conditional probabilities, as shown in Table II, where P_{jk} (j,k = 0,1) is the probability of being classified as exposed (j = 1) or not (j = 0) given a true exposure status k (k = 0,1). P_{00} is often referred to as exposure specificity and P_{11} as exposure sensitivity. By definition,

1047-322X/91/0606-534/$2.00/7

TABLE I. True Distribution of Cases and Controls by Exposure Status

		Case (1)	Control (0)
Exposure	Yes (1)	C_{11}	C_{01}
	No (0)	C_{10}	C_{00}

TABLE II. Distribution of Subjects by True and Observed Exposure Status

		True Exposure (k) Yes (1)	No (0)
Observed Exposure (j)	Yes (1)	P_{11}	P_{10}
	No (0)	P_{01}	P_{00}
		1.0	1.0

when the exposure errors are nondifferential, cases and controls have the same sensitivity and specificity. The observed proportion of nonexposed cases or controls, c_{i0}, is derived as $P_{00} \cdot C_{i0} + P_{01} \cdot C_{i1}$. The observed proportion of exposed cases or controls, c_{i1}, is derived as $P_{11} \cdot C_{i1} + P_{10} \cdot C_{i0}$. The observed exposure odds ratio, ψ, is $(c_{11}c_{00})/(c_{01}c_{10})$; c_{10} and c_{00} are the observed proportion of cases and controls not exposed, respectively, and c_{11} and c_{01} are the observed proportion of cases and controls exposed, respectively.

For the purpose of discussion, bias to the odds ratio, θ, is defined in terms of excess risk. If the true excess risk is $\Psi - 1$, then bias is

$$\theta = (\psi - \Psi)/(\Psi - 1) \quad [\Psi \neq 1] \qquad (1)$$

Bias, θ, has a negative value if Ψ is underestimated and a positive value if it is overestimated.

Assuming nondifferential errors and $\Psi > 1.0$, the relationship between C_{ij}, P_{jk}, and θ is described for a binary exposure variable under a range of conditions common to occupational case–control studies. Since the choice of values for C_{ij} and P_{jk} can affect the nature and magnitude of bias, values thought to be reasonable for occupational studies have been selected.

The observed proportion of controls not exposed, c_{00}, is often between 0.85 and 0.96 in occupational case–control studies; these observed values reflect the sum effect of false positive and false negative exposure errors. Values less than 0.85 are uncommon except when a definition of "ever exposed" is used, where it can be assumed that the number of false positive exposures is likely to be high. Values at or above 0.95 are common when more rigorous definitions of exposure are used. For illustration purposes, values ranging between 0.80 and 0.96 are imputed for C_{00}.

Estimates of the proportion of subjects correctly classified as not exposed, P_{00}, and as exposed, P_{11}, are likely to vary considerably depending on the exposure, the data collected, and the method of assessing exposure. Nonetheless, reasonable estimates of the lower bound for the P_{00} can be obtained. It can be shown that $P_{00} \geq c_{00}$ when exposure classification is equal to or better than chance (see Appendix). That is, even if exposure status is broadly defined and all subjects designated as exposed are, in fact, false positives, the lower bound for P_{00} will be greater than or equal to $1 - c_{01}$. Values between 0.80 and 0.96 are, therefore, considered as reasonable bounds for P_{00}.

The lower bound for P_{11}, the proportion correctly classified as exposed, is likely to be relatively low compared to P_{00}, especially as more rigorous definitions of exposure are used. For the purpose of illustration, a broad range of values from 0.5 to 1.0 is used (i.e., classification based on chance to perfect classification of the exposed).

Nondifferential Errors and Bias to the Odds Ratio

In this section, the bias caused by exposure errors when the exposure variable is dichotomous and $\Psi > 1.0$ is reviewed; this topic has received considerable attention elsewhere.[3–9] In contrast to previous work, the primary focus in this section is on the magnitude of the bias given that the exposure prevalence is low.

The true odds ratio, Ψ, is always underestimated when exposure errors are nondifferential.[3–9] The magnitude of the bias or degree to which Ψ is underestimated is shown in Figure 1 and depends on C_{01} (the proportion of controls exposed), P_{11} (sensitivity), and P_{00} (specificity). For the plots in Figure 1, Ψ is equal to 4.0. Two points are noteworthy. First, when the exposure prevalence is relatively high ($C_{01} > 10\%$), significant reductions in Ψ are caused by a decrease from 100 percent in either P_{00} or P_{11}. Second, as the exposure prevalence decreases, even a very small decrease in P_{00} (e.g., from 100% to 99% in Figure 1) results in a substantial bias, whereas even a very large decrease in P_{11} has a relatively small effect on the bias, especially at the lower levels of exposure prevalence.

What are the practical implications of exposure errors when working with a dichotomous exposure variable? When the prevalence of exposure among controls, C_{01}, is low, misclassifying exposed subjects as not exposed is not critical. The number of subjects with a false negative exposure designation is small relative to the number of true negatives, and the odds of exposure among cases and controls are not appreciably altered. Decreasing the number of false positives and thereby increasing P_{00}, however, is central to developing efficient strategies to reduce bias, even at the cost of a reduction in P_{11}. Efforts directed to reducing the number of false positive exposure designations is logistically efficient in that one need only consider the relatively small proportion of subjects designated as exposed.

Exposure Errors: Sources and Remedies

Exposure errors can be introduced at the time of interviewing or exposure assessment. In this section, the errors introduced at each step are reviewed, and the various strategies to remedy or adjust for such errors (Table III) are presented.

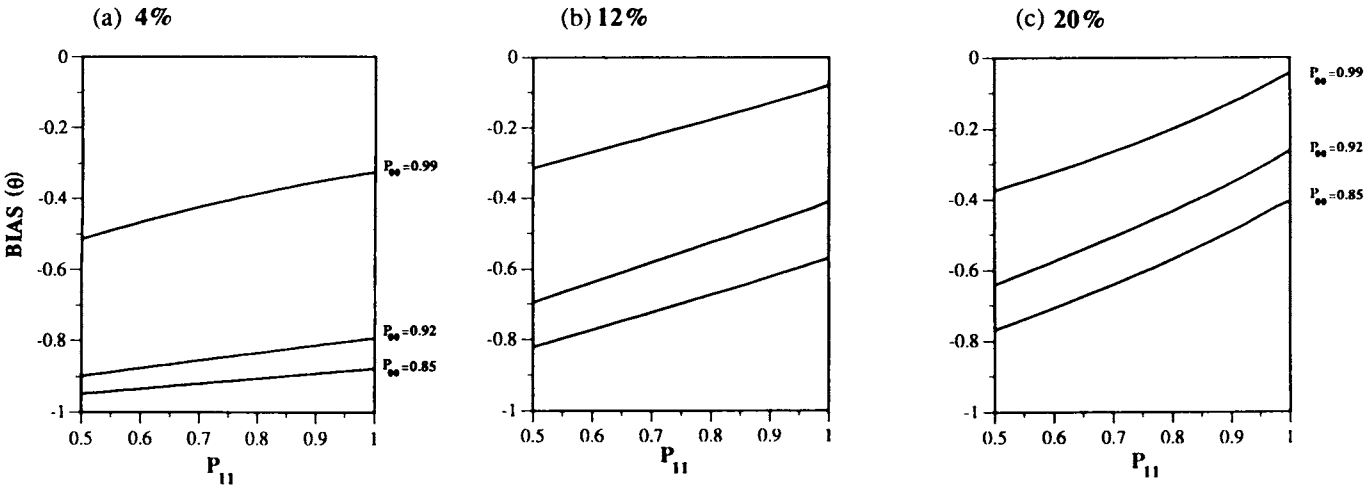

FIGURE 1. The bias to the exposure odds ratio in relation to the exposure prevalence, the proportion correctly classified exposed (P_{11}), and the proportion correctly classified as not exposed (P_{00}), given a true exposure odds ratio of 4.0.

TABLE III. Source and Effect of Exposure Errors in Occupational Case–Control Studies

Source of Error	Cause of Error	Effect on P_{00}	Effect on P_{11}	Strategy to Reduce Bias
Respondent	Omission of an occupation.	None	Lower	• Provoke recall by preparing subject in advance to think about their occupational history.[1,2]
	False report of an occupation.	Lower	None	• Validation of employment for jobs designated as exposed. • Repeat interviews of apparently exposed subjects.
	Dates of employment are in error or the information about the occupation is vague.	Lower	Lower	• Ascertain respondents job-specific certainty of recall. • Obtain details on tasks and duties. • Repeat interviews about tasks and duties of apparently exposed subjects.
Data coding	Coding is not repeatable between coders or within coders.	Lower	Lower	• Explicit coding rules to avoid arbitrary assignment of codes to job titles. • Use an uncertainty code where applicable. • Conduct tests of coding repeatability within and between coders. • Periodic validation of assigned codes. • Do not code occupations. Instead, review the complete occupational history and descriptions as needed.
	Information about the occupation is inadequate to assign a code with certainty.	Lower	Lower	• Obtain details about the employer, job, tasks performed, and equipment, tools, and materials used. • Identify occupational codes assigned with low certainty. • Validation of jobs designated as exposed to confirm the job codes.
	An occupation code is heterogeneous with regard to exposure status.	Lower	None	• Develop ad hoc codes to separate occupational titles into more homogeneous groups.
Exposure assessment	Job title falsely designated as exposed.	Lower	None	• Validate exposure for jobs designated as exposed. • Obtain independent judgments from three or more industrial hygienists. • Obtain industrial hygienist's self-assessment of familiarity with job.
	Job title falsely designated as not exposed.	None	Lower	• Obtain independent judgments from three or more industrial hygienists. • Obtain industrial hygienist's self-assessment of familiarity with job.
	Exposure for the job title is arbitrarily assigned.	Lower	Lower	• Identify exposure decisions made with low certainty. • Obtain industrial hygienist's self-assessment of familiarity with job.

Reporting Errors

Respondents can omit an occupation, falsely report an occupation, or offer information about the occupation that is too vague to make valid exposure decisions. Omitting an occupation can only result in a false negative error, thereby reducing the proportion correctly classified as exposed (P_{11}). This occurs when an occupation associated with the exposure of interest is not reported. As previously noted, the underestimation bias that this type of error causes to the odds ratio is small. On the other hand, if an occupation is not associated with the exposure of interest, exposure status is not affected by whether or not the occupation is reported.

Falsely reporting an occupation can result in a false positive error if the falsely reported occupation is associated with the exposure of interest and the true occupation is not. As a result, P_{00} is reduced and the odds ratio may be seriously underestimated. If the falsely reported occupation is not associated with exposure, then, as noted above, it has no effect on the subject's exposure status as long as the true occupation is not associated with exposure.

Finally, vague or uncertain information offered by the respondent can result in a false positive or false negative error that is introduced either during the coding step or when deciding on the subject's exposure status. Because false positive errors are of greatest concern when exposure prevalence is low, the following strategies may be useful in treating this type of error: validating exposure status; scientific assessment of interview methods for obtaining occupational histories; and obtaining indirect measures of exposure errors.

Positive exposure status for cases and controls judged from reported work histories may possibly be validated by confirming reports with employers,[11] by obtaining details on the place of work and specific occupation(s) reported, and, where feasible, through biologic assessment of the subject. Validation of exposure in this way has several advantages. It is efficient, in that the validation step is only limited to occupations and subjects designated as exposed. Second, P_{00} can be effectively increased without reducing P_{11}.

The potential limitations to validation, given available resources, should be recognized. Confirming that a subject held a reported occupation may violate confidentiality unless the individual is informed beforehand that the employer may be contacted. As an alternative, the employer can be contacted to inquire about specific occupations without reference to the employee. While information can be obtained about the occupation, including information on duties, tasks, and materials, there is no direct validation that the subject was employed in the reported industry or held the reported occupation.

While validation of reported occupations and exposures may be useful, epidemiologists are more or less in the dark in knowing what is practically feasible; relatively little has been invested in understanding the resources available or developing the methodology for this purpose. The extent to which validation is feasible depends on a number of factors including: the proportion of employers who still exist; the number of employees who must be contacted; and whether sufficient resources and information are available, both within the questionnaire and on the part of the employer, to identify the occupation of interest and the associated materials and tasks.

Biologic assessment of exposure is a new and untested frontier, especially as a part of case–control studies. Biologic exposure assessment in case–control studies of cancer is feasible when the case-fatality rate for the tumor is low (e.g., bladder cancer) and when the agent is sequestered in accessible body tissues or fluids for a relatively long period of time. The extent to which biologic assessment can be used as a simple binary marker or as a quantitative marker of exposure is likely to depend on the retention half-life.

Reporting errors can also be reduced by an improvement in interviewing techniques. The greatest limitation here is that relatively little is known about the validity with which occupational information is reported, and even less is known about how interviewing techniques affect the validity of reports. To date, very few validation studies have been reported.[12–17] While a number of new innovations have been introduced in the last decade,[1] little is known about the repeatability of various techniques and their effect on the validity of reporting occupations.

Two types of information are of potential value in adjusting for the effects of reporting errors or, at a minimum, estimating the uncertainty in risk estimates attributable to exposure errors: 1) repeated within subject interviews and 2) "markers" of ability to recall dates, places of employment, and occupations with associated tasks. Repeated interviews are likely to be most valuable in obtaining information reported with poorest validity and reliability. This could include occupational titles from subjects who have held a number of jobs[12,14] or tasks and duties associated with the reported occupation. The latter may be highly relevant for judging exposures associated with a reported occupation. However, with the exception of one study of a select group of current workers,[18] little is known about how validly tasks and duties are reported. Nonetheless, if repeat interviews are conducted, the information on within subject variability in reporting can be incorporated into statistical models to adjust for errors in exposure status, an approach which has received considerable attention in the use of dietary information in epidemiologic studies.[19,20]

Memory should be a topic of central interest to epidemiologists. Short of identifying cumulative biologic markers of exposure, investigators who conduct case–control studies will continue to rely on the memory of their subjects for risk information. Omitting information from subjects identified as having poor or vague recall of historical events is likely to reduce bias, assuming that such recall is nondifferential. Designation of subjects as poor reporters can be done by having the interviewer rate the quality of the subjects' reports. More formal strategies might be developed through the collaboration of epidemiologists and psychologists to explore existing instruments or to

develop standardized measures that can be employed to identify unreliable reporters. A small-scale effort for survey research, in general, has been initiated at the National Center for Health Statistics[21] and should be extended on a much larger scale given the fundamental importance of memory to epidemiologic research.

Reporting errors are the most permanent and the most costly to resolve. In addition, because disease status may affect the quality and reliability of reporting occupational histories, reporting errors are likely to be the primary source of differential misclassification. In contrast, errors in deciding exposure status are less permanent, usually nondifferential (the judgment is made without knowledge of the subject's disease status), and primarily under the control of the investigator. As such, there is considerable latitude for improving P_{00} and P_{11} and making use of the option to increase P_{00} at the cost of decreases in P_{11}.

Exposure Assessment

Case–control analysis of occupational risks is typically done either by forming exposure groups (EG) or by forming occupational title groups (OTG). With EGs, the exposure or agent is explicitly defined which results in a composite of different occupational titles that presumably have exposure features in common. For example, an EG for lead exposure might comprise several occupations including plumbers, pressmen, and selected foundry workers. On the other hand, the exposure or agent is not explicitly defined for an OTG. Instead, risks are estimated for individuals having a common occupation. Each of these methods has strengths and limitations.

The Exposure Group Method

The EG method is typically done in one of two ways: by assigning a numeric code to an occupation on the basis of information reported by the case or control, followed by an exposure decision for the specific occupational code;[22] alternatively, exposure decisions are made at the individual level by reviewing each occupation and associated descriptors reported by a subject. The former strategy is less labor intensive, while the latter makes the best use of all available information in making exposure decisions. A relatively efficient approach is to combine the two procedures. That is, each occupation is numerically coded, followed by a computer identification of the occupational codes that have possible exposure to the agent of interest. The individual occupations (including the associated descriptors) designated as possibly exposed are individually assessed for exposure to the specific agent of interest. At this point, minimizing or controlling for the effects of exposure decision errors is a major concern.

Several strategies have been employed to minimize the occurrence and effect of exposure decision errors. First, at the time of coding, occupations reported with uncertainty can be flagged by the use of a supplemental code to indicate that a job code itself cannot be completely resolved. In this way, the option is preserved to omit uncertain data from the analysis. Second, while it has not been standard practice, measures of coding repeatability should be conducted and reported with analytic results. Finally, periodic checks should be made to confirm the validity of occupational codes.[2]

An alternative to coding occupational titles is simply not to code. Instead, the subject's complete work history, including any and all associated descriptions, is retained intact either in computerized or hard copy form. This form of data is most suitable for the exposure decision step and effectively eliminates coding errors. It does not, however, eliminate the inevitable exposure decision errors due to vague reporting.

Because exposure judgments or decisions themselves are also prone to error, the certainty with which such decisions are made can also be coded or measured. This can be done by asking the "exposure judges" to systematically score the adequacy with which exposure decisions can be made, given the reported information. The score itself may be a composite of both the quality of the information reported and the "judge's" own knowledge of the job. Exposure decisions based on inadequate information can be categorized as "uncertain" at the time of analysis, thereby minimizing both false positive and false negative decisions. Alternatively, uncertainty and within and between "judge" variability can be directly measured. As noted above, this information can be used to directly adjust for the bias due to nondifferential errors introduced at the exposure decision step; however, this does not correct for errors introduced at the time of interview. As a less formal approach, the decisions obtained independently from more than one judge or expert can be used to identify specific occupations that are problematic. This can be followed by a resolution step where a meeting of all judges is convened to discuss conflicting decisions. Uncertainty is explicit when differences of opinion cannot be resolved. Uncertainty can also be resolved by a follow-up interview with the subject to answer questions regarding a previously reported occupation.

In general, it is assumed that a more valid exposure decision will be made as more detailed information is obtained about a subject's occupation. However, little is known about the validity with which information on tasks and duties is reported. Whether or not P_{00} and P_{11} can be improved by using such ancillary information depends on the weight given to it, the duration of time the occupation was held, the number of occupations held, and in general, the reliability with which job duties, tasks, and the like are reported.[12,14]

The strategies described above improve P_{00} and P_{11} independently. P_{00} can also be increased at the cost of reducing P_{11}. This is accomplished by simply using more rigorous criteria in establishing who is exposed. As a result, more subjects are both correctly designated as not exposed (higher P_{00}) and incorrectly designated as exposed (lower P_{11}). An alternative to simply increasing P_{00} at the cost of lowering P_{11} is to categorize subjects separately with designated but uncertain exposure status or to derive the odds ratio with and without such subjects.[22] In this way, both

P_{00} and P_{11} are improved because the nonexposed group is less likely to be contaminated with exposed subjects.

The Occupational Title Group Method

The most common strategy for examining risks in occupational CBCC studies is to derive the odds ratio by occupation or OTGs rather than by exposure status. The exposure decision step is effectively eliminated when using this method, and thus, less time and fewer resources are required. There are, however, a number of limitations inherent to this strategy and, as such, limitations to improving P_{00} and P_{11}.

An OTG is simply a composite of occupational titles that are presumably the same or associated with the same tasks. OTGs do not necessarily designate a specific exposure, and even when defined as such, a single OTG is not all inclusive for a specific exposure; i.e., workers who are members of other OTGs may also have the same exposure. In the analysis, however, each OTG is implicity treated as an exposure group, and some workers who are truly not "exposed" are included in the "exposed" OTG because their occupational title appears to be the same. The observed odds ratio, ψ, is derived for each OTG by treating all other occupations as the reference or, in effect, the nonexposed group. Analyses are first presented for broadly defined OTGs (e.g., construction workers) and then by more specific subgroups (e.g., carpenters) either because there is a specific concern *a priori* or because an excess risk was found for the more broadly defined OTG.

Broadly defined OTGs are heterogeneous. Under such conditions, P_{00}, the proportion correctly classified as not exposed, is always relatively low. P_{11} may be high depending on the proportion of exposed individuals in occupations outside the OTG of interest. P_{00} increases and P_{11} decreases as a more definitive OTG is used (e.g., using a three-digit versus a two-digit SIC/SOC code). While this trade-off is desirable, it is not optimal, especially when compared to that for defining exposure groups. Moreover, there are undesirable consequences of this strategy. Work histories are not closely examined by experts but are more or less reviewed by trained coders, and the exposure decision and speculation about the exposure are made after the fact. Innumerable odds ratios are estimated. Ultimately, the confidence in exposure-specific findings from such studies is low.

Estimating Truth

A strategy which has received little attention is to estimate the range of values for the true odds ratio, Ψ, that could give rise to the data.[23] That is, given an observed odds ratio, ψ, and a range of estimates for the P_{00} and P_{11}, possible values for Ψ are derived. An example of this is shown in Table IV for a dichotomous variable, given an observed odds ratio of 1.4, an observed exposure prevalence (c_{01}) of 10 percent, and values for P_{00} and P_{11} ranging from 0.91 to 1.0 and from 0.5 to 1.0, respectively. In this example, the true odds ratio can range from the observed value of 1.4 to as high as 4.57.

Discussion

The motivation for surveillance of occupational cancer is the early detection of excess cancer cases and identification of their causes. Prospective monitoring of workers or former workers has become more common but is typically restricted to occupational settings with large numbers of workers. For workers employed in smaller companies (< 500 employees), about 85 percent of the United States workforce,[3] the CBCC design is a potentially effective surveillance tool for detecting occupationally related cancers.

In this regard, the CBCC study has several important features as a surveillance tool. First, in contrast to cohort studies, efforts are made to obtain a complete occupational history. In addition, data are obtained on confounders, such as smoking and other lifestyle factors. Second, the specific hazard of interest need not be defined *a priori,* although the occupational history should be adequate to infer specific exposures. Third, CBCC data sets can be used long after data collection is completed to answer emerging etiologic questions. Fourth, preventable cancer cases, i.e., cases caused by known carcinogens which should be controlled in the workplace, can be detected. Fifth, attributable risk can be estimated.[24] Finally, where an excess risk is shown to occur, follow-up investigations of cases with apparent exposure can be conducted to confirm exposure status, to define the exposure environment more precisely, and possibly to develop effective prevention strategies.

The potential application of CBCC for occupational surveillance, however, is limited by two aspects of exposure. First, putative carcinogens are likely to be distributed across a number of different work environments and occupations. Second, condensing an occupational history into a single measure of exposure is complex and prone to error, where false positive exposure errors are the primary source of bias to the odds ratio.

A number of strategies can be employed to control or minimize the effect of these errors. First, as resources permit, the EG method should be used. Second, follow-up interviews with cases and controls should be employed either to measure reporting repeatability or to clarify previously reported occupations. In general, this means that case–control studies must be conducted prospectively (or at most retrospectively for one year with a prospective component) with concurrent exposure assessment such that, as questions arise, the subject can be contacted within

TABLE IV. True Relative Odds Given an Observed Relative Odds of 1.4, an Exposure Prevalence of 10%, and a Range of Values for P_{00} and P_{11}

	P_{00}					
P_{11}	0.91	0.93	0.95	0.97	0.99	1.00
0.50	4.57	2.33	1.85	1.63	1.52	1.47
0.60	4.49	2.29	1.81	1.60	1.49	1.45
0.70	4.43	2.26	1.79	1.58	1.47	1.43
0.80	4.39	2.24	1.77	1.56	1.45	1.42
0.90	4.36	2.22	1.76	1.56	1.45	1.41
1.00	4.34	2.21	1.75	1.55	1.44	1.40

a relatively short interval. Finally, and perhaps most importantly, the methods employed in obtaining occupational histories and in assessing exposures should be scientifically examined. A concerted investment of resources in understanding the methods currently employed in case–control studies and in developing new methods is essential before one can improve on the quality of findings from such studies. The demand on the part of risk assessors for more specific and confident findings requires that our instruments and methods be calibrated if the case–control study for occupational exposure is to be successful in detecting lower level risks.

References

1. Siemiatycki, J.; Day, N.E.; Fabry, J.; Cooper, J.A.: Discovering Carcinogens in the Occupational Environment: A Novel Epidemiologic Approach. Natl. Cancer Inst. 66(2):217–225 (1981).
2. Gerin, M.; Siemiatycki, J.; Kemper, H.; Begin, D.: Obtaining Occupational Exposure Histories in Epidemiologic Case–Control Studies. J. Occup. Med. 27(6):420–426 (1985).
3. Shy, C.M.; Kleinbaum, D.G.; Morgenstern, H.: The Effect of Misclassification of Exposure in Epidemiological Studies of Air Pollution Health Effects. Bull. N.Y. Acad. Med. 54:1155–1156 (1978).
4. Copeland, K.T.; Checkoway, H.; McMichael, A.J.; et al.: Bias Due to Misclassification in the Estimation of Relative Risk. Am. J. Epidemiol. 105:488–495 (1977).
5. Diamond, E.; Lilienfeld, A.M.: Effects of Errors in Classification and Diagnosis in Various Types of Epidemiological Studies. Am. J. Public Health 52(11):37–44 (1962).
6. Bross, I.: Misclassification in 2 × 2 Tables. Biometrics 10:478–486 (1954).
7. Gladen, B.; Rogan, W.J.: Misclassification and the Design of Environmental Studies. Am. J. Epidemiol. 109:607–616 (1979).
8. Greenland, S.: The Effect of Misclassification in the Presence of Covariates. Am. J. Epidemiol. 112:564–569 (1980).
9. Flegal, K.M.; Brownie, C.; Haas, J.D.: The Effects of Exposure Misclassification on Estimates of Relative Risk. Am. J. Epidemiol. 123:736–751 (1986).
10. Cornfield, J.: A Method of Estimating Comparative Rates from Clinical Data. Applications to Cancer of the Lung, Breast, and Cervix. J. Natl. Cancer Inst. 11:1269–1275 (1951).
11. Marsh, G.M.; Enterline, P.E.: A Method for Verifying the Completeness of Cohorts Used in Occupational Mortality Studies. J. Occup. Med. 21(10):665–670 (1979).
12. Baumgarten, M.; Siemiatycki, J.; Gibbs, G.W.: Validity of Work Histories Obtained by Interview for Epidemiologic Purposes. Am. J. Epidemiol. 118:583–591 (1983).
13. Fidler, A.T.; Baker, E.L.; Letz, R.E.: Estimation of Long-Term Exposure to Mixed Solvents from Questionnaire Data: A Tool for Epidemiological Investigations. Br. J. Ind. Med. 44:133–141 (1987).
14. Bond, G.G.; Bodner, K.M.; Sobel, W.; et al.: Validation of Work Histories Obtained from Interviews. Am. J. Epidemiol. 128:343–351 (1988).
15. Rosenberg, C.R.; Mulvihill, M.N.; Fischbein, A.; Blum, S.: An Analysis of the Validity of Self-Reported Occupational Histories Using a Cohort of Workers Exposed to PCBs. Br. J. Ind. Med. 44:702–710 (1987).
16. Stewart, W.F.; Tonascia, J.A.; Matanoski, G.M.: The Validity of Questionnaire-Reported Work History in Live Respondents. J. Occup. Med. 29(10):795–800 (1987).
17. Bourbonnais, R.; Meyer, F.; Gheriault, G.: Validity of Self-Reported Work Histories. Br. J. Ind. Med. 45:29–32 (1988).
18. Weiss, D.J.; Dawas, R.V.: An Objective Validation of Factual Interview Data. J. Appl. Psychol. 44(6):381–385 (1960).
19. Liu, K.; Stamler, J.; Dyer, A.; et al.: Statistical Methods to Assess and Minimize the Role of Intraindividual Variability in Obscuring the Relationship Between Dietary Lipids and Serum Cholesterol. J. Chronic Dis. 31:399–418 (1978).
20. Marshall, J.R.; Graham, S.: Use of Dual Responses to Increase Validity of Case–Control Studies. J. Chronic Dis. 37(2):125–136 (1984).
21. National Center for Health Statistics: Autobiographical Memory for Health-Related Events. Vital and Health Statistics, Series 6, No. 2. U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control, Atlanta, GA (August 1989).
22. Siemiatycki, J.; Richardson, L.; Gerin, M.; et al.: Associations Between Several Sites of Cancer and Nine Organic Dusts: Results from Hypothesis-Generating Case–Control Study in Montreal, 1979–1983. Am. J. Epidemiol. 123:235–249 (1986).
23. Blettner, M.; Wahrendorf, J.: What Does an Observed Relative Risk Convey About Possible Misclassification? Meth. Inform. Med. 23:37–40 (1984).
24. Wilcosky, T.; Wing, S.: The Healthy Worker Effect, Selection of Workers and Work Forces. Scand. J. Work. Environ. Health 13:70–72 (1987).

APPENDIX

Consider a population of size $n_{..}$ with a joint true and observed exposure distribution, as shown in the following two by two table:

Observed Exposure (j)	True Exposure (k): Yes (1)	True Exposure (k): No (0)	
Yes (1)	n_{11}	n_{10}	$n_{1.}$
No (0)	n_{01}	n_{00}	$n_{0.}$
	$n_{.1}$	$n_{.0}$	$n_{..}$

$$\text{Let } n_{..} = n_{11} + n_{10} + n_{01} + n_{00}$$

$$C_{11} = (n_{11} + n_{01})/n_{..}$$

$$c_{11} = (n_{11} + n_{10})/n_{..}$$

$$c_{00} = 1 - c_{11}$$

$$P_{00} = n_{00}/(n_{10} + n_{00})$$

$$\text{Odds(observed} + \mid \text{true} +) = n_{11}/n_{01}$$

$$\text{Odds(observed} + \mid \text{true} -) = n_{10}/n_{00}.$$

If exposure classification is better than chance ($n_{11}/n_{01} > n_{10}/n_{00}$),

$$n_{11}n_{00} > n_{10}n_{01}$$

$$n_{11}n_{00} + n_{10}n_{00} + n_{01}n_{00} + n_{00}^2 > n_{10}n_{01} + n_{01}n_{00} + n_{10}n_{00} + n_{00}^2$$

$$n_{..}n_{00} > (n_{01} + n_{00})(n_{10} + n_{00})$$

$$n_{00}/(n_{10} + n_{00}) > (n_{01} + n_{00})/n_{..}$$

$$n_{00}/(n_{10} + n_{00}) > (n_{..} - [n_{11} + n_{10}])/n_{..}$$

$$P_{00} > 1 - ([n_{11} + n_{10}]/n_{..})$$

$$P_{00} > 1 - c_{11}$$

$$P_{00} > c_{00}$$