Scand J Work Environ Health 1990;16(4):284-288

<https://doi.org/10.5271/sjweh.1783>

Issue date: 01 Aug 1990

Validity of exposure data obtained by questionnaire. Two examples from occupational reproductive studies.

by [Ahlborg GA Jr](#)

Affiliation: Orebro Medical Center Hospital, Department of Occupational Medicine, Sweden.

The following article refers to this text: [2010;36\(3\):231-241](#)

This article in PubMed: www.ncbi.nlm.nih.gov/pubmed/2389136

This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

Validity of exposure data obtained by questionnaire

Two examples from occupational reproductive studies

by Gunnar A Ahlborg, Jr, MD^{1,2}

AHLBORG GA Jr. Validity of exposure data obtained by questionnaire: two examples from occupational reproductive studies. *Scand J Work Environ Health* 1990;16:284-8. Exposure data from self-administered questionnaires were compared with independent information on occupational exposures in two studies of reproductive outcome. Agreement in the case-referent study concerning dry-cleaning work and tetrachloroethylene exposure was good. However, exposure reporting was indicated to be more accurate for the cases than the referents. Correction for misclassification slightly changed the odds ratio from 1.02 to 1.27 for nonspecific exposure and from 0.92 to 0.82 for tetrachloroethylene exposure. Missing information on the latter exposure was more crucial, since adding the employer information for such exposure increased the risk estimate to 1.24. In a prospective follow-up study, exposure information was validated in a sample of the study population. Reporting of heavy lifting appeared to be fairly correct, whereas the underreporting of chemical exposures was a problem. Validation of self-reported exposure data is desirable, and the direction and magnitude of possible misclassification bias should be evaluated in each specific situation.

Key terms: epidemiology, misclassification, occupation, pregnancy, sensitivity, specificity.

Correct assessment and classification of exposure is as essential in occupational studies as it is in other fields of epidemiology (1). Investigations of reproductive hazards may be thought of as particularly problematic in this respect, since the timing of a potentially toxic exposure in relation to reproductive events is often crucial to its effect (2). Several methods for obtaining exposure data are available, each having its advantages and limitations.

Self-administered questionnaires may provide exposure information of acceptable validity when compared with data from clinical interviews or from a review of the information by an industrial hygienist (3, 4). However, if it is possible to obtain "objective" information on exposure from the workplace, at least for a sample of the study population, the bias resulting from any misclassification of exposure can be evaluated and corrected for in a more direct way (5, 6). In this paper two such examples from reproductive studies are presented. The first example relates to a case-referent study of adverse pregnancy outcome in relation to tetrachloroethylene exposure in dry-cleaning work (7). The second example concerns heavy lifting and chemical exposures at work and fetal death in a prospective follow-up study of pregnant women (8).

When exposure data obtained from questionnaires are being compared with those obtained from an independent source, the probabilities of misclassification can be expressed as the sensitivity and specificity of the questions in the questionnaire. Sensitivity is the probability of correct classification among those who are exposed according to the "objective" source, and specificity is the probability of correct classification among those who were truly not exposed (6).

Example from a case-referent study

In a case-referent study of women who worked in a laundry or dry-cleaning shop during pregnancy, information on work conditions and exposures was obtained with a self-administered questionnaire (7). The case group comprised women with pregnancies ending in a hospitalized spontaneous abortion, perinatal death, or the birth of a child with a birthweight of less than 1500 g or a congenital malformation. The outcome of the reference pregnancies was a seemingly healthy infant. All pregnancies were identified in national medical registers and had occurred in the period 1974-1983.

The questionnaire included questions on type of production at the workplace (only laundry, laundry and dry cleaning, only dry cleaning) and which cleaning agents were used. Specifically the woman was asked to state whether or not tetrachloroethylene was used for the dry-cleaning process.

Cases and referents were recruited from two cohorts, one of which had its origin in personnel records from 475 laundries or dry-cleaning shops in Sweden. Along

¹ Örebro Medical Center Hospital, Department of Occupational Medicine, Örebro, Sweden.

² University Hospital, Department of Occupational Health, Linköping, Sweden.

with these records the employers supplied information on type of production, amount of dry cleaning, and which cleaning agents were used in 1973–1983. If tetrachloroethylene had been used, but not during the whole period, the exact time when such use started and ended was given. All information from the employers was obtained before the pregnancies included in the study were identified.

A comparison was made between the information obtained from the employer and that reported in the questionnaire concerning the workplace situation at the time corresponding to the first trimester of each pregnancy studied. The result with regard to the question of whether any dry cleaning was performed at the workplace or not (regardless of the woman's own work tasks) is shown in table 1.

Most of the women who reported that they had been working during the first trimester of pregnancy gave a correct answer, albeit many of them worked at large workplaces where dry cleaning comprised a minor part of the production. Two cases (5 %) and 11 referents (11 %) gave information which was contradictory to that obtained from the employers. Pregnancies from the early part of the study period were not over-represented among these women. If the pregnancies of the women who reported dry cleaning were classified as exposed (excluding those who were unable to answer the question), then the number of exposed cases would be 32, the number of unexposed cases 13, the number of exposed referents 75, and the number of unexposed referents 31. The computed odds ratio (OR) was 1.02 with a 95 % confidence interval (95 % CI) of 0.47–2.20.

If, instead, those who worked during pregnancy and answered yes or no to the question concerning exposure are classified according to the information from the

employers, there are 70 exposed and 36 unexposed referents, while the case numbers remain the same. The odds ratio increases to 1.27 (95 % CI 0.60–2.71). The conclusion is that the slight differential misclassification in this example biases the risk estimate towards unity. Adding the information (obtained from the employers) that was missing in the questionnaire data gives a similar result (OR 1.28, 95 % CI 0.61–2.68).

The more specific question concerning tetrachloroethylene use at the workplace rendered a larger number of uninformative ("don't know") or missing answers (table 1). The proportion of affirmative statements from the employers was about the same for the women who gave such answers as for the women who gave an informative answer. Calculation of the odds ratio for tetrachloroethylene use at the workplace yielded an estimate of 0.92 (95 % CI 0.36–2.33) on the basis of the questionnaire information (14 exposed and 11 unexposed cases, 40 exposed and 29 unexposed referents). Adjustment for the independent information from the employers lowered the estimate to 0.82 (95 % CI 0.32–2.07). The few misclassified pregnancies were found among the referents.

The proportion with possible exposure among the women who did not know if tetrachloroethylene was used at the workplace was larger in the case group (86 %) than in the reference group (73 %) according to the employer information. Adding the exposure information that was missing in the questionnaires resulted in a change of the odds ratio from 0.82 to 1.24 (95 % CI 0.59–2.61). This estimate compared well with that obtained for dry cleaning. The crucial point in this example was thus the missing information, and not erroneous reporting of exposure (eg, the presence of tetrachloroethylene at the workplace) leading to misclassification.

Table 1. Comparison between self-reported data and information from the employers concerning the presence of dry cleaning and the use of tetrachloroethylene in the workplace at the time corresponding to the first trimester of each studied pregnancy (48 cases and 110 referents).

Employer	Questionnaire							
	Cases				Referents			
	Yes	No	Don't know	Not working	Yes	No	Don't know	Not working
Dry cleaning								
Yes	31	1	2	5	67	3	2	8
No	1	3	1	4	8	18	2	2
Sensitivity		0.97				0.96		
Specificity		0.75				0.69		
Tetrachloroethylene ^a								
Yes	14	0	19	5	39	3	30	8
No	0	2	3	4	1	16	11	2
Sensitivity		1.00				0.93		
Specificity		1.00				0.94		

^a Information missing from the employer concerning one of the cases.

Example from a follow-up study

Information was collected on occupational exposures from pregnant women receiving prenatal care in Örebro County from October 1980 to June 1983 with a self-administered questionnaire (8). A total of 3901 working women completed the questionnaire, usually during the first trimester of pregnancy. An industrial hygienist visited the workplaces of 101 of these women shortly after registration in order to validate the exposure data obtained in the questionnaires. After having seen the woman (or one of her colleagues) demonstrate her work duties, and in some cases having consulted the supervisor or safety engineer, the industrial hygienist independently classified exposure in a standardized way corresponding to the questions in the questionnaire. Chemical exposure was recorded in situations where biological uptake was considered plausible.

In the study from which this example was taken, the concern was with nondifferential misclassification, since exposure information was collected before the outcome of pregnancy was known to the women. Two types of exposure have been considered, heavy lifting and contacts with a broad category of chemicals other than solvents.

Heavy lifting

The women were asked if their work involved the lifting of objects weighing 12 kg or more, and, if so, the frequency of the lifting. In order to allow for some uncertainty in the women's estimates, the industrial hygienist classified a weight of 10 kg or more as heavy. Eighty-eight women (23 exposed and 65 unexposed) were accurate in reporting exposure to heavy lifting 10 times per week or more if the classification made by the industrial hygienist is considered an accurate reference. Disagreement was found for eight women who claimed they were exposed and five who said they were unexposed. The sensitivity was 0.82 and the specificity 0.89. If the frequency of lifting is disregarded and only the occurrence of any heavy lifting is considered, the sensitivity becomes 0.79 and the specificity 0.91.

In the original study 1101 women reported exposure to heavy lifting at least 10 times per week. Seventy-seven of these women experienced a fetal death (spontaneous abortion or stillbirth). One hundred and fifty of the 2275 pregnancies among the women who said they were unexposed to heavy lifting terminated in a fetal death. The crude risk ratio for heavy lifting based on these figures was 1.06 (95 % CI 0.82–1.38). The bias from misclassification was evaluated with the formula presented by Flegal et al (6), but the apparent relative risk was entered into the equation to yield the "true" or corrected risk estimate (formula and calculation given in the appendix). With an exposure prevalence of 0.33, the corrected risk estimate would be 1.09 (ie, not much different from the previous estimate).

Chemical exposure

Less specific questions (eg, concerning exposure to "other chemical substances") had much lower sensitivity. The specificity remained high, however, according to the validation procedure employed. In earlier analyses the crude rate ratio for spontaneous abortion + perinatal death had been found to be 1.26 (95 % CI 0.83–1.91) for exposure to chemicals other than solvents. The prevalence of this exposure in the study population was about 0.10. If it is assumed that the sensitivity was 0.30 and the specificity was 0.98 in this case (average figures from several questions on different chemical exposures other than solvents, including one concerning "other chemical substances"), then the corrected risk estimate becomes 1.49. With a somewhat lower specificity of 0.95 or 0.90 the computed "true" relative risk would be 1.85 and 2.78, respectively. This result means, of course, that the apparent relative risk is highly dependent on specificity in this situation, and not that the "true" risk was related to the classification of exposure. If a relatively crude measure (eg, occupation) is used as a proxy for a rare specific exposure, then the risk of large misclassification bias is obvious.

Discussion

In the presented examples some misclassification of exposure was indicated. The direction of the possible misclassification bias was towards unity, which was expected in the prospective study but may be thought of as more noteworthy in the case-referent study. The risk estimates calculated from the questionnaire data were not, or only slightly, elevated. The possible underestimation of the degree of association between exposure and effect (if such existed) due to nondifferential misclassification was not negligible in the case of exposure to chemicals other than solvents in the prospective study. This misclassification was mainly due to an underreporting of exposure. As expected, specificity was crucial in the situation where exposure was rather uncommon (chemical exposure) (5).

The exposure information from the employers in the case-referent study was general and did not specifically apply to the pregnancies considered in the study. For example, some of the women reported that they had not been working during their first trimester of pregnancy. A few of these had actually terminated employment just prior to pregnancy; others were on temporary leave. They were thus unexposed to dry cleaning and tetrachloroethylene, even in the cases in which the employers stated that such use had occurred in the workplaces at the time corresponding to the pregnancies. If only employer information is used nondifferential misclassification may occur, and this misclassification can be of particular concern in studies that focus on acute exposures during a critical period in early pregnancy (2).

The substantial number of uninformative or missing answers from the women with respect to tetrachloroethylene use is a greater problem, especially as this proportion was larger among the cases than among the referents. Most of the cases concerned spontaneous abortions, and it might have been difficult for these women to recall which specific chemical was used during a few months of pregnancy several years ago. Two other reasons for the loss of information on this question may be proposed, and these reasons apply both to the cases and to the referents. The first is that many of the women worked at large workplaces where dry cleaning was a minor part of the total production and only engaged a few workers (often men). Another explanation would be that female laundry and dry-cleaning workers usually have a low educational level and include some immigrants with limited knowledge of Swedish.

The self-reported exposure data obtained by questionnaire compared relatively well with information obtained from "objective" sources in these two examples, but some effect of the correction for possible misclassification was also illustrated. It is important to keep in mind that the information used as the "truth" cannot be claimed to represent the exact truth, but may be used to indicate the direction and magnitude of the misclassification bias.

Acknowledgments

The author thanks Dr L Bodin for his statistical advice, and professors C Hogstedt and O Axelson for

their critical comments on the manuscript.

The original studies were financially supported by the Örebro County Council and the Swedish Work Environment Fund.

References

1. Axelson O. Dealing with the exposure variable in occupational and environmental epidemiology. *Scand J Soc Med* 1983;13:147—52.
2. Lemasters GK, Selevan SG. Use of exposure data in occupational reproductive studies. *Scand J Work Environ Health* 1984;10:1—6.
3. Rosenstock L, Legerfo J, Heyer NJ, Carter WB. Development and validation of a self-administered occupational health history questionnaire. *J Occup Med* 1984; 26:50—4.
4. Eskenazi B, Pearson K. Validation of a self-administered questionnaire for assessing occupational and environmental exposures of pregnant women. *Am J Epidemiol* 1988;128:1117—29.
5. Copeland KT, Checkoway H, McMichael AJ, Holbrook RH. Bias due to misclassification in the estimation of relative risk. *Am J Epidemiol* 1977;105:488—95.
6. Flegal KM, Brownie C, Haas JO. The effects of exposure misclassification on estimates of relative risk. *Am J Epidemiol* 1986;123:736—51.
7. Ahlberg G Jr. Pregnancy outcome among women working in laundries and dry-cleaning shops using tetrachloroethylene. *Am J Ind Med* (in press).
8. Ahlberg G Jr, Hogstedt C, Bodin L, Bárány S. Pregnancy outcome among working women. *Scand J Work Environ Health* 1989;15:227—33.

Appendix

Calculation of bias from the misclassification of heavy lifting

Formula 1 according to Flegal et al (1):

$$R^* = \frac{[URP(E) + (1 - V)P(\bar{E})] \times [1 - V]P(E) + VP(\bar{E})}{[UP(E) + (1 - V)P(\bar{E})] \times [(1 - U)RP(E) + VP(\bar{E})]}, \text{ where}$$

R^* = apparent relative risk

U = sensitivity

V = specificity

$P(E)$ = prevalence of exposure in the population

$P(\bar{E})$ = complement of $P(E)$

R = true relative risk.

The formula may be translated as:

$$R = \frac{(1 - V)P(\bar{E})[(1 - U)P(E) + VP(\bar{E})] - R^*VP(\bar{E})[UP(E) + (1 - V)P(\bar{E})]}{R^*(1 - U)P(E)[UP(E) + (1 - V)P(\bar{E})] - UP(E)[(1 - U)P(E) + VP(\bar{E})]} \quad (\text{continued})$$

In the example on page 10 the equation becomes:

$$R = \frac{0.11 \times 0.67 (0.18 \times 0.33 + 0.89 \times 0.67) - 1.06 \times 0.89 \times 0.67 (0.82 \times 0.33 + 0.11 \times 0.67)}{1.06 \times 0.18 \times 0.33 (0.82 \times 0.33 + 0.11 \times 0.67) - 0.82 \times 0.33 (0.18 \times 0.33 + 0.89 \times 0.67)}$$

Computing R yields 1.09.

Reference

1. Flegal KM, Brownie C, Haas JO. The effects of exposure misclassification on estimates of relative risk. *Am J Epidemiol* 1986;123:736—51.

Received for publication: 11 December 1989