Scand J Work Environ Health 1994;20(3):200-205

<https://doi.org/10.5271/sjweh.1408>

Issue date: 01 Jun 1994

Effect of incomplete exposure assessment on epidemiologic dose-response analyses.

by [Loomis DP](#), [Savitz DA](#)

Affiliation: Department of Epidemiology, School of Public Health, University of North Carolina, Chapel Hill 27599-7400.

This article in PubMed: www.ncbi.nlm.nih.gov/pubmed/7973492

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

Effect of incomplete exposure assessment on epidemiologic dose-response analyses

by Dana P Loomis, PhD, David A Savitz, PhD¹

LOOMIS DP, SAVITZ DA. Effect of incomplete exposure assessment on epidemiologic dose-response analyses. *Scand J Work Environ Health* 1994;20:200-5.

OBJECTIVES — When potentially hazardous agents have multiple environmental sources, failure to include all exposure sources can constitute a type of measurement error. In addition, the effects of exposure from one source can also be confounded by exposure to other sources of the same agent. In this study clarification of these concepts is sought, and the direction and magnitude of the resulting bias in epidemiologic measures of association are examined.

METHODS — The bias in dose-response functions when the exposure data omit some sources of the agent was estimated with linear and log-linear models to compute risk differences and risk ratios under different assumptions about the magnitude and correlation of exposures from measured and unmeasured sources.

RESULTS — With unmeasured exposure of constant magnitude, there is no bias when a measure of association of the appropriate form (difference measures for additive dose-response processes, ratios for multiplicative ones) is selected. When the magnitude of unmeasured exposure varies, the result is nondifferential measurement error that can bias observed dose-response relations upward or downward, depending on the pattern of measurement error and the measure of association.

CONCLUSIONS — Failure to measure all sources of exposure to an agent and account for them in the analysis can bias the results of epidemiologic studies. When it is not feasible to measure all exposure sources, the magnitude of bias can be predicted by estimating the distribution of omitted exposures from external data or substudies. Sensitivity analyses are particularly useful for estimating the direction and magnitude of potential bias from incomplete exposure assessment.

KEY TERMS — epidemiologic methods, exposure assessment, measurement error, misclassification.

Many hazardous agents to which people are exposed have more than one source in the environment. Lead, for example, reaches people through food, air, and water, and carbon monoxide dose is a function of exposure to motor vehicle exhaust plus personal and environmental tobacco smoke. Some agents, such as benzene, radon, and formaldehyde, can be encountered in both workplaces and residential environments.

Epidemiologic studies nevertheless frequently examine only one among many sources of exposure to the agent because the information on the contributions of other sources is insufficient or difficult to obtain. Occupational studies typically consider only work exposures, although cumulative residential exposures to some agents, like radon, can equal exposures acquired on the job (1).

Exposure to electromagnetic fields provides a useful illustration. This agent is encountered in both

workplaces and residential environments, and within each of these areas there are many discrete sources of exposure, including building wiring, household appliances, power lines, video display terminals, and electrical machinery, of which none is clearly predominant over the others. Yet among the dozens of epidemiologic studies conducted on electromagnetic fields, few have simultaneously considered appliances and other residential exposures, and none have considered both occupational and nonoccupational exposures (2, 3).

Epidemiologic studies whose goal is to quantify dose-response relations between specific agents and health outcomes require accurate measurements of exposure. When the agent itself is of interest, without regard to the source of exposure, epidemiologic studies should clearly incorporate all sources of exposure to the agent in order to yield accurate results. Constructing dose-response functions using some but not all sources constitutes measurement error.

Epidemiologic studies may also seek to isolate the effects of a particular source of exposure to an agent. Regulation and individual decision making frequently focus on the benefits of changing a single source of exposure; occupational health standards, in particular, consider only workplace exposures. If expo-

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

Reprint requests to: Dr D Loomis, Department of Epidemiology, CB 7400, McGavran-Greenberg Hall, University of North Carolina, Chapel Hill, NC 27599-7400, USA

sure from the source of concern is measured accurately, omission of other sources does not, in principle, constitute measurement error. Nevertheless, exposures from sources other than the one of primary interest are a potential cause of the same health effects. In such situations, exposures from such secondary sources can act as confounders of the effect of the source of primary concern. The potential importance of the omitted sources of exposure clearly increases when their magnitude is large, relative to exposures from the primary source.

The preceding goals of measuring the overall effect of exposure and measuring the effect of a single source are often blended in practice. This kind of mixture is particularly common in occupational studies, in which workers' experience is observed to develop dose-response data for the agent in general, and not only for workplace exposures.

Several types of measurement error can be involved when the assessment of exposure is incomplete. The simplest type is systematic omission of an exposure component of constant magnitude, such that the true level of exposure is uniformly understated for every person or group in a study. Such an omission might occur, for example, if exposures to cosmic radiation were omitted from estimates of exposure to ionizing radiation from other sources.

The other type of measurement error which has been considered in this study is the systematic omission of a component of exposure whose magnitude varies among persons or groups with different measured exposures and which therefore results in an underestimation of true exposure by a nonconstant amount. This type of error might occur, for example, if, in addition to occupational exposure to formaldehyde, workers also receive a variable contribution from residential formaldehyde exposure. The magnitude of this type of measurement error can be positively or negatively correlated with true exposure, or independent of it.

In this paper, we examine the consequences of failure to account for multiple sources of exposure to a single agent. Using simple dose-response models and hypothetical examples from studies of exposure to electromagnetic fields, we have first considered the problem of empirically estimating an unknown dose-response function when the available exposure data omit some sources and then examined the potential for multiple exposure sources to act as confounders when the independent effect of a single source is of interest. Although the principles are simple, these issues have not to our knowledge been directly addressed in the epidemiologic literature.

Methods

Risk models

Two simple mathematical models are frequently used to describe the exposure-disease relationship in a

population exposed to a specific, disease-causing agent. In linear form, population risk is modeled as $R = r_0 + \beta X$, where r_0 is the background risk, X is the dose of the agent, and β is a coefficient expressing the increase in risk per unit of dose. The log-linear model, $R = \exp[r_0 + \beta X]$, is similar, except that r_0 and β are on the log-risk scale.

Measures of association

The effect of agents on disease occurrence is often assessed by estimating the parameters of linear or log-linear risk models from population data. The risk difference (RD) and risk ratio (RR) are common measures of association in epidemiologic studies. Under the linear dose-response model, $RD = \beta X$ and $RR = 1 + \beta X/r_0$ for the comparison of a population with exposure X to an unexposed referent population. The same comparison under a log-linear dose-response model yields $RR = e^{\beta X}$ and $RD = e^{r_0}(e^{\beta X} - 1)$.

Components of exposure

When the measurement of exposure is incomplete, a person's true exposure, x , can be viewed as $x_m + x_u$, the sum of measured and unmeasured components of exposure, respectively. Since there may be multiple sources of unmeasured exposures, the average true exposure of a population with measured exposure X_m can be written as $X_m + X_u$, where $X_u = \sum_i P_i X_{ui} / \sum_i P_i$, with P_i as the prevalence of unmeasured exposure source i , and X_{ui} as its average magnitude.

In the slightly modified form $X_i + \sum_j P_j X_{nj} / \sum_j P_j$, where X_{nj} is the intensity of each of n secondary exposure sources and P_j is its prevalence, this expression can also describe the situation in which multiple sources of exposure exist, but one is considered to be of primary interest. In such cases, variations in the prevalence and magnitude of the secondary sources in relation to the level of the primary exposure are of concern, with confounding being present when P_i or X_{nj} vary across levels of the primary source, X_i .

The preceding expressions treat exposures as continuous variables, and this treatment allows them to be conveniently summarized by population means. However, we have focused subsequent discussion on examples using discrete, ordinal exposure data; they are more typical of epidemiologic analyses and make the illustrations clearer.

Effect of unmeasured exposure

In practice, measurements of exposure are often used to estimate dose, and these exposures themselves may be measured with error, as when some sources of the agent are not taken into account. In these situations it is of interest to know how severely observed risk coefficients or related effect measures like the risk difference and risk ratio are biased relative

to the values that would be obtained with complete, accurate exposure data.

To examine the impact of incomplete exposure measurement on observed dose-response relations, we used the linear and log-linear risk models to calculate measures of association based on varying assumptions about the magnitude and relationship of measured and unmeasured exposures. The following two types of systematic measurement error resulting from the omission of some sources of an agent were considered: (i) underestimation of exposure by a constant amount and (ii) underestimation by an amount which varies with the measured level of exposure and which may be positively or negatively correlated with it or independent of it. We have assumed, in addition, that the agent causes disease, that there is no random error in measuring exposure nor any bias from sources other than incomplete treatment of exposure, and that exposures from all sources act quantitatively and qualitatively in the same way.

Results

Constant exposure component omitted

The consequences of measuring only one among several sources of exposure to an agent are easily predicted when the quantity of unmeasured exposure is constant across the study population and simply adds an increment to the average dose. If the true dose-response function is assumed to be linear, it can be shown algebraically that the observed risk difference is equal to βX , the true risk difference. Underestimation of exposure by a constant amount therefore has no effect on risk differences observed for measured exposures. However, risk ratios are consistently underestimated in this situation. (See the appendix.) The results are reversed for a log-linear dose-response relation in that observed risk ratios equal

the true risk ratio $e^{\beta X}$, whereas risk difference measures overstate the excess risk at any level of measured exposure. (See the appendix.) Since the coefficient β represents the change in risk or relative risk per unit of exposure, the same principles apply whether exposure is considered on a continuous or categorical scale.

Now consider a hypothetical study of leukemia among electrical workers (table 1). Workers with measured occupational exposures to magnetic fields of 1, 2, 4, 8, or 16 mG, as shown in the first column of table 1, have cumulative exposures ranging from 2 to 29 mG-years, on the assumption of 40 h of exposure per week in the measured field for 10 years. If all of the workers were also exposed over the same interval to average background fields of 1 mG during the 128 h they spent away from work each week, the occupational exposure data would underestimate their total exposure by an equal amount (about 7 mG-years) for workers in each of the five exposure groups.

Comparison of the risk ratios and risk differences in table 1 shows that uniform underestimation of exposure under the log-linear risk model causes the risk difference for each exposure category to be overstated and steepens the dose-response curve it describes (table 1). The risk ratios and the slope they describe are measured correctly, however. The observed risk coefficient estimated from the data in continuous, rather than categorized, form is also unbiased (table 1).

Nonconstant exposure component omitted

Failure to measure exposures from all sources has more complex effects when the intensity or prevalence of unmeasured exposures is not constant but varies relative to the measured exposure. The direction and magnitude of the resulting bias then depend on the magnitude of the unmeasured exposure and its relationship to the measured component (appendix). In general, unmeasured exposures whose magnitude is positively correlated with measured exposure tend to amplify dose-response relationships, while negatively correlated unmeasured exposures diminish them.

As an illustration of this type of bias, consider again the hypothetical electrical workers in the previous example. The intensity of measured and unmeasured exposures remains as before, but this time the duration and intensity of occupational exposure are inversely correlated, and the workers are followed for 20 years (table 2). With this pattern of employment, workers with the most intense occupational exposures tend to remain on the job for the shortest time and consequently have the largest component of unmeasured cumulative exposure. Risk differences in this situation are consistently overstated at every level of exposure (table 2). However, this particular pattern of unmeasured exposure

Table 1. Example of the effect of constant 1-mG unmeasured exposure on a dose-response relation observed in a hypothetical study of leukemia among workers exposed to magnetic fields (log-linear risk model). (RD = risk difference, RR = risk ratio)

Measured occupational exposure intensity (mG)	Correct association with measured exposure ^a		Apparent association with misclassified exposure ^b	
	RD	RR	RD	RR
1.0	0.000	1.0	0.000	1.0
2.0	0.001	1.2	0.001	1.2
4.0	0.003	1.6	0.004	1.6
8.0	0.008	3.0	0.013	3.0
16.0	0.039	10.5	0.065	10.5

^a Apparent risk = $\exp[-5.66 + 0.066 \text{ mG} \cdot \text{y}^{-1} (X_m + X_u)]$, with measured exposure intensity in milligauss converted to cumulative exposure in milligauss-years, on the assumption of 10 years' exposure for 40 h a week at constant levels.

^b True risk = $\exp[-5.66 + 0.066 \text{ mG} \cdot \text{y}^{-1} X_m]$, cumulative exposure measured as in footnote a.

results in negligible overestimation of risk ratios in the three intermediate exposure groups, with a more substantial overestimate in the high-exposure category (table 2).

Confounding by other exposure sources

If an agent causes disease and there are several correlated sources of that agent, then each source satisfies the definition of a confounder for the effects of the others. Thus the effects of exposure from one source considered in isolation may be confounded by exposures received from other sources (which may be measured or unmeasured). Where the focus of a study is the effect of exposure from a specified source, exposures to the same agent are conceptually identical to confounding exposures to other disease-causing agents.

A hypothetical study of childhood cancer in relation to exposure to magnetic fields from electric blankets during gestation provides a useful illustration of this problem (table 3). Suppose that high exposures from electric blankets are correlated with high ambient magnetic field exposures from other residential sources because geographic conditions that encourage extended use of electric blankets through much of the year also lead to high use of electric heat, lighting, and the like. In the example, estimated exposures from blankets and other sources (4) are dichotomized into high and low categories, the prevalence of high ambient exposure being assumed to be 80% among those with high blanket exposure and 20% among those with low blanket exposure.

Considering only exposure from electric blankets — ignoring exposures from other residential sources — is equivalent to evaluating the crude risk ratio for high versus low blanket exposure, which has a value of 2.7 under the log-linear risk model (table 3). However, the risk ratio for the same comparison stratified by the level of exposure to ambient magnetic fields is 2.0 in each stratum of ambient exposure (table 3). Since the stratum-specific ratios are identical, a risk ratio for high versus low blanket exposure adjusted for other residential exposures would also equal 2.0. Thus, like other confounders (5), exposures from different sources produce differences between the crude and adjusted estimates. The risk ratio for blanket exposure adjusted for other residential exposures is the more valid estimate of the independent effect of magnetic fields from electric blankets. Evaluation of the health consequences of modifying exposures from electric blankets would be distorted by failure to consider ambient magnetic fields, just as it would by confounding by any other factor.

Discussion

We have shown that failure to account fully for exposure to the agent of interest can challenge the

Table 2. Example of the effect of 1.0-mG unmeasured exposure with varying duration of measured exposure on a dose-response relation observed in a hypothetical study of leukemia among workers exposed to magnetic fields (log-linear risk model, estimation of risk as in table 1). (RD = risk difference, RR = risk ratio)

Occupational exposure intensity (mG)	Duration in job (years)	Correct association with measured exposure ^a		Apparent association with misclassified exposure ^a	
		RD	RR	RD	RR
1	20	0.000	1.0	0.000	1.0
2	18	0.001	1.3	0.004	1.3
4	15	0.004	1.9	0.014	2.0
8	12	0.011	3.3	0.036	3.7
16	9	0.029	7.0	0.096	8.3

^a Correct and apparent associations estimated with the log-linear risk model as in table 1, with measured exposure converted to cumulative exposure on the assumption of 20 years of total exposure with 40 h of exposure per week at the indicated level and duration and the remainder of exposure at 1 mG.

Table 3. Example of confounding by multiple sources of exposure to the same agent in a hypothetical study of childhood cancer in relation to prenatal exposure to magnetic fields (log-linear risk model). (RR = relative risk)

Exposure from ambient sources ^a	Exposure from blankets ^a	Crude RR for blanket exposure ^b	RR for blanket exposure by stratum of ambient exposure ^c
Low	Low	2.7	2.0
	High		
High	Low		2.0
	High		

^a Numerical values for blanket and ambient field exposures estimated from reference 4.

^b Risk = $\exp[-11.5 + 0.69 \text{ d}]$.

^c Reference group is low blanket exposure within stratum of ambient exposure.

quantitative interpretation of exposure-disease associations obtained from epidemiologic data.

When the true exposure is understated by a constant amount, bias can be avoided by specifying the correct analytical model — ratio measures for log-linear dose-response relationships and difference measures for linear dose-response relationships. Unfortunately, the appropriate model may not always be apparent. The biological action of toxic agents is rarely well enough understood to dictate the correct model uniquely (6), and measurement error makes it more difficult to distinguish empirically between several alternatives.

When the magnitude or prevalence of the unmeasured exposure components varies depending on the magnitude of measured exposure, the effect on the observed results is not as easily predicted, and selecting an appropriate analytic model will not necessarily remove the bias due to measurement error. This observation adds to the evidence from recent papers (7–9) that it is not universally true that non-

differential error in measuring exposure attenuates exposure-disease associations (10).

The tendency for purely random nondifferential exposure measurement error to attenuate observed exposure-disease associations is well known (11). However, our analysis of situations in which the magnitude of nondifferential measurement error is associated with measured exposure indicates that the slope of dose-response relations can be biased upward or downward, even though the measurement error is independent of individual disease status and always understates the true level of exposure. Although standard texts note that coefficients in regression models can be biased in a similar way by correlated errors in measuring predictor variables (12), the consequences for epidemiologic research do not appear to be widely appreciated.

We have also shown that exposures received from multiple sources of a single agent may act as confounders of an association that has been chosen as the one of primary interest. Exposures received at different times would act analogously. For example, given a focus on exposure within particular "windows" of age (13), the effect may be confounded by earlier or later exposures that are correlated with the one of interest (14).

Our analysis suggests some strategies for eliminating or reducing the potential bias from incomplete assessment and analysis of exposure. When multiple exposure sources are known or suspected, the preferred solution is obviously to identify and measure all of the sources. Some biological markers, blood lead, for example, provide a measure of exposure from all sources combined, although they also make it impossible to separate their effects for assigning risk to specific sources of exposure, or benefits to their elimination.

When data describing exposure from all sources are obtained, they must also be analyzed with methods that are appropriate to the scientific goals. If the principal goal is to characterize a dose-response relation for a particular agent-disease combination, it can be met the most directly by summing measured exposures from all sources to estimate total exposure. In studies that seek to isolate the independent effect of a particular exposure source, control for potential confounding by exposures from other sources is necessary.

When direct information cannot be obtained for all exposure sources, external data or a smaller substudy may help to characterize the relative contributions of the measured and unmeasured sources and also the relationship between them. If the exposure sources are uncorrelated, then, ideally, knowledge of the underlying dose-response relation would allow selection of an analytic model not likely to be biased by the unmeasured source. At a minimum, estimates of the magnitude of exposures from various sources and

their relationships can be used to estimate the bias from studying the measured source alone. A sensitivity analysis that considers a range of possibilities would place bounds on the uncertainty in measuring the dose-response relation in the population. The principles developed in this study provide the basis for more systematic speculation about potential bias from incompletely evaluated exposure.

Acknowledgments

An earlier version of this work was presented at the American Industrial Hygiene Conference, Salt Lake City, Utah, in May 1991. We are grateful to Drs B Armstrong and S Greenland for their provocative conversations, which encouraged us to investigate this issue. We also thank Drs H Brenner, H Kromhout, I Hertz-Picciotto, and A Rowland for their constructive reviews.

References

1. Committee on the Biological Effects of Ionizing Radiations. Health risks of radon and other internally deposited alpha-emitters (BEIR IV). Washington, DC: National Academy Press, 1988.
2. Ahlbom A. A review of the epidemiologic literature on magnetic fields and cancer. *Scand J Work Environ Health* 1988;14:337-43.
3. Savitz DA, Calle EE. Leukemia and occupational exposure to electromagnetic fields: review of epidemiologic surveys. *J Occup Med* 1987;29:47-51.
4. Preston-Martin S, Peters JM, Yu MC, Garabrant DH, Bowman JD. Myelogenous leukemia and electric blanket use. *Bioelectromagnetics* 1988;9:207-13.
5. Greenland S, Robins JM. Identifiability, exchangeability, and epidemiological confounding. *Int J Epidemiol* 1986;15:412-8.
6. Siemiatycki J, Thomas DC. Biological models and statistical interactions: an example from multistage carcinogenesis. *Int J Epidemiol* 1981;10:383-7.
7. Brenner H, Savitz DA, Jöckel K-H, Greenland S. Effects of nondifferential exposure misclassification in ecologic studies. *Am J Epidemiol* 1992;135:85-95.
8. Dosemeci M, Wacholder S, Lubin JH. Does nondifferential misclassification always bias a true effect toward the null value? *Am J Epidemiol* 1990;132:746-8.
9. Flegal KM, Keyl PM, Nieto FJ. Differential misclassification arising from nondifferential errors in exposure measurement. *Am J Epidemiol* 1991;134:1233-44.
10. Rothman KJ. *Modern epidemiology*. Boston, MA: Little, Brown & Co, 1986:86.
11. Armstrong BG. The effects of measurement error on relative risk regressions. *Am J Epidemiol* 1990;132:1176-84.
12. Snedecor GW, Cochran WG. *Statistical methods*. Ames, IA: Iowa State University Press, 1980:172.
13. Checkoway H, Pearce N, Hickey JLS, Dement JM. Latency analysis in occupational epidemiology. *Arch Environ Health* 1990;45:95-100.
14. Pearce N. Multistage modelling of lung cancer mortality in asbestos textile workers. *Int J Epidemiol* 1988;17:747-52.

Appendix

Effect of unmeasured exposure on measures of association

Using the notation given in the text, let $R(X_m)$ be the risk due to measured exposure alone and $RD(X_m)$ and $RR(X_m)$ be the corresponding measures of the excess risk difference and risk ratio per unit of exposure (or dose). When all exposure is measured, $RD(X_m) = \beta X_m$ under the linear model and $RD(X_m) = e^{r_0}(e^{\beta X_m} - 1)$ under the log-linear model; $RR(X_m) = 1 + \beta X_m / r_0$ under the linear model and $e^{\beta X_m}$ under the log-linear model. These relationships apply whether X is expressed in continuous, ordinal, or binary form, since β represents the increment in risk or relative risk per unit of exposure.

When some sources of exposure are not measured, the observed risk of disease in an exposed population, R^* , equals $r_0 + \beta X_m + \beta X_u$ under the linear risk model and $R^* = \exp[r_0 + \beta X_m + \beta X_u]$ under the log-linear risk model, where $X_u = \sum_i P_i X_{ui} / \sum_i P_i$.

Let $RD(X_m)^*$ and $RR(X_m)^*$, respectively, indicate difference and ratio measures expressing the appar-

ent association of disease with measured exposure, when some unmeasured exposure is also present. If X_u is a constant, then $RD(X_m)^* = (r_0 + \beta X_m + \beta X_u) - (r_0 + \beta X_u) = RD(X_m)$ under the linear model, but, under the log-linear model $RD(X_m)^* = e^{r_0} e^{\beta X_u} (e^{\beta X_m} - 1) > RD(X_m)$. For the ratio measure of association, $RR(\beta X_m)^* = (r_0 + \beta X_m + \beta X_u) / (r_0 + \beta X_u) = 1 + \beta X_m / (r_0 + \beta X_u) < RR(\beta X_m)$ under the linear model, whereas under the log-linear model $RR(\beta X_m)^* = \exp[(r_0 + \beta X_m + \beta X_u) - (r_0 + \beta X_u)] = RR(\beta X_m)$.

The preceding expressions can be expanded if X_u is not constant (ie, if it has more than one value). For example, if X_m has levels 0 and 1, respectively, associated with unmeasured exposure of X_{u0} and X_{u1} , then under the linear model $RD(X_m)^* = \beta X_m + \beta(X_{u1} - X_{u0})$. $RD(X_m)^*$ can then be greater than, less than, or equal to $RD(X_m)$, depending on the magnitude of X_{u1} relative to X_{u0} . Analogous expressions can be developed for other measures of association.

Received for publication: 30 July 1993