# UCLA
## UCLA Previously Published Works

**Title**
Basic Problems in Interaction Assessment

**Permalink**
https://escholarship.org/uc/item/7sf91363

**Journal**
Environmental Health Perspectives, 101(Suppl 4)

**ISSN**
1542-4359

**Author**
Greenland, Sander

**Publication Date**
1993-12-01

**DOI**
10.2307/3431660

Peer reviewed

# Basic Problems in Interaction Assessment

## Sander Greenland

Department of Epidemiology, UCLA School of Public Health, Los Angeles, CA 90024-1772

This paper reviews problems with the definition and estimation of interactions in epidemiologic studies. Methods for modeling interactions and dose–response also are reviewed, and references to more detailed literature are provided. Concepts are illustrated in the context of evaluating the joint effects of household radon exposure and environmental tobacco smoke. — Environ Health Perspect 101(Suppl 4):59–66 (1993).

Key Words: Epidemiologic methods, interaction, relative risk, risk assessment

## Introduction

In any study of the health effects of exposure mixtures, it is natural to ask whether or not observed effects are due to interactions of the mixtures' components; for example, one may inquire whether or not the effect of one component is modified by the effect of another (effect modification). Four major problems in addressing such a question are: *a*) The term "interaction" has no single definition but requires precise definition in order to be studied; *b*) even when it is precisely defined, there is often little study power to detect interaction; *c*) interactions are inevitably confounded with dose–response and latency relationships; and *d*) measurement errors, even if independent and nondifferential (random), may severely distort interaction assessment. This paper reviews these four problems and provides references to detailed literature on the problems. Definitions of the central concepts are reviewed first in order to provide a basis for precise problem statements. Next, the problems are described and illustrated in the context of evaluating effects of household radon exposure and environmental tobacco smoke (passive smoking). Finally, methods for dealing with the problems are reviewed.

Issues concerning mechanisms of interaction are not addressed here. As recently discussed by Thompson (*1*), epidemiologic data are limited inherently in their ability to discriminate among such mechanisms, because different mechanisms may predict identical patterns of disease. This problem is a logical one and persists even if the problems discussed here are eliminated.

## Definitions

### Main Effects and Causal Effects

A source of ambiguity in the study of interactions (and indeed in the study of any effects) is the existence of multiple definitions of the term effect. Two major definitions exist. Ironically, both seem to have developed from the pioneering work on experimental design conducted by Fisher, Neyman, and others during the period between the first and second world wars.

The first definition, the parametric definition, is by far the most common today: An effect is a coefficient of a study exposure in a generalized linear model for the outcome of interest. [A generalized linear model is simply a linear model for some transformation of the expected outcome (*2*).] As an example, consider a log-linear (multiplicative) model for the rate $R$ (in cases per person-year) of lung cancer in a cohort of married nonsmokers, given a certain exposure to spousal tobacco smoke $x$ and radon level $z$, within a stratum $k$ defined by some combination of age, sex, and (possibly) other determinants of lung cancer:

$$\log_e R_{kxz} = \alpha_k + \beta x + \gamma z \qquad [1a]$$

or, equivalently,

$$R_{kxz} = \exp(\alpha_k + \beta x + \gamma z). \qquad [1b]$$

Here, $k = 1, 2, 3,...$ simply indexes the various strata created by subdividing the cohort into subcohorts homogeneous on age, sex, etc., and $\alpha_k$ represents the log rate among stratum-$k$ subjects who have no smoke or radon exposure ($x = 0$ and $z = 0$).

The coefficients $\beta$ and $\gamma$ of $x$ and $z$ traditionally are called the main effects of smoke and radon. This term suggests that $\beta$ and $\gamma$ represent some sort of causal action of smoke and radon on lung cancer rates. Such an interpretation could, however, be misleading. For example, the magnitude of $\beta$ and $\gamma$ would be affected by a failure to stratify on some cause of lung cancer that is distributed differentially across levels of radon and smoke exposure. For example, if asbestos exposure were associated with radon and smoke exposure in this cohort but the stratification did not include asbestos, one would say that the causal effect of smoke and radon was confounded with the asbestos effect, or that there was confounding by asbestos in the above model and in effect estimates derived from the model.

The parametric definition arose in the context of randomized experiments in agricultural research. Given randomization, the definition is not very misleading. If subjects had been randomized to the various smoke and radon levels, one would not expect smoke or radon to be associated with any potential confounder such as asbestos. Unfortunately, a causal interpretation of $\beta$ and $\gamma$ requires (among other things) absence of confounding; given the difficulty of assuring the latter condition, references to $\beta$ and $\gamma$ as main effects should be regarded as traditional rather than careful usage.

The difficulty with the parametric definition stems from the fact that model 1 describes many different subcohorts of the same cohort (one subcohort for every level of smoke and radon in the total cohort). That is, model 1 is a descriptive model with no causal or biological content. It only describes how the rate varies across subcohorts exposed to different levels of smoke and radon; it does not describe how any of the subcohorts would respond if smoke or radon levels were altered (unless, fortuitously, there is no confounding within the analysis strata). If, say, $x$ is measured in "pack-decades smoked by spouse," a coefficient of $\beta = 0.182$ only says that the rate in subcohorts (strata) with one pack-decade of spousal smoking is on average $\exp(0.182) = 1.2$ times higher than in subcohorts with no spousal smoking; it does not say that this elevated rate is caused by the environmental tobacco smoke.

The second major definition of effect, the counterfactual definition, attempts to deal explicitly with the preceding problem. A causal effect is defined as a contrast of the outcome of a single subject under two different exposure conditions. Consider a married man in our cohort of nonsmokers. Suppose this man would have contracted lung cancer at age 85 if he had no smoke or radon exposure. However, since his marriage at age 20, he has been living with his wife who smokes a pack a day in a house that produces 1 WLM/year of radon-progeny exposure; these exposures result in his developing lung cancer at age 55. Thus, the causal effect of his actual smoke–radon exposure (relative to no exposure) on his incidence time is −30 years; that is, he contracted lung cancer (became an incident case of lung cancer) 30 years sooner than he would have in the absence of both exposures. Note that the condition of no exposure is counterfactual: It refers to what would have happened if, contrary to fact, the man had not been exposed to either smoke or radon.

Counterfactual models for causal effects extend at least as far back as the 1920s but have only seen extensive development over the last few decades (3). Modern development began in philosophy literature (4) and in educational statistics (5); another line of development was introduced into epidemiology by Rothman (6). In the ensuing decade, epidemiologists have employed counterfactual models to define biological interactions (7–9), confounding (10), and intermediate effects (11,12).

The present discussion ignores the problem of competing risks, that is, outcome events that remove a subject from risk of the outcome of interest. For lung cancer, all such competing risks are deaths from other causes, such as a car crash. Proper conceptualization of competing risks in a causal framework is somewhat controversial (13,14). To avoid complexities in presentation, the remaining discussion assumes that within levels of radon, smoking, age, sex, and other controlled covariates, competing risks occur independently of lung cancer. This assumption allows one to interpret all lung cancer incidence times as expected times, given no competing risks occur. Nevertheless, in any application, the assumption would need to be evaluated critically against any available background knowledge.

## Statistical Interaction

In the theory surrounding generalized linear modeling, one commonly sees interactions defined as the coefficients of exposure products in the model. ("Product" here

refers to multiplication, not the result of a chemical reaction.)  Continuing our smoke–radon example, consider the model

$$R_{kxz} = \exp(\alpha_k + \beta x + \gamma z + \delta xz). \quad [2]$$

In the context of this model, the interaction of smoke and radon usually refers to the coefficient $\delta$ of the product $xz$ of smoke and radon level; often, the entire product term $\delta xz$ is called an interaction term. If model 1 is correct, it is said that no exposure interactions or nonlinearities are present on the log-linear or multiplicative scale.

Such usage of the term interaction has been criticized on several grounds (15–17). One criticism is that such usage is algebraic, divorced from any consideration of what constitutes interaction or synergy on the biological level. Another criticism is that such usage renders the presence or absence of interaction entirely dependent on the form of the statistical model one chooses; for the same data, interaction may appear to be present when using one model but absent when using another.

To illustrate the last point, suppose lung cancer rates in our cohort follow the no-interaction log-linear model [1] with $\beta = 0.182$ per pack-decade spousal use and $\gamma = 0.693$ per 100 working-level months (WLM) radon-progeny exposure. Then the expected rates in stratum $k$ will be $R_{k00} = \exp(\alpha_k)$ among subjects with no exposure,

$$R_{k10} = \exp(\alpha_k + 0.182) = 1.2\exp(\alpha_k) \quad [3]$$

among subjects with one pack-decade of spousal-smoke exposure but no radon-progeny exposure,

$$R_{k01} = \exp(\alpha_k + 0.693) = 2.0\exp(\alpha_k) \quad [4]$$

among subjects with no spousal-smoke exposure but 100 WLM radon-progeny exposure, and

$$R_{k11} = \exp(\alpha_k + 0.182 + 0.693)$$
$$= 2.4\exp(\alpha_k) \quad [5]$$

among subjects with one pack-decade of spousal-smoke exposure and 100 WLM radon-progeny exposure. When expressing these four rates in the format of a linear excess-rate-ratio model

$$R_{kxz} = (1 + \beta^* x + \gamma^* z + \delta^* xz)\exp(\alpha_k), \quad [6]$$

one finds that

$$R_{k10} = 1.2\exp(\alpha_k) = (1 + \beta^*)\exp(\alpha_k), \quad [7]$$

$$R_{k01} = 2.0\exp(\alpha_k) = (1 + \gamma^*)\exp(\alpha_k), \quad [8]$$

and

$$R_{k11} = 2.4\exp(\alpha_k)$$
$$= (1 + \beta^* + \gamma^* + \delta^*)\exp(\alpha_k). \quad [9]$$

The rate among the unexposed, $\exp(\alpha_k)$, cancels out of these three equations; this results in three simple linear equations with solutions $\beta^* = 0.2$, $\gamma^* = 1.0$, and $\delta^* = 1.2$. In other words, although no interaction is present when the rates are expressed in a log-linear model (i.e., $\delta = 0$), interaction is present when the rate ratios are expressed in a linear model (i.e., $\delta^* \neq 0$).

## Causal Interactions

A different concept of interaction arises under the counterfactual model of effects. Consider again the man who developed lung cancer at age 55 after living 35 years with a wife who smoked a pack a day, in a house that produced 1 WLM/year of radon-progeny exposure. It was assumed that this man would have survived to develop lung cancer at age 85 only if all smoke particles and radon progeny in his household air had been removed (e.g., filtered) from the air he breathed.

Now ask whether or not the lung tumor he developed (at age 55) would have occurred later (if at all) if all the smoke particles but none of the radon progeny had been removed from the air. If the answer is yes, one says that spousal smoke advanced the incidence time of the subject's lung cancer. Also ask whether or not the tumor would have occurred later (if at all) if none of the smoke particles but all the radon progeny had been removed. If the answer to this question is yes, one may say that the radon advanced the incidence time. If the answer to both questions is yes, so that both exposures contributed to the advance in incidence time, one may say that the factors exhibited cooperative interaction (causal coaction, or synergism) in advancing the subject's incidence time.

To extend the example, suppose the subject would have developed lung cancer at age 70 if only the smoke particles had been removed and at age 65 if only the radon progeny had been removed. The advance in incidence time from 65 in the presence of smoke alone to age 55 in the presence of both exposures represents a portion of the total advance (of 30 years) that required the presence of both exposures to occur. Thus, the portion of the advance from 65 to 55 may be called the interaction effect or coaction of the two exposures.

Coaction is a special case of a more general concept of causal interaction or interdependence of causal effects, which

formalizes (in the counterfactual framework) concepts such as synergy, antagonism, and competitive action. Greenland and Poole (9) review this counterfactual theory and derive its connection to the sufficient-component causal theory of Rothman. Under the counterfactual theory, an instance of synergism between two factors is defined as an instance of disease in an individual that would not have occurred (by a specified time) if either or both factors had been absent. The connection to the above example is that lung cancer would not have occurred by age 55 if either or both factors (35 pack-years of spousal smoke exposure and 35 WLM of radon exposure) had been absent.

Note that the preceding counterfactual concept of synergism does not correspond to mechanism-based concepts of interaction [for example, see (1)]. Certain mechanisms do, however, predict response patterns consistent with this concept when interaction is present.

## Connections among Definitions of Effects and Interaction

The definition of coaction just given bears no resemblance to the statistical definition of interaction; in particular, the concept of coaction is connected only indirectly to the statistical model for the rates. In terms of incidence time, the definition of coaction conflicts with the common definition of synergy as a total effect greater than the sum of the separate effects: In the example, the advance of lung-cancer time when both exposures are present (30 years) is less than the sum of the advance when only radon is present (85 − 70 = 15 years) and the advance when only smoke is present (85 − 65 = 20 years). Nevertheless, there is a connection among these concepts when the problem is reformulated in terms of incidence proportions (i.e., average risks of disease).

As an illustration of this connection, consider the subcohort of male nonsmokers whose exposure histories (up to the time they contract lung cancer) are, say, $x = 1$ pack/day spousal cigarette use and $z = 1$ WLM/year radon-progeny exposure, starting at age 20. Let $R_{xz}(t)$ be the actual proportion of this subcohort contracting lung cancer by age $t$. Define the three counterfactual proportions $R_{x0}(t)$ = proportion of the subcohort contracting lung cancer by age $t$ if only the radon progeny had been removed from the environment; $R_{0z}(t)$ = proportion of the subcohort contracting lung cancer by age $t$ if only the tobacco smoke had been removed; and $R_{00}(t)$ = proportion of the subcohort contracting lung cancer by age $t$ if both the radon progeny and the smoke had been removed. From

the four proportions just defined, one can compute two average-risk differences as measures of the effects radon and smoke would have had in the absence of the other,

$$RD_{x0}(t) = R_{x0}(t) - R_{00}(t) \text{ (radon)} \quad [10]$$

and

$$RD_{0z}(t) = R_{0z}(t) - R_{00}(t) \text{ (smoke)} \quad [11]$$

which are entirely counterfactual, and a difference that measures their actual combined effect,

$$RD_{xz}(t) = R_{xz}(t) - R_{00}(t). \quad [12]$$

It can be shown that superadditivity of the differences,

$$RD_{xz}(t) > RD_{x0}(t) + RD_{0z}(t) \quad [13]$$

can occur only if, in some subjects, radon and smoke causally interact in some of the cohort members; that is, only if coaction has occurred in some members (8,9). Note, however, that the converse is not true: Coaction may take place without superadditivity occurring (8,9).

It follows that an upper one-sided test of the additivity condition

$$RD_{xz}(t) = RD_{x0}(t) + RD_{0z}(t) \quad [14]$$

may be regarded as a test for the occurrence of coaction. Various forms of this conclusion, and tests of additivity (model 14) as a test for synergism, can be found in the pharmacology literature as far back as the 1920s (18). The idea did not seem to attract notice in the epidemiologic literature until the 1970s; see Rothman (15), Koopman (7), and Miettinen (8) for some early formulations. Inequality 13 conforms to the common notion of synergy as a combined effect exceeding the sum of separate effects; note, however, that the effect referred to here is the effect of the exposures on an entire, homogeneously exposed subcohort. In contrast, the effect referred to in the definition of coaction refers to a single subject.

Inequality 13 also conforms to the definition of statistical interaction if one adopts an additive model for the average risks. To see this, define

$$\alpha(t) = R_{00}(t), \beta(x,t)$$
$$= RD_{x0}(t), \gamma(z,t) = RD_{0z}(t), \quad [15]$$

and

$$\delta(x,z,t) = \left[ RD_{xz}(t) - RD_{x0}(t) - RD_{0z}(t) \right]. \quad [16]$$

Then, with a little algebra, we see that inequality 13 implies

$$R_{xz}(t) = \alpha(t) + \beta(x,t) + \gamma(z,t) + \delta(x,z,t)$$

with

$$\delta(x,z,t) > 0. \quad [17]$$

Thus, as before, superadditivity of effects (in particular, an additive-risk model with two causal exposures and a positive product term) implies the presence of interaction. Although the counterfactual and statistical definitions do not otherwise coincide, the superadditive case is, fortunately, the one of primary concern in the study of environmental and occupational hazards, for it is this case that is usually of most public-health concern (16,17).

The counterfactual proportions $R_{x0}(t)$, $R_{0z}(t)$, and $R_{00}(t)$ used for empirical testing of additivity would ordinarily be estimated from comparison groups that are subject to the various combinations of exposure. For example, $R_{00}(t)$ would be estimated from a subcohort with no (or negligible) smoke and radon exposure. This estimate must be adjusted for possible confounding.

In observational epidemiology, adequate adjustment may be difficult or impossible to achieve. There are usually too few subjects to allow simultaneous stratification on all important adjustment variables (confounders) and detailed comparison of exposure groups (although this problem generally is dealt with by using statistical models to estimate the average risks). More intractably, some important confounders may be impractical to measure accurately or to measure at all, and thus may remain uncontrolled. Problems arising from confounder mismeasurement are well recognized in the epidemiologic literature, however (19–21), and will not be a point of focus here. Instead, later sections will discuss the implications of exposure measurement problems for the assessment of interaction.

## Some Problems in Interaction Assessment

### The Power and Precision Problem

In epidemiologic settings, the power to detect statistical interactions is typically an order of magnitude less than the power to detect main effects; see Greenland (22) and Breslow and Day (23) for examples. Similarly, the variance of the interaction estimate will be an order of magnitude greater than the variance of the main-effects estimate under a no-interaction model.

An intuition for these results may be obtained by comparing variance formulas for estimates of main effect and interaction when both exposures $x$ and $z$ are dichotomous with levels 1 (exposed) and 0 (unexposed). Here we consider the basic linear-risk model

$$R_{kxz} = \alpha_k + \beta x + \gamma x + \delta xz \qquad [18]$$

which may be viewed as a special case of model 17. If there is only one stratum and $\delta$ is assumed to be zero (no interaction), the usual estimates of $\beta$ will have a variance approximately equal to $V_1 V_0/(V_1 + V_0)$ where $V_1$ and $V_0$ are the variances of the estimates of $R_{k11} - R_{k01}$ and $R_{k10} - R_{k00}$. In contrast, the usual estimates of $\delta$ will have a variance equal to $V_1 + V_0$. The ratio of the latter variance to the first is $(V_1 + V_0)2/V_1 V_0$, which equals 4 if $V_1 = V_0$ and will be larger if $V_1 \neq V_0$. Thus, in this simple case, the precision of the interaction estimate will be no more than a quarter that of the usual main-effect estimate. An identical result is obtained if one considers a log-linear rate model such as model 1 (23).

Situations involving continuous exposure measurements are considerably more complex, but nevertheless reveal that considerably larger study sizes are needed to study interactions than are required to detect effects (24). We will return to this issue in the discussion of designs for the study of interactions.

## Confounding of Interaction and Dose–Response

In common epidemiologic usage, dose–response refers to the changes in risk produced by changes in a single exposure, whereas interaction refers to changes in risk produced by two or more exposures. Thomas (25) has pointed out that a major problem in the assessment of both dose–response and interaction is their tendency to confound one another, as well as their tendency to confound and be confounded with latency estimates. For example, consider the full quadratic generalization of model 18 to continuous exposures,

$$R_{kxz} = \alpha_k + \beta_1 x + \beta_2 x^2 + \gamma_1 z + \gamma_2 z^2 + \delta xz. \qquad [19]$$

In practice, $x$ and $z$ may be centered (that is, have their sample means subtracted off their observed values) to minimize correlation among the coefficient estimates. Even if this is done, however, the quadratic dose–response terms $x^2$ and $z^2$ will usually be highly correlated with the interaction (product) term $xz$; consequently, if $\beta_2$ and $\gamma_2$ are nonzero, $x^2$ and $z^2$ will act as confounders for

$xz$, so that a biased estimate of $\delta$ will result if $x^2$ or $z^2$ is omitted from the model. In a symmetric fashion, omission of $xz$ will bias the $\beta_2$ and $\gamma_2$ estimates if $\delta$ is nonzero.

More generally, failure to adequately model dose–response and latency can lead to bias in interaction estimates and vice-versa. Perhaps a more illuminating way to view this problem is to recognize that dose–response, latency, and interaction assessment are actually facets of a single task, namely assessment of the shape of the joint time-dependent dose–response surface relating incidence to both exposures. For example, model 19 specifies that this surface is quadratic; without specific prior knowledge about combined smoke and radon effects, there would be no basis for omitting any term from the model (unless the data clearly indicated a term was negligible).

Of course, model 19 is fairly restrictive as is its log-linear analogue (obtained by replacing $R_{xz}$ with $\log_e R_{xz}$), and does not encompass the possibility of transforming $x$ and $z$ to improve model accuracy and to model latency. Some alternative modeling approaches will be discussed below. The present point is that dose–response and interaction should be viewed in a unified fashion if one wishes to avoid higher-order confounding.

## Measurement Errors

In ordinary language, a measurement error is simply the act of recording an incorrect value for some variable on some subject. Statistical theory is concerned with the distribution of these errors in the study population and the relationship between true and measured values. For example, one may ask a number of questions involving the measured and true values for environmental tobacco smoke, such as: a) What is the distribution of true values $x$ among subjects with measured values $x_m$? b) What is the distribution of measured values $x_m$ among subjects with true values $x$? c) Do the errors in the measured values $x_m$ vary systematically across levels of the true values $x$ of smoke? (If not, the smoke errors are said to be additively homogeneous.) d) Do the errors in the measured values $x_m$ vary systematically across levels of other variables? (If so, the errors are said to be differential; if not, the errors are said to be nondifferential.) e) Are the errors in the measured values $x_m$ of smoke associated with the errors in the measured values $x_m$ of radon? (If not, the errors in the two variables are said to be independent of each other.)

An analogous list can be made for the errors in measuring lung cancer incidence time. Traditionally, however, disease out-

comes have been treated as dichotomies (diseased/not diseased), and errors in disease measurement have been treated as diagnostic errors, which are evaluated in terms of sensitivity (probability of true positive among cases) and specificity (probability of true negative among noncases).

The above listing does not exhaust the possibilities, and hence it may be clear that the topic of measurement error, and all its possible effects, can become exceedingly complex. It should not be surprising then that most studies on the topic are limited in scope and usually make several simplifying assumptions. Most commonly, errors are assumed to be independent and nondifferential, so that the answers to questions $d$ and $e$ and the analogous questions for disease are negative. One rationale for such an assumption in methodologic studies is that if some bias arises from well-behaved (independent nondifferential) errors, the same sort of bias or worse should be expected if the errors are not well behaved. Although this rationale is not valid universally (26), investigators often attempt to ensure that these errors will be independent and nondifferential, and so such errors are worth studying in detail.

Nevertheless, it should be recognized that optimistic conclusions based on assuming independent nondifferential errors cannot be extended to dependent or differential errors, and that the errors actually occurring in a study can become differential under ordinary circumstances. Consider, for example, exposure measurements over time. Such measurements often are based on historical records or, worse, subject memory. In such situations, exposure measurements for the more distant past may be less accurate than measurements for more recent exposure; if so, accuracy of cumulative exposure measurement will vary with any variable correlated with calendar time, such as another exposure. Even if the intrinsic accuracy of the exposure measurements do not vary over time, the degree of bias produced by measurement errors may still vary over time (27). Similar problems will arise if accuracy of outcome measurement (e.g., disease diagnosis) varies over time.

## The Impact of Measurement Errors

The impact of measurement errors on main-effect estimates has been studied extensively, especially for situations involving independent nondifferential error. One well-known result is that independent nondifferential errors in the classification of a dichotomous exposure and covariate cannot produce bias away from the null value

of the exposure effect; for example, any bias produced by such error in the estimate of $\beta$ in model 1 will be towards zero. This result, while useful, is often stated without mention of the assumptions of independent errors and dichotomous exposure.

Unfortunately, violations of either assumption can result in bias away from the null; Dosemeci et al. (28) show that independent nondifferential classification error can produce bias away from the null if the exposure has as few as three levels. It is not clear, however, how often such bias occurs in practice, and there are a number of special error models under which the estimated coefficients in linear or log-linear models can only be biased towards the null. For example, this is so under the classical model, in which the measured value $x_m$ is given by $x_m = x + \varepsilon_x$, where $x$ is the true value, $\varepsilon_x$ is the $x$ error, and $x$ and $\varepsilon_x$ are normally distributed with $\varepsilon_x$ independent of all other variables (including $x$)—that is, the error is independent, additively homogeneous, non-differential, and normal. Although these conditions are restrictive, the result extends to various cases involving nonnormal exposures and errors. Extension to multiplicative errors, with $x_m = x \cdot \varepsilon_x$ and $x$ and $\varepsilon_x$ strictly positive, follows by using $\log(x_m) = \log(x_t) + \log(\varepsilon_x)$ in place of $x_m$ as the regressor variable. These and other results for special models are reviewed by Armstrong (29). Lubin et al. (25) specifically consider models for radon measurement to evaluate the impact of measurement errors in studies of tobacco smoke, radon, and lung cancer.

The impact of measurement errors on interaction estimates has been studied less thoroughly. Independent nondifferential classification errors can produce spurious appearances of interaction and can mask true interactions, depending on other features of situation (19). More generally, the interaction coefficient $\delta$ in models 17 and 18 may be biased towards or away from the null by independent nondifferential errors in the study covariates (regressors); errors in disease classification may further aggravate such biases, thus distorting the entire shape of the dose–response surface. These results easily extend to situations involving arbitrary polytomous or continuous exposures (Appendix). Nevertheless, there are a number of special cases in which nondifferential independent error will not affect the validity of tests for interaction, and may rarely or never produce bias away from the null; for example, if the true values were distributed jointly and normally and if the errors were independent, additively homogeneous, non-differential, and normal (that is, if $x_m = x +$

$\varepsilon_x$ and $z_m = z + \varepsilon_z$, where $x$, $z$ are bivariate normal and the errors $\varepsilon_x$, $\varepsilon_z$ are normal and independent of $x$, $z$, and each other), or if the errors were independent, nondifferential, and $x$ and $z$ were not associated with each other (Appendix).

The distortion of dose–response and interaction estimates produced by measurement error depends heavily on the particulars of the study distribution of exposures and errors. Thus, rather than rely on any general (and possibly misleading) conclusions, it may be best to evaluate the effects of measurement error on a study-specific basis, using methods of the sort discussed in the next section. In the particular case of environmental tobacco smoke and radon, measurement errors may render the study of interactions infeasible due to attenuated power (24); a similar conclusion may apply to most other epidemiologic studies of environmental exposures.

## Coping with the Problems

### Designs for Assessing Interactions and Dose–Response

In studies involving primary subject selection, power for detection of interactions can be increased by using special sampling plans. Unfortunately, a major obstacle in employing such designs is that they require a priori specification of a number of parameters that may be only vaguely known, if at all. For cohort studies, one must be able to specify likely values for the intercept and main-effect parameters (e.g., $\alpha$, $\beta$, $\gamma$ in model 18) in the model of interest, as well as a value for the interaction parameter ($\delta$) for which one wishes to maximize power or precision. For case–control studies, the intercept need not be specified, but one must have some idea of the exposure distributions in the population serving as the source of cases and controls.

A considerable amount of literature exists for choosing optimal designs, at least in the cohort framework; Seber and Wild (30) provide references to the linear-model literature and also review design methods for nonlinear models. Although this literature is highly technical, a few general conclusions can be drawn, especially in the special case of studying departures from risk or rate additivity.

The optimal design for detecting departures from additivity will not correspond to the optimal design for detecting departures from linearity of the dose–response curve for each exposure. Nor will either of these designs correspond to the optimal design for detecting main effects; however, the presence of main effects will hopefully have been established before embarking on a specialized study of interactions.

Because one will have to simultaneously consider interaction and dose–response, as explained earlier, it may be best to select subjects to maximize precision of the estimated dose–response surface. In this approach, interaction represents but one of several potentially important departures from linearity of the joint dose–response surface relating smoke ($x$) and radon ($z$) to risk. For example, consider the quadratic-risk model given in model 19. A good design for studying such a model would select subjects to enhance the precision of estimates for $\beta_2$ and $\gamma_2$, as well as $\delta$.

More generally, one would want to allow for response surfaces other than quadratic, including perhaps unanticipated shapes. One simple cohort design to help achieve this end would try and insure that subjects are distributed evenly across the joint range of smoke and radon levels (that is, across the combinations of $x$ and $z$).

The case–control situation is not addressed as easily, for it is the case–control ratio rather than the joint exposure distribution that is controlled by the investigator. Nevertheless, if one is willing to sacrifice the ability to estimate the main effect of one of the exposures, one also may manipulate the marginal distribution of that exposure by, for example, case–control matching; see Smith and Day (31) and Thomas and Greenland (32) for some elementary studies of the impact of matching on interaction assessment in the context of log-linear interactions. For interaction assessment, one can expect that certain highly variable matching ratios will offer more precision than fixed ratios: Relatively few controls per case would be needed in strata with many cases, but relatively many controls per case would be needed in strata with few cases.

If one already knows the joint distribution of disease and one of the exposures in the source population, it may be most efficient to employ a two-stage design rather than a conventional matched design; see Cain and Breslow (33) for further discussion of this point.

### Modeling Interactions and Dose–Response

The confounding of interactions and dose–response can be overcome if one has accurate information on the values of the variables (here, smoke and radon) over a reasonably broad range of combinations of the variables. Even with accurate and broad-ranging measurements, however, one must take care to employ a model form flexible enough to accurately approximate the true dose–response surface. Because the shape of the true surface usually is

unknown (and is in fact what is under study), a safe strategy would be to employ as flexible a model form as practical.

The most flexible approaches available are nonparametric regression methods, such as bivariate smoothers; for examples, see Hastie and Tibshirani (34). Unfortunately, these methods are not yet implemented widely in software, are impractical for handling more than a few regressors, and can require fairly large samples for reasonable performance. An easier approach, with somewhat less flexibilty, is generalized additive modeling (34). As an example, the generalized-additive analogue of model 1 would be

$$\log_e(R_{xz}) = \alpha_k + \beta(x) + \gamma(z), \quad [20]$$

where $\beta(x)$ and $\gamma(x)$ are now unspecified functions of $x$ and $z$ that will be estimated from the data. Unlike model 1, which constrains dose–response to be log linear, model 20 allows the dose–response for smoke and radon to be any shape at all. Both models 1 and 20 do, however, imply that the shape for the smoke dose–response does not change across levels of radon or covariates, and the shape for the radon dose–response does not change across levels of smoke or covariates; this set of constraints is called the no-additive-interaction or parallelism condition. Model 20 is easily fit using the GAIM software package (35). To generalize model 20 to allow for departures from additivity, one may add a product-term function to obtain

$$\log_e(R_{kxz}) = \alpha_k + \beta(x) + \gamma(z) + \delta(xz). \quad [21]$$

This is one of several possible generalized-additive analogues of model 1. Unlike model 20, it does not constrain the dose–response surface to contain parallel dose–response curves.

All the models given so far imply that the shape of the dose–response surface does not change across the covariate strata (i.e., there is no additive interaction with covariates). To get around this restriction, one could model the covariate effects in detail and add interaction terms between the covariates and exposures to the model. Among the drawbacks of this strategy is that the resulting model may have too many terms for the fitting procedure to work. Even if the model can be fit, the individual terms may be estimated with little accuracy. The individual terms also may be difficult to interpret, although this need not be a problem if one focuses on graphs of the response surfaces instead of on model terms.

Further extensions of the above models may be obtained by considering other transformations of the outcome measure, as in the additive logit model in which

$$\text{logit} R_{kxz} = \alpha_k + \beta(x) + \gamma(z), \quad [22]$$

where $\text{logit } R = \log_e[R/(1-R)]$. One also may employ incidence times or rates in place of risks as the outcome measure in the above models. The latter models often fit better and may even obviate the need for product terms in the model. They also allow for straightforward incorporation of time-dependent exposures in the model, an obvious advantage in longitudinal studies of exposures such as smoke and radon. Nevertheless, tests of the no-coaction hypothesis still correspond to testing the fit of an additive-risk model (such as model 17 or 18) (36).

Unfortunately, additive-risk models cannot be fit to case–control data unless one has sufficient external information to reconstruct the population risks from the data. For unmatched studies, all one needs is an estimate of the crude disease rate in the source population of cases and controls or knowledge of the case and control sample fractions. For matched studies, one must have the crude rates or sampling fractions within levels of the matching factors. Given this information, however, one may fit the same variety of model forms as used for cohort data (37).

For further discussions of modeling issues and techniques see Breslow and Day (23), McCullagh and Nelder (2), and Hastie and Tibshirani (34). Less technical overviews of modeling are given by Greenland (38) and Checkoway et al. (39).

## Evaluating and Correcting for Measurement Error

The best means of coping with measurement error is, of course, not to have it. Because this ideal is not attainable in typical environmental and occupational studies, evaluation of measurement error and its effects is an essential component of any informative study. Most evaluations are limited to narrative review of factors influencing errors and the implications for bias; most commonly, these evaluations comprise arguments that exposure-measurement errors were independent and nondifferential and hence produced only bias towards the null. As shown earlier, however, such arguments are of little use in interaction assessment, because independent, nondifferential misclassification may bias interaction terms in any direction.

Much more can be done if data are available about the accuracy of the exposure and covariate measurements in the study. In the best situation a validation substudy is conducted in which exposure is remeasured in a subsample of subjects using criterion methods, that is, methods more accurate than the general methods applied to all subjects. The association of the criterion and general measurements, as estimated from the validation substudy, may then be used to correct coefficient estimates obtained from the full study cohort. Correction methods also may be applied if the criterion-general measurement association is estimated from data external to the study (although, in the latter case, one must assume that this association is the same in both the study and the external data). There is now an extraordinary variety of validation-based correction methods available (for example 40–42).

If a criterion measurement is unavailable, it still may be possible to obtain a more limited correction of coefficient estimates using a reliability substudy in which replications of the general measurement are obtained on a subsample of subjects. Again, there is a variety of reliability-based correction methods (e.g., ref. 43).

If neither validation data nor reliability data are available, but some educated guesses can be made about the distributions of exposure and covariate errors, one may conduct a sensitivity analysis of the study results. In such an analysis, various hypothesized error distributions are used to correct the study results; one thus sees how sensitive estimates are to assumptions about the error distribution. This analysis is conducted easily under various simplifying assumptions (29). If the study variables are discrete, matrix formulas for correcting contingency-table results can be applied (40, Appendix), and these are programmed easily in matrix languages such as GAUSS, SC, S-PLUS, and SAS IML.

## Conclusions

Given the difficulties inherent in attempting to study interactions with epidemiologic data, design and analysis is best focused on accurate estimation of the entire dose–response surface relating incidence to covariates, rather than on isolated aspects of this surface, such as statistical interaction. One may, of course, test the departure of the data from surfaces predicted by various causal models, such as the no-coaction model (7,9) or the simple independent-action model (44), but the power and validity of these tests will be nearly optimal under the same conditions that insure accuracy of dose–response estimation, such as well-balanced exposure distributions and accurate exposure measurement.

Flexible modeling and, where possible, quantitative evaluation of measurement error

will help achieve the most accurate assessment of interaction possible with available data. Nevertheless, because of limitations of power and because of distortions produced by measurement error, one should be cautious about the potential of environmental epidemiology for interaction assessment.

## Appendix

For simplicity, suppose we have just one stratum, and let $P(x,y \mid x_m, z_m)$ be the probability that a subject with measured smoke and radon exposures $x_m$ and $z_m$ has true levels $x$ and $z$; note that $\Sigma_{xz} P(x,y \mid x_m, z_m) = 1$ (here, $\Sigma_{xz}$ indicates the sum over all possible values of $x$ and $z$). Let

$$\bar{x}(x_m, z_m) \equiv \Sigma_{xz} \, xP(x,y \mid x_m, z_m)$$

and

$$\bar{z}(x_m, z_m) \equiv \Sigma_{xz} \, zP(x,y \mid x_m, z_m) \quad [23]$$

be the means of the true smoke and radon levels among subjects with measured levels $x_m$ and $z_m$; let $R_{xz}$ be the average risk among subjects with true levels $x$ and $z$; and suppose $R_{xz}$ follows the no-interaction linear-risk model (model 18 with $\delta = 0$). Then the average risk among subjects with measured levels $x_m$ and $z_m$ will be

$$R(x_m, z_m) = \Sigma_{xz} P(x,z \mid x_m, z_m) R_{xz}$$
$$= \Sigma_{xz} P(x,z \mid x_m, z_m)(\alpha + \beta x + \gamma z)$$

$$= \alpha \cdot 1 + \beta \Sigma_{xz} \, xP(x,z \mid x_m, z_m)$$
$$+ \gamma \Sigma_{xz} \, zP(x,z \mid x_m, z_m)$$
$$= \alpha + \beta \bar{x}(x_m, z_m) + \gamma \bar{z}(x_m, z_m).$$

$$[24]$$

Now let $RD(x_m, z_m) = R(x_m, z_m) - R(0,0)$ be the risk difference between subjects with measured levels $x_m, z_m$ and subjects measured as having no exposure. Then

$$RD(x_m, z_m)$$
$$= \alpha + \beta \bar{x}(x_m, z_m) + \gamma \bar{z}(x_m, z_m)$$
$$\quad - [\alpha + \beta \bar{x}(0,0) + \gamma \bar{z}(0,0)]$$
$$= \beta [\bar{x}(x_m, z_m) - \bar{x}(0,0)]$$
$$\quad + \gamma [\bar{z}(x_m, z_m)] - \bar{z}(0,0)]; \quad [25]$$

in contrast, for subjects measured as exposed to only one of the two exposures, we have

$$RD(x_m, 0) + RD(0, z_m)$$
$$= \beta [\bar{x}(x_m, 0) - \bar{x}(0,0)]$$
$$+ \gamma [\bar{z}(x_m, 0) - \bar{z}(0,0)]$$
$$+ \beta [\bar{x}(0, z_m) - \bar{x}(0,0)]$$
$$+ \gamma [\bar{z}(0, z_m) - \bar{z}(0,0)]$$
$$= \beta [\bar{x}(x_m, 0) + \bar{x}(0, z_m) - 2\bar{x}(0,0)]$$
$$+ \gamma [\bar{z}(x_m, 0) + \bar{z}(0, z_m) - 2\bar{z}(0,0)].$$

$$[26]$$

Thus, except in certain special cases,

$$RD(x_m, z_m) \neq RD(x_m, 0) + RD(0, z_m). \quad [27]$$

that is, the risks based on the measured exposures need not be additive, and this is so even if the measurement error is independent and nondifferential and the risks based on the true exposures are additive.

Additivity will be preserved (i.e., 25 will equal 26 under model 18 with $\delta = 0$) if the mean true levels $\bar{x}$ and $\bar{z}$ depend on the measured levels $x_m$ and $z_m$ in an additive fashion, for then

$$\bar{x}(x_m, z_m) - \bar{x}(0,0)$$
$$= \bar{x}(x_m, 0) + \bar{x}(0, z_m) - 2\bar{x}(0,0)$$
$$[28]$$

and

$$\bar{z}(x_m, z_m) - \bar{z}(0,0)$$
$$= \bar{z}(x_m, 0) + \bar{z}(0, z_m) - 2\bar{z}(0,0)$$
$$[29]$$

This would occur, for example, if the errors were independent nondifferential and $x$ and $z$ were unassociated, or if $x$ and $z$ were bivariate normal and their respective errors were independent normal with homogeneous variance. Additivity also will be preserved under "Berkson error" [see Armstrong (29) for discussion of Berkson error in the context of main-effect estimates]. ◄║

---

## REFERENCES

1. Thompson WD. Effect modification and the limits of biological inference from epidemiologic data. J Clin Epidemiol 44:221–232 (1991).
2. McCullagh P, Nelder JA. Generalized linear models, 2nd ed. New York: Chapman and Hall, 1989.
3. Rubin DB. Comment: Neyman and causal inference in experiments and observational studies. Stat Sci 5:472–480 (1990).
4. Lewis D. Causation. J Philos 70:556–567 (1973).
5. Rubin DB. Estimating causal effects of treatment in randomized and nonrandomized studies. J Educ Psych 66:688–701 (1974).
6. Rothman KJ. Causes. Am J Epidemiol 104:587–592 (1976).
7. Koopman J. Interaction between discrete causes. Am J Epidemiol 113:716–724 (1981).
8. Miettinen OS. Causal and preventive interdependence: elementary principles. Scand J Work Environ Health 18:159–168 (1982).
9. Greenland S, Poole C. Invariants and noninvariants in the concept of interdependent effects. Scand J Work Environ Health 14:125–129 (1988).
10. Greenland S, Robins JM. Identifiability, exchangeability, and epidemiological confounding. Int J Epidemiol 15:412–418 (1986).
11. Robins JM. Control of confounding by intermediate variables. Stat Med 8:679–701 (1989).
12. Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. Epidemiology 3:143–155 (1992).
13. Slud EB. Dependent competing risks and the latent-failure model (letter). Biometrics 44:1203–1204 (1988).
14. Prentice RL, Kalbfleisch JL. Letter to editor. Biometrics 44:1205 (1988).

15. Rothman KJ. Synergy and antagonism in cause–effect relationships. Am J Epidemiol 99:385–388 (1974).
16. Rothman KJ, Greenland S, Walker AM. Concepts of interaction. Am J Epidemiol 112:467–470 (1980).
17. Rothman KJ. Modern epidemiology. Boston: Little, Brown, 1986.
18. Greenland S. Elementary models for biological interaction. J Haz Materials 10:449–454 (1985).
19. Greenland S. The effect of misclassification in the presence of covariates. Am J Epidemiol 112:564–569 (1980).
20. Kupper LL. Effects of the use of unreliable surrogate variables on the validity of epidemiologic research results. Am J Epidemiol 120:643–648 (1984).
21. Savitz DA, Barón AE. Estimating and correcting for confounder misclassification. Am J Epidemiol 129:1062–1071 (1989).
22. Greenland S. Tests for interaction in epidemiologic studies: a review and a study of power. Stat Med 2:243–251 (1983).
23. Breslow NE, Day NE. Statistical methods in cancer research II: the design and analysis of cohort studies. Lyon: International Agency for Research on Cancer, 1987.
24. Lubin JH, Samet JM, Weinberg CR. Design issues in epidemiologic studies of indoor exposure to radon and risk of lung cancer. Health Physics 59:807–817 (1990).
25. Thomas DC. Are dose-response, synergy, and latency confounded? Alexandria, VA: American Statistical Association, 1981.
26. Greenland S, Robins JM. Confounding and misclassification. Am J Epidemiol 122:495–506 (1985).

27. Greenland S. Interpreting time-related trends in effect estimates. J Chron Dis 40 (Suppl 2):17S–24S (1987).

28. Dosemeci M, Wacholder S, Lubin JH. Does nondifferential misclassification always bias a true effect toward the null value? Am J Epidemiol 132:746–748 (1990).

29. Armstrong BG. The effects of measurement error on relative risk regressions. Am J Epidemiol 132:1176–1184 (1990).

30. Seber GAF, Wild CJ. Nonlinear regression. New York: Wiley, 1989.

31. Smith PG, Day NE. The design of case–control studies: the influence of confounding and interaction effects. Int J Epidemiol 13:356–365 (1984).

32. Thomas DC, Greenland S. The efficiency of matching in case–control studies of risk-factor interactions. J Chron Dis 38:569–574 (1985).

33. Cain KC, Breslow NE. Logistic regression analysis and efficient design for two-stage studies. Am J Epidemiol 128:1198–1206 (1988).

34. Hastie TJ, Tibshirani RJ. Generalized additive models. New York: Chapman and Hall, 1990.

35. Almudevar T, Tibshirani R. GAIM version 1.1. Toronto: S.N. Tibshirani Enterprises, 1991.

36. Greenland S. Additive risk versus additive relative risk models. Epidemiology 4:32–36 (1993).

37. Greenland S. Multivariate estimation of exposure-specific incidence. J Chron Dis 34:445–453 (1981).

38. Greenland S. Modeling and variable selection in epidemiologic studies. Am J Public Health 79:340–349 (1989).

39. Checkoway H, Pearce NE, Crawford-Brown D. Research methods in occupational epidemiology. New York: Oxford University Press, 1989.

40. Greenland S. Variance estimation for epidemiologic effect estimates under misclassification. Stat Med 7:745–757 (1988).

41. Rosner B, Spiegelman D, Willet WC. Correction of logistic regression relative risk estimates and confidence intervals for measurement error: the case of multiple covariates measured with error. Am J Epidemiol 132:734–745 (1990).

42. Pepe MS, Fleming TR. A nonparametric method for dealing with mismeasured covariate data. J Am Stat Assoc 86:108–113(1991).

43. Thompson WD. Kappa and attenuation of the odds ratio. Epidemiology 1:357–369 (1990).

44. Weinberg CR. Applicability of the simple independent-action model to epidemiologic studies involving two factors and a dichotomous outcome. Am J Epidemiol 123:162–173 (1986).