
Measurement Issues in Environmental Epidemiology

Maureen Hatch^{1*} and Duncan Thomas²

¹Columbia University School of Public Health, 600 West 168th Street, New York, NY 10032; ²University of Southern California, School of Medicine, Department of Preventive Medicine, 1420 San Pablo Street, Los Angeles, CA 90033

This paper deals with the area of environmental epidemiology involving measurement of exposure and dose, health outcomes, and important confounding and modifying variables (including genotype and psychosocial factors). Using examples, we illustrate strategies for increasing the accuracy of exposure and dose measurement that include dosimetry algorithms, pharmacokinetic models, biologic markers, and use of multiple measures. Some limitations of these methods are described and suggestions are made about where formal evaluation might be helpful. We go on to discuss methods for assessing the inaccuracies in exposure or dose measurements, including sensitivity analysis and validation studies. In relation to measurement of health outcomes, we discuss some definitional issues and cover, among other topics, biologic effect markers and other early indicators of disease. Because measurement error in covariates is also important, we consider the problems in measurement of common confounders and effect modifiers. Finally, we cite some general methodologic research needs. — *Environ Health Perspect* 101(Suppl 4):49–57 (1993).

Key Words: Biologic markers, dose, environment, exposure, mathematical modeling, measurement error, psychosocial factors, sensitivity analysis, susceptibility

Measuring Environmental Exposure and Dose

Concepts

Environmental exposures can occur as a result of contact with a variety of elements (air, water, soil) that, in turn, influence the pathways for exposure (inhalation, ingestion, dermal). Individuals' interactions with these elements are complex, and therefore it is not surprising that exposure assessment and dose estimation are formidable challenges to those investigating the health effects of environmental agents.

The concepts of exposure and dose have been elaborated in a series of recent publications issued by the Board on Environmental Studies and Toxicology of the National Academy of Sciences (1,2). The term exposure refers to the concentration of an agent at the boundary between an individual and the environment as well as the duration of contact between the two, but dose refers to the amount actually deposited or absorbed in the body over a given time period. Although internal dose is the ideal measure from the scientific standpoint, regulation can deal only with external exposures, and therefore one may want to measure both exposure and dose.

Individuals' exposures may be modified by factors such as activity patterns, which determine encounters with various sources of exposure; bioavailability of the agent in time and place; and the rate at which exposure occurs (e.g., a relatively constant rate versus a variable rate). From a given exposure, a person's resultant dose will depend on host characteristics, such as age, sex, and metabolism. It also will reflect the susceptibility of target tissue at the time of exposure; any shielding provided by the body (e.g., the placenta, the blood-brain barrier) or modulation by buildings that attenuate exposure to electric fields and gamma radiation but can be a source of exposure to radon; and the effect of concurrent exposures, such as cigarette smoking or medications. In addition, only particular components of the dose may be relevant to health effects. For calculating dose-response relationships, this biologically effective dose is what ought to be quantified. But in many instances it may be difficult to define what the biologically effective dose is, much less measure it. In any event, the definition is time-dependent and subject to change along with the state of scientific knowledge, just as measurement capabilities change with new technology. Epidemiologists undoubtedly need to prepare for a new generation of studies in which measurement of variables will involve data at the level of the gene. A commitment of resources, such as talent and funding, could improve the state of the art in exposure and dose assessment and potentially yield better estimation of exposure-response

relationships and more effective measures of environmental protection.

In the past, the methods used to assign exposures in environmental health studies were quite crude, and to some extent they still are (e.g., pesticide usage patterns, residence near a point source of pollution). Even in studies where disease has been ascertained at the individual level, exposure measures may be ecologic in nature and based on average levels for a group. When the group is defined in geographic terms, exposure levels might be estimated from values recorded by environmental sampling in a subject's general vicinity. However, recent research has shown that correlations sometimes are weak between readings from area monitors and subjects' exposures measured using personal monitors (3), which are presumed to relate more closely to the true dose. Discrepancies between readings from personal and areawide samples can result from heterogeneity of exposures, from poor placement of samplers (e.g., air monitors at elevations well above the breathing zone), or from failure to take account of human activity patterns and other sources of exposure.

Exposure monitoring systems can be and are being improved, however. Newer approaches include sampling the microenvironments where exposure principally occurs, including indoor environments (e.g., bedrooms and living rooms in studies of radon and electric and magnetic fields), as well as total exposure monitoring in which all potentially relevant microenvironments are

This manuscript was prepared as part of the Environmental Epidemiology Planning Project of the Health Effects Institute, September 1990–September 1992.

*Author to whom correspondence should be addressed.

This work was supported in part by grant R01-HD24659 from the National Institute for Child Health and Human Development.

sampled (4,5). The latter approach is particularly important for ubiquitous compounds like the polycyclic aromatic hydrocarbons. To some extent, personal exposure monitoring is also beginning to be incorporated into environmental health studies. In addition to these attempts to improve externally derived measures of exposure, efforts are being made to estimate internal dose using strategies like empirical dosimetric modeling, pharmacokinetic modeling, and biologic markers.

Such efforts are important. The failure to assign individual exposure and dose accurately leads to measurement errors with consequent effects on measures of association (and, ultimately, risk assessments) that will differ depending on whether the error is random or systematic and whether the unit of analysis is the individual or the group. Systematic error in exposure measurement can introduce bias either toward or away from the null. Random error tends to bias results toward the null, although exceptions to the rule can be found in unusual circumstances (6). For ecologic studies in which exposure is a binary variable derived from combinations of individual observations, the rule stating random error generally biases results toward the null may not hold (7).

Given the consequences of error in estimating exposure, it is important to try to increase accuracy of measurement at the design stage of a study. How, then, does an investigator decide when the use of a surrogate exposure measure (i.e., an error-prone measure) is acceptable, and when it is not? Rosner et al. have shown (8) that for correlations between surrogate and true measures of exposure less than 0.8, the odds ratios estimated by logistic regression will differ markedly for the surrogate and the true exposure measure, while much less bias will occur when correlations between the two measures are 0.8 or greater. *In vivo* tibia lead levels measured by X-ray fluorescence have been proposed as a good surrogate for cumulative blood lead levels on the basis of a correlation coefficient of 0.84 (9). For dietary exposures, however, the correlation between food frequency questionnaires and less error-prone methods (food records, measurements in food or biological samples) is only around 0.5 (10); yet food frequency questionnaires continue to be applied in large-scale studies, only occasionally with correction of risk estimates for error in measurement. On the other hand, the failure to find a correlation (actual coefficients not given) between current adipose tissue or serum dioxin levels and surrogate measures of past exposure to

Agent Orange in Vietnam (11,12) affected a decision not to conduct further research using exposure surrogates based on troop location and herbicide spraying records. These examples underscore the need to be explicit about criteria for acceptable surrogate measures, as well as the need to take error into account when surrogates are used, even while emphasizing the development of better approaches to exposure-dose assessment.

In the following section, we describe methods designed to reduce error in exposure measurement insofar as is currently possible (approaches such as dosimetric modeling, pharmacokinetic modeling, biologic markers, and use of multiple measures), as well as approaches to assessing the residual uncertainties in the estimated dose. Even the best of the current methods will not yield a measure that is completely error-free, and it is therefore important to recognize and characterize the residual error in measurement so that it can be considered in analysis of the data.

Measurement Approaches

Exposure or Dose Modeling

Estimating a subject's exposure to an environmental agent involves combining information about possible sources of exposure (usually obtained from the subject, from some other respondent, or from records) with an assessment of the likely degree of exposure from each source.

When an exposure under study is environmental, there may be multiple pathways by which a person might be exposed and it can be important to consider all elements and all routes. For example, residents downwind of the Nevada Test Site could have been exposed to external gamma radiation from the passing fallout cloud itself, from ingesting contaminated milk or vegetables, or, in the case of infants, from in utero exposures or breast-feeding. For each of these pathways, several different radionuclides might need to be considered. After eliminating pathways that would be expected to make a negligible contribution to the total dose, one can estimate the likely dose rate per unit of exposure to each pathway. In the fallout example, this involved consideration of *a*) source term, the amount and type of radionuclide released; *b*) the environmental transport, dispersion from the source to sites of deposition; *c*) rate of radioactive decay and environmental dispersion of the radionuclides; *d*) farm management practices leading to contamination of dairy cattle or vegetables; *e*) estimates of the uptake of radionuclides by vegetables and milk; *f*) distribution of

milk and vegetables to consumers; and *g*) uptake by the target organ from ingested radionuclides. To calculate an individual's dose, this information was then combined with extensive questionnaire data on breast-feeding and maternal and individual consumption of milk and vegetables at various ages. For some subjects, modifications were needed to allow for homegrown vegetables or backyard cows or goats. For subjects with incomplete exposure information, distributions of default values specific to their particular circumstances (age, sex, location, etc.) were developed. Similar calculations were performed for each of over 100 nuclear tests, and the results then were summed to produce estimates of each subject's total dose (13).

The process described above is far more complex than has been the norm in environmental epidemiology, but it represents the current state of the art in environmental dose assessment. Less refined, but perhaps less costly, approaches to exposure-dose modeling (often for households or geographic areas rather than for individuals) have been based on Gaussian-dispersion modeling of airborne emissions (14-16), hydrogeologic modeling of waterborne exposures (17), and isopleth modeling of soil contaminants (18). Assuming that dosimetry models are reasonably accurate, such approaches should decrease bias arising from measurement error and increase precision. Assessment of the validity of dosimetry models should be made whenever possible. For example, an environmental dispersion model of emissions at the time of the accident at the Three Mile Island nuclear plant was validated by the readings from off-site thermoluminescent dosimeters.

Dosimetric modeling methods are likely to be used more frequently in future environmental health studies. A question is whether the effort required both in terms of the information that must be collected from study subjects and/or by environmental sampling and the effort involved in development of the dosimetric model itself are warranted by the gain in precision or reduction in bias of the exposure estimates. Information on this point could be obtained by comparing the point and interval estimates of associations observed using gold standard dose estimates with those that would be obtained using cruder methods. Such comparisons could be made in existing data sets. Understanding when the gains from dosimetric modeling are substantial and when they are only marginal would be useful in establishing methodologic standards of practice.

Some other issues related to dosimetry are exemplified by studies of cancer and electric and magnetic fields (EMFs). The initial hypothesis about EMFs was derived from observations showing apparent excesses of leukemia (and some other cancers) both in children living near electric power lines that would be expected to generate high magnetic fields (19) and in certain classes of electrical workers (20). In both the residential and occupational settings, it has been difficult to establish whether the magnetic fields are the responsible agent. While subsequent studies have demonstrated that certain electrical wiring configurations and certain categories of electrical work are associated with higher than average fields, so far no convincing associations have been found between leukemia risk and individuals' exposure to electric or magnetic fields determined by area measurements. No studies using personal dosimetry have yet been reported.

Four possible explanations are suggested for the failure to establish a clear association between cancer and measured field strengths. First, it may be due to their extreme variability in space and time. Any necessarily short-term measurement (24 hr or a week in a small number of locations) is a poor surrogate for lifetime dose; under this explanation, household wiring classifications and job titles may be more stable measures of long-term exposure. Second, the failure to detect an association with measured fields may reflect a failure to measure the biologically relevant parameter (e.g., peaks, transients, resonance between static and oscillating fields rather than the time-weighted average). Studies of reproductive outcomes, where the period of exposure is much shorter than for cancer and where there may be a particular time window of vulnerability, could help indicate whether the discrepancy in associations with wire codes and measured fields is due to their capturing different time frames or different dimensions of EMFs. A third explanation for the associations of cancer with wiring configurations, but not with measured fields, relates to selection bias (lower selection probabilities for controls living near wiring with high current configurations). Fourth, the surrogate exposure measures (wire codes, job titles) may be confounded by other correlated risk factors. This controversy is still far from resolved, but consideration of selection bias and possible confounders together with careful assessment of all potentially salient aspects of electric and magnetic fields and of the variability of the different measurements should shed light on the issue.

The EMF example underscores the need for making multiple measures of exposure. In particular, it argues for continuing to include surrogate measures along with gold standard measures in studies of health effects until the relations between the surrogate and criterion measures are well understood and there is certainty about the true gold standard (i.e., until the correct biologic mechanism is known). Substituting an incorrect gold standard for a surrogate measure can actually increase measurement error. One analytic approach to using multiple measures that has been proposed as a means of increasing validity is to restrict analysis to subjects who are classified as exposed or unexposed by two different, if imperfect, exposure measures (21). This clearly risks some loss in power since subjects with discordant results on the two measures are excluded from analysis. Another proposed approach is to estimate the misclassification probabilities for each measure and from them to estimate the prevalence of exposure (22).

Some mention of personal monitors should also be made. While these do not provide a measure of resulting body burden, as biologic markers are meant to do, personal monitors may measure the intensity of an individual's total exposure to airborne agents better than fixed-site area monitors. This is not always the case, however, particularly in studies of long-term exposures or where areawide concentrations are fairly uniform. The TEAM study (Total Exposure Assessment Methodology) conducted by the U.S. Environmental Protection Agency (EPA) found that personal air monitors were acceptable to subjects from 7 to 85 years of age (23). Investigators studying effects of exposure to EMFs and indoor air pollutants on children are anxious to develop personal monitors that can be used with children under age seven, including toddlers. At present, personal monitors for EMFs are in the form of wristbands and may not be suitable for very young children. Technology for personal exposure monitoring is still evolving, but it will rarely be feasible to apply personal exposure monitoring to all subjects and all relevant time periods. Therefore, methodologic approaches are needed for combining collected exposure data with personal samplers and environmental monitors.

Pharmacokinetic Modeling

Pharmacokinetic modeling is an approach to dosimetry that incorporates information about the internal pharmacologic processes that ensue once an agent reaches the portal(s)

of entry into an individual's body (24). These include uptake into the circulation; distribution within the body; and metabolism, storage, and elimination. These models can be simple, involving only one body compartment, or complex, involving multiple body compartments. In either case, compartmental rate relationships are used in the model's equations to estimate concentrations at critical tissues. Such models are also useful as guides to temporally relevant and efficient ambient sampling (24). Pharmacokinetic modeling of exposure and dose may be viewed as a counterpart to biologically based disease models.

Biologic Markers

Because of the difficulty of obtaining accurate and unbiased exposure information from study subjects and the difficulty of estimating the doses that such exposures might produce, there has been great interest in the development of biologic markers. These may be defined as "cellular, biochemical, or molecular alterations that are measurable in biological media, such as human tissue, cells, or fluids" (25). If used appropriately, biologic markers allow for considerable improvement in measurement of dose. First, they may obviate the errors arising from subjects' lack of knowledge, memory failure, biased recall, or deliberate misinformation (26). Second, even when subject reports of exposure are accurate, individuals may vary considerably in uptake and handling of a material; the error introduced by such individual variation can be reduced or removed by using markers that provide an estimate of the dose to a particular individual. Third, some markers can be used to detect biological interactions between the exposure of interest and critical tissues; DNA adducts are an example of this type of marker. In studying environmental tobacco smoke, for instance, one can—in addition to asking about maternal smoking during pregnancy—actually measure smoking-related DNA adducts in placentae (27) and, where the fetus is lost, in critical organs such as fetal lung or liver (28). Another advantage of biologic markers is that generally they give a quantitative, or at least semi-quantitative, estimate of dose. They also can serve as the gold standard for other information sources, thus providing a basis for error allowance procedures in studies that rely on less accurate exposure measures due to the cost of the marker.

Other Biologic Dosimeters

Certain signs or symptoms can also be viewed as biologic dosimeters. For example,

in the cohort of atomic bomb survivors, it has been reported that subjects with a history of epilation have a 2.5-fold steeper dose-response curve for leukemia than those without (29). This can be interpreted either as an indicator of their greater radiosensitivity or as an indicator of miscalculation of their doses, perhaps as a result of differences in shielding not accounted for by available dosimetry data.

To be useful in environmental epidemiology studies, a biologic exposure marker should be clearly better than anamnestic data or environmental measures; should allow for differentiation between exposure levels; should be applicable on a large scale; or if too costly for large-scale use, should at least be acceptable to subjects in a validation sub-study. Before markers are used in epidemiologic research, their sensitivity and specificity should be known from both the laboratory and epidemiologic perspectives; reproducibility of results within and between laboratories must also be known; and, very importantly, the particular time frame they reflect and during which they can be measured *in vivo* must be established (25) so that they provide interpretable data regarding time and dose.

At present, few exposure markers satisfy these requirements. Some markers may provide a record of cumulative exposure (e.g., bone lead measurement, mercury or cocaine measurements in hair), but most can assess only relatively recent exposures. Studies of biologic markers that use a case-control design and a cross-sectional marker of exposure can be difficult to interpret because of ambiguity about the temporal sequence of the marker and the disease [e.g., whether selenium levels in breast cancer cases are cause or consequence (30)]. Indeed, such studies can be misleading. Vineis and Caporaso (31) have described how a case-control study nested in a cohort allowed Wald and his colleagues (32) to make use of the time between initial collection of specimens from members of the cohort and subsequent onset of cancer to clarify the time order in the relationship with blood retinol. Although analysis considering only the early cases of cancer suggested that blood retinol might be protective, ultimately it was apparent that some metabolic change associated with the disease was acting to reduce retinol levels, rather than vice versa. In addition to such problems in interpretation, biological measurements are often costly to perform. Furthermore, the need to obtain specimens can reduce the cooperation of subjects and introduce the potential for selection bias to occur through initial refusal or later attrition, although these problems

are probably not insurmountable if they are anticipated and addressed.

Use of Multiple Measures

When the biological basis of an association is poorly understood, it can be very helpful to have various types of exposure measurements available. Or, as mentioned previously in connection with personal exposure monitoring, it may be necessary to rely on another source of exposure information for portions of the study period. The obvious approach is to analyze each type of measurement separately, but there may be merit in combining them into an index, if only to reduce measurement error. Complications can arise if all measurements are not available on the same subjects. Any associations observed might be due to differences in the measurements or to differences in the subgroups of subjects for whom the measurements are available. In a study of childhood leukemia and electric and magnetic fields, London et al. (33) reported the results separately for various summaries of 24-hr bedroom dosimetry, spot measurements at various locations, and wiring configurations. However, drawing on all of these data, they also developed regression models for magnetic fields at various locations based on attributes of the wiring and used the values predicted by these models as the time-weighted average fields for all houses lived in. Thus, predicted values were used both to replace existing measurements and to impute missing values. The rationale behind the approach is to avoid the loss of information and possible selection bias associated with restricting analysis to subjects with data for all measurements made (34). One alternative is to retain measurements where they exist and to impute only the missing values, leaving open the possibility of stratifying on data quality in the analysis. Other approaches undoubtedly can be devised, and it would be desirable to compare their validity using data sets in which exposure-response relationships are well understood and where more than one measure of exposure exists.

Other Issues in Measurement of Exposure

Taking Account of Critical Periods for Exposure

A principal problem in environmental epidemiology has been that the inaccuracy in measurement generally (although not always) operates in the direction of overestimating exposure and therefore underestimates risk or perhaps misses health effects altogether. For example, when assigning the same level of exposure to all 1000 resi-

dents living within five miles of a toxic dump site when only 100, say, were truly exposed and the other 900 were either unexposed or exposed at very low levels, one would be certain to calculate an observed relative risk for exposure that would be lower than the true risk. Hence the importance of increasing the accuracy of exposure definitions and measurement is obvious. Rothman and Poole have pointed out (35) that it is also important to use information on critical periods for exposure, either in the design phase of a study, in the analysis phase, or in both. For example, in a study of Down's syndrome, parental exposures occurring after the fertilization period are presumably irrelevant to the outcome; in fact, there is mounting evidence that most cases of Down's are traceable to errors at the time of the first meiotic division in the maternal germ cell (36). By removing all exposures that are not of biologic consequence from the estimate of association, one can expect the magnitude of the estimated association to increase. Moreover, information on known critical periods might be used to test whether an association appears to be spurious. If an association were found not only during the critical period but also for exposure during noncritical periods, then the association might be due to recall bias, or it could be reflecting autocorrelations in exposure status. Multivariate analysis of the effects of exposure in various critical and noncritical periods could, in principle, overcome this problem, provided there are enough exposed subjects with different temporal patterns of exposure to be informative.

Taking Account of Migration In and Out of Exposed Areas

The problem of in- and out-migration is frequently raised as an issue in interpreting results of studies that define exposure in terms of time and place. Although several studies have considered the effects of population migration on the validity and precision of estimated associations between exposure and disease (37) and have described when and in what direction bias is likely to arise, these issues are still not understood well. Perhaps more simulations or empirical demonstrations are needed to improve the general level of comprehension about the effects of population mobility on geographic studies. In the case of specific studies, it would help to know something about duration of residence or at least age-specific duration patterns in an area. One recent suggestion is to estimate by various means the fraction (f) of time spent by a subject in a particular place and to assign for

the remaining fraction $(1-f)$ the average exposure for some total referent area (38).

Assessing Past Exposure

A major problem in many environmental health studies is the difficulty of estimating past exposures when only present-day measurements are available. Often, some data on subjects' past exposures can be obtained by questionnaire or review of existing records. For example, in occupational studies, payroll records are used to assemble a job history. The use of records from years past to establish exposure status has the important advantage of obviating recall bias, although it may introduce its own problems (e.g., missing records or less specificity in records from early years). Estimating the actual historical exposure levels is more difficult than simply classifying exposure status, and it often involves a large degree of judgment. Clearly, the more historical data there are on variation in exposure levels over time and place, the better. Study of such patterns of variation can suggest models for predicting exposures at times for which no measurements are available. For example, in a study of salivary tumors and dental X-rays, Preston-Martin et al. (39) reviewed 58 studies that described doses from various procedures at various times and, while taking into account the dates of introduction of new technologies, used regression analysis to develop models for the expected dose as a function of calendar year. In occupational settings, the subjective experience of long-service workers has been used to compare current exposures with those in the distant past. Similar strategies (i.e., tracking technological developments, use of knowledgeable informants) need to be applied in the assessment of past environmental exposures. For example, in a case-control study of colorectal cancer and water chlorination among women teachers in New York State, Lawrence et al. (40) used current water sampling in conjunction with records from water treatment plants covering the previous 20 years in a mathematical model to estimate cumulative exposures to chloroform in drinking water at home and at work.

Uses of Existing Environmental Databases

One limitation on assessing past environmental exposures is that reviews of existing data bases at the national and state level repeatedly have found them to be inadequate for epidemiologic purposes because of insufficient data points to assess variability, lack of a standardized Quality Assessment/Quality Control protocol, incomplete geographic coverage, and missing information (41). Efforts

are underway to modify the major air and water data bases to make them more useful for future environmental health studies.

Existing environmental data banks could also be used to define strata within which to conduct sample surveys. Surveys of individuals within these ecological exposure groupings would help document human activity patterns and could indicate the distribution of exposure and important confounding or effect-modifying variables in each stratum. Potentially, such stratified-sample surveys might provide the basis for constructing an environment-exposure matrix similar to the job-exposure matrices used in occupational studies. Such exposure matrices are generally assumed to have a "Berkson error" structure (42), in which the average of the true doses for all subjects in an exposure assignment group is equal to the assigned value. As a consequence, if the true dose-response is linear, the estimated slope of a linear relationship will not be biased toward the null.

Estimating Dose Uncertainties

A major concern among environmental epidemiologists is the influence of errors in exposure estimates on associations with disease and methods of dealing with such errors. The best cure for this problem is to avoid measurement error in the first place. When this is not feasible (and it often may not be, particularly in investigating common source exposures such as toxic dump sites), it is helpful to be able to quantify the direction and magnitude of the errors. This can be done in a number of ways, including *a*) validation studies on a subset of the study sample or a pilot sample to compare the measurements to be made in the field with a gold standard, *b*) replication of measurements to assess within-subject variability, *c*) multiple types of measurements to assess validity, and *d*) sensitivity analysis to estimate the influence of various unknowns or uncertain parameters on the estimated doses. The goal might be either to describe the distribution of exposure errors across the population (or subgroups there of) or to obtain an estimate of the precision of each subject's exposure assignment.

Because a gold-standard assay is often not feasible for use in the field (because of cost, time, acceptability, etc.), validation studies usually must be limited to a relatively small number of subjects. The resulting estimates of error distributions may be imprecise (43), although this will be less of a problem if the data are treated as continuous and if parameters for sensitivity and specificity do not have to be estimated (8). Nonetheless, sample sizes for validation studies that are needed to insure

good estimates of the error rates in field measurements should be calculated carefully. Other considerations are to insure that the measurement error process in the sample used for validating the field measure is similar to that in the target population for the full study and to avoid selection bias in the validation study, which might arise if requirements associated with use of the gold standard measure are very demanding and participation rates are consequently low. In the New Jersey case-control study of radon and lung cancer among women, in-home radon measurements were obtained for only 40% of the houses targeted, and smoking rates differed among those with measured and unmeasured homes, raising the possibility of selection bias (44). If data on disease are collected on validity study participants, potential selection bias can be examined by testing for heterogeneity in the risk estimates.

Replicate measurements are useful for describing repeatability (45) but cannot assess other components of error, such as subjects' tendency to consistently overreport or underreport exposures. Having different types of measurements available may be more useful in estimating misclassification probabilities, even if none of the measures is error-free. See, for instance, Hui and Walter's maximum likelihood method for estimating error rates with two independent assessments of exposure (22).

Sensitivity analyses can take a number of forms. The basic idea is to consider a range of plausible values for each of the unknowns in the exposure assignment process. If there are only a few unknowns, one might consider each of them and evaluate their influence on either the individual exposure assignments or the final dose-response relation. If there are many, one can estimate the distribution of assigned doses, either analytically or by Monte Carlo simulation. The latter approach was used in the studies around the Nevada Test Site because of the complexity of the dosimetry algorithm. Components of uncertainty that were considered include the source term, environmental transport, farming practices and distribution, and default values for individuals' missing data. A series of sensitivity analyses were also carried out on a mathematical model that estimated the relative geographic distribution of exposure to accident emissions at Three Mile Island by examining variations in modeling assumptions for their effect on the base case (46). Parameters considered were the source term, the degree of plume rise, wind shifts, and residual error weighting. In addition, a Bayesian analysis was used to quantify uncertainty about the time-release pattern.

Measuring Outcome of Environmental Exposures

Definitional Issues

As strong effects of environmental exposure have been identified and dealt with, environmental epidemiology increasingly has become a search for weaker associations. It is all the more important, therefore, to improve measurement of outcome through careful definition and avoidance or reduction of error (35). In defining study end points, the aim should be to specify the health outcome of interest as precisely as possible in order to avoid further dilution of a weak association through inclusion of irrelevant cases. In fact, it may be desirable to consider subgroups of disease that are etiologically homogeneous and that are believed to be responsive to the exposure of interest on the basis of theory or prior observations (e.g., certain histopathologic types of lung cancer and radon; leukemia types and subtypes with ionizing radiation and EMFs). This can present something of a dilemma, however, because statistical power for examining subgroups is likely to be low unless the difference in effect size among subgroups is sufficient to offset the reduced sample size.

The virtues of lumping versus splitting frequently come up for discussion in the context of studies of congenital anomalies. It is unlikely that an exposure would affect all types of congenital defects. With maternal cocaine use during pregnancy, for example, defects involving vascular disruption seem to be implicated. However, a biological basis for positing subgroups of interest is often lacking; empirical Bayesian approaches may be useful in helping to formulate relevant subgroupings. In any event, the numbers in particular case groups are likely to be small for all but a few categories. If sufficiently large series cannot feasibly be accrued in a single study, multisite (even multinational) projects may need to be mounted, or more reliance may need to be placed on meta-analyses combining results from several studies. Which of these strategies to pursue should be discussed by groups of investigators studying the same exposure, and their potential funding sources.

Disease outcomes in environmental epidemiology can be measured on a continuous scale or categorically as incident or prevalent cases or as deaths. Incidence data are usually preferable for investigating etiology since prevalence or mortality data may be influenced by factors affecting duration of disease and survival as well as

those relating to cause. However, incidence data are often less easily accessed than mortality data, and they can be subject to artifactual variations in ascertainment—as a result of screening programs, for example. Whether incidence or mortality is the more reliable indicator of health status and in what age groups it is reliable have been discussed extensively but not resolved. See, for example, the recent papers by Doll (47) and by Davis et al. (48) about cancer time trends. It might be helpful to have a set of recommended approaches for trend analysis that were developed by a group of dispassionate methodologists. For etiologic studies, incidence data seem conceptually superior; when mortality data are used, consideration needs to be given to accounting for influences on survival since these might correlate with exposure.

In some areas of research, such as reproduction and development, different outcomes can occur depending on the timing and dose of exposure. In such circumstances, it may be important to examine several end points. Extending population-based registration systems to cover more outcomes than cancer and birth defects and to cover more geographic areas potentially could be useful for environmental studies in several respects: in identification of cases, in validation of self-reported information, and in ascertaining disease status of migrants.

Biologic Effect Markers and Other Early Indicators of Disease

Biologic effect markers potentially have a number of advantages as study end points, particularly if they are strongly prognostic of disease in ways not explained by available exposure information—for example, by reflecting susceptibility or the action of cofactors (26). While some effect markers are actually subclinical events (e.g., biochemical tests of occult pregnancy loss), often markers of effect correlate only weakly with disease. Serum alpha-fetoprotein is a useful marker for liver cancer as well as a prenatal marker for neural tube defects. Markers that are not as clearly predictive of risk, particularly at the individual level, can lead to problems of interpretation and to needless anxiety for those individuals found to have elevated levels. The premature application of a poorly standardized cytological assay on a group of already concerned residents at Love Canal is a case in point. Calls have been made repeatedly to carry out longitudinal studies, in experimental animals and humans, that will measure the positive predictive value of such

markers before applying them in field studies; but these have been largely ignored. The Scandinavian countries, however, have mounted a collaborative prospective study of cancer in a cohort of 3190 individuals who have been tested for sister chromatid exchanges (SCEs), structural chromosome aberrations, or both. A report based on a 13-year follow up of 800 subjects in the Finnish portion of the data (49) found a moderate, statistically significant positive association between cancer risk and chromosome aberrations (SMR = 2.65; 95% CI 1.2, 5.0); there was a positive trend (SMR = 2.06; 95% CI 0.8, 4.2) for SCEs. Additional prospective studies of this kind are needed to establish the relationships between markers and disease in order to assure their appropriate use and interpretation. In addition, determining when a marker could serve as the basis for preventive health measures directed at a distal end point such as cancer is an important issue; see Prentice (50) for a useful discussion of this and a proposed operational criterion for surrogate response variables.

Other potential advantages of biologic effect markers are their use in classifying disease more precisely and in suggesting mechanisms of action, such as those relating to susceptible subpopulations. For example, biologic markers that distinguish slow from fast acetylators have indicated that the enzyme *N*-acetyltransferase plays an important role in bladder cancers induced by exposure to aromatic amines (51,52). Methodologic needs in the area of effect markers include attention to sources of variability, both biological and laboratory-related, and to logistical issues, such as how to achieve reasonable participation rates when the effect marker requires a demanding regimen. Three current studies of early pregnancy loss illustrate this latter problem. Two of the studies ask participants for daily urine samples. The third study uses a modified specimen collection scheme requiring urine samples only twice monthly, at the beginning of menses. Preliminary data indicate higher response rates for the study with the simplified collection protocol. Whether the variability in enrollment is due to the differing demands on study subjects or to other variable aspects of the three studies (such as the perceived salience of the topic in the target population) is not known. Systematic research is needed to determine how to achieve cooperation in studies that use biologic markers and how to provide for calculating or estimating the extent and magnitude of selection bias.

Subclinical End Points

What role should physiologic changes (e.g., nerve conduction velocity, T-cell subsets, sperm count) have in environmental health assessments? It has been argued that functional alterations and nonspecific symptoms are likely to be more frequent consequences of low-level environmental exposures than frank disease (53). However, baseline rates and normal ranges for such end points may be lacking. Objective methods of assessment to remove the potential for biased recall may be at an early stage of development, and interpretation of results in terms of risk to groups and to individuals frequently is problematic, particularly as assay improvement allows for discriminating function more and more minutely. These methodologic limitations can be addressed—semen evaluation is a case in point (although the clinical significance of altered semen quality is still not clear-cut)—however, substantial time and effort will be required.

Measuring Confounders and Effect Modifiers

Effect on Risk Estimates If Inadequately Controlled

A confounding variable is one that, if not controlled appropriately, will tend to distort the exposure–disease association. For example, when studying whether household exposure to radon is a cause of lung cancer, one should be concerned about the possible confounding effect of smoking. Smoking is clearly a major risk factor for lung cancer. If houses with high radon levels are more likely to be inhabited by smokers, then this would produce an apparent relationship between radon and lung cancer even if there were no causal effect. The converse also could happen; if smokers tended to live in low-radon houses, then one might fail to find an association between radon and lung cancer if it really were present.

The strategies commonly used by epidemiologists to control confounding include restriction (e.g., to nonsmokers), matching, or statistical adjustment. All of these approaches presume that the confounding variable has been correctly measured. Greenland (54) has pointed out that errors in measurement of a confounding variable will tend to cause partial loss of an ability to eliminate confounding bias; for example, if the true odds ratio (adjusted for the true confounder) is 2.0 and the crude odds ratio (unadjusted) is 4.0, then the odds ratio adjusted for an incorrectly or crudely measured confounder might be 3.0. This intermediate outcome can only be counted upon in a case in which the errors in mea-

suring the confounder are random (unrelated to exposure or disease status); in other cases, the adjusted odds ratio could be further from the truth than the unadjusted odds ratio. Kupper (55) has shown that an inaccurate surrogate confounder can produce seriously misleading inferences.

A factor like smoking, in addition to being a confounder, could also act as an effect modifier—that is, a variable that modifies the strength of the association between exposure and disease. A major question in the radon literature is whether the joint effects of smoking and radon exposure are multiplicative, additive, or some intermediate possibility. If they act additively, for example, then radon exposure would produce the same additional risk of lung cancer in smokers and nonsmokers; but because lung cancer is rare in nonsmokers, it would follow that radon exposure might account for a much larger proportion of lung cancers in that group. Conversely, if the two exposures act multiplicatively, the proportional increase in lung cancer rates due to radon exposure would be the same in smokers and nonsmokers; but because of the higher rates in smokers, the absolute increase would be larger in smokers. This issue therefore has important risk assessment and public health policy implications. Again, Greenland (54) has shown that errors in measurement of a covariate can distort its modifying effect and possibly introduce an apparent interaction where none exists. Diet and cooking habits in relation to aflatoxin exposure, and showering habits in relation to radon are additional examples of potentially important confounding or effect-modifying variables in environmental epidemiology.

Approaches to Measuring Common Confounders and Modifiers

The implications of the previous section are that careful measurement of strong confounders or modifiers should be given as much attention as the exposure and disease variables. It follows that some of the same approaches discussed in the sections on measurement of exposure and disease, such as use of multiple measures and biologic markers, will pertain here as well.

Continuing with the example of smoking, it is not sufficient simply to classify subjects by their present status as current, former, or never smokers. As long as smoking is a risk factor for the disease under study, one usually tries to obtain information on at least the ages at starting and stopping and the average daily amount of smoking. These data can be used to compute pack-years (the

product of amount and duration), which is a stronger predictor of lung cancer risk than current status. In some other cases, however, such a product term may actually increase error. Better yet, nonlinear multivariate models could be used to allow for the joint effects of age at starting, duration and intensity of smoking, and time since quitting. Other modifying factors might include changes in level of smoking over time, use of filter cigarettes, and depth of inhalation. However, incorporating multiple modifying factors into an analysis needs to be done with considerable thought to produce models that are biologically plausible. Routine inclusion of interaction terms in a multiple logistic regression analysis can produce models in which ex-smokers eventually become at lower risk than never smokers, or light smokers have the same dependence on duration or age at start as heavy smokers. Use of general risk models based on biologically plausible theories is an attractive alternative.

Even the most complete smoking history is still likely to be misclassified, and the errors might well be related to the exposure or disease variables under study. In an occupational study of radon exposure and lung cancer, for example, miners with lung cancer might preferentially underreport their smoking histories to avoid prejudicing a compensation claim. For these reasons, there has been great interest in developing unbiased methods of assessing potential confounders. Biological measures, such as urinary cotinine for smoking or 4-amino-biphenyl-DNA adducts, are very attractive for this purpose. Other approaches were discussed above, in the section on exposure measurement. The disadvantage of most of these methods is that they measure only recent exposure and lifetime exposure will still be misclassified. The development of methods for combining information from different types of measurements could be very useful. Also discussed previously in the exposure measurement section, and equally relevant here, is the need to assess and allow for measurement error in confounders and effect modifiers whenever possible. Therefore, consideration should be given to mounting validation substudies to quantify measurement error in important covariates.

Susceptibility

Variation within a population in sensitivity to an exposure of interest can be substantial. Khoury et al. (56) estimated the proportion of susceptible individuals in the population for cigarette-induced cancers at

several sites; the proportions varied from <1% for oral and esophageal cancer up to 13% for cancer of the lung. Bias in risk estimates will arise if individuals with similar exposures but different susceptibilities are treated the same. There are a number of epidemiologic designs for assessing sensitivity to environmental exposures. As a measurement problem, the central issue is whether the marker for sensitivity being examined is a measurement of the genotype itself, some host characteristic, or family history.

The ability to classify genotypes directly has profound implications for identifying sensitive individuals. The obvious difficulty is that there are millions of genetic loci, for which only a relatively small number have probes available and only a few might be relevant to any particular disease. Thus, some prior knowledge that a locus has a role in the disease process is essential before embarking on a search for interactions with possible environmental exposures. Even so, the information for identifying genetically susceptible individuals may involve invasive and costly tests.

Recognition of phenotypically distinguishable subgroups of the population that have different baseline risks of disease or sensitivities to environmental exposures can therefore be very useful for public health protection. The measurement issues that arise here are essentially no different from those for any other effect modifier, as discussed above.

For family history as a marker of susceptibility to a disease, the basic minimal information that needs to be collected is the identification of the family members with the disease and the number, ages, and relationships of family members at risk. This information should be collected systematically for all first-degree relatives (parents, siblings, and offspring), and possibly for all second-degree relatives. As the objective is to examine family history as a marker of sensitivity to an environmental exposure, every effort should be made to obtain exposure information on all relatives, not just the affected ones.

Psychosocial Stress as Confounder, Effect Modifier, and Mediator

The psychosocial stress that may be associated with exposure to a perceived environmental hazard can potentially confound, mediate, or modify any associations between the exposure and disease. Stress might operate indirectly and cause exposed individuals to alter risk behaviors. Stress also could have an artifactual association with the end point of concern because of changes in care seeking, diagnostic practices, or self-reported health states. Alternatively, concern about environmental exposures could cause adverse outcomes other than those potentially associated with the perceived hazard. For example, studies around the Three Mile Island and Chernobyl nuclear plants indicate that the perception of danger can increase distress levels or clinical states like anxiety and depression (57,58), irrespective of whether radiation-induced increases in cancer actually occur.

The issue of stress as a confounder, effect modifier, mediator, indicator of some methodologic bias—or even as an exposure or outcome—needs to be explicitly addressed in future environmental epidemiologic research conducted on sensitized populations. Some relevant methodology has been developed in studies of communities near toxic wastes to distinguish between biologic effects of exposure to hazardous substances at such sites and either symptoms of stress or altered symptom reporting (59,60). These preliminary efforts include use of a scale to measure hypochondriasis and stratified analysis of self-reported symptoms to take account of subjects' perception about the source of pollution. Environmental epidemiologists need to learn when and how to address the issue of psychosocial stress in order to clarify interpretation of health effects studies and to estimate the importance of stress in its own right. Consideration should be given to measuring perceived stress and physiologic indicators of stress as well as to collecting data on methodological covariates such as motivation to participate,

interest in receiving health care, and beliefs about the exposure in question as a cause of adverse health effects.

Methodologic Needs and Recommendations

The aspect of study design that involves measurement of variables is critical, especially in fields like environmental epidemiology where the risks from exposure are likely to be small, difficult to detect, and perhaps not clinically significant, yet may be of public health importance. Methodologic research in this area should emphasize the further development and application of dosimetric modeling. Existing data sets representing a range of research problems within environmental epidemiology could be used to assess the gains from dosimetry algorithms compared with cruder, more conventional methods of exposure assessment.

Dosimetry models invariably will use a combination of questionnaire data, environmental measurements, and biologic markers; this underscores the need for development and refinement of methods for handling multiple measures. Biologic markers themselves, as measures of exposure, effect, or susceptibility, are an area where additional methodologic development would be desirable.

A second important aspect of methodologic research relates to sensitivity analyses and other approaches for estimating the uncertainty in measurement of exposure and dose. Included in this category would be validation studies to compare a gold standard with a more error-prone exposure measurement in order to allow for correction of bias in the analysis stage of research. Consideration needs to be given to the costs and benefits of investigating measurement error in the primary study or in a sub-study (which could be carried out internally or externally in relation to the primary study). A final area that deserves attention is measurement error in covariates, which can be as important as measurement error in the exposure or outcome variables. □

REFERENCES

1. National Research Council. Biologic markers in reproductive toxicology. Washington DC: National Academy Press, 1989.
2. National Research Council. Human exposure assessment for airborne pollutants: advances and opportunities. Washington, DC: National Academy Press, 1991.
3. Wallace LA, Ziegenfuss RC. Comparison of carboxyhemoglobin concentrations in adult nonsmokers with ambient carbon monoxide levels. *J Air Pollution Control Assoc* 35:944–949 (1985).
4. Liroy PJ. Exposure analysis and assessment for low-risk cancer agents. *Int J Epidemiol* 19:S53–S61 (1990).
5. Wallace L, Pellizzari E, Hartwell TD, Perritt R, Sparacino C, Whitmore R, Sheldon L, Zelon H. The TEAM (total exposure assessment methodology) study: personal exposures to toxic substances in air, drinking, and breath of 400 residents of NJ, NC, and ND. *Environ Res* 43:290–307 (1987).
6. Dosemeci M, Wacholder S, Lubin JH. Does nondifferential misclassification of exposure always bias a true effect toward the null value? *Am J Epidemiol* 132:746–748 (1990).
7. Brenner H, Savitz DA, Jöckel KH, Greenland S. The effects of nondifferential exposure misclassification in ecological studies. *Am J Epidemiol* 135: 85–95 (1992).

8. Rosner B, Spiegelman D, Willett 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* 132:734-45 (1990).
9. Somerville LJ, Chettle DR, Scott MC, Tennant DR, McKiernan MJ, Skilbeck A, Trethowan WN. In vivo tibia lead measurements as an index of cumulative exposure in occupationally exposed subjects. *Br J Ind Med* 45:174-181 (1988).
10. Willett WC, Sampson L, Stampfer MJ, Rosner B, Bain C, Witschi J, Hennekens CH, Speizer FE. Reproducibility and validity of a semiquantitative food frequency questionnaire. *Am J Epidemiol* 122:51-65 (1985).
11. Kang HK, Watanabe KK, Breen J, Remmers J, Conomos MG, Stanley J, Flicker M. Dioxins and dibenzofurans in adipose tissue of US Vietnam veterans and controls. *Am J Pub Health* 81:344-349 (1991).
12. Centers for Disease Control Veterans' Health Studies. Serum 2,3,7,8-tetrachlorodibenzo-p-dioxin levels in U.S. Army Vietnam-era veterans. *J Am Med Assoc* 260:1249-54 (1988).
13. Stevens W, Thomas DC, Lyon JL, Till J, Kerber RA, Simon SL, Lloyd RD, Elghany NA and Preston-Martin S. Leukemia in Utah and radioactive fallout from the Nevada Test Site: a case-control study. *J Am Med Assoc* 264:585-591 (1990).
14. Hatch M, Beyea J, Nieves J, Susser W. Cancer near the Three Mile Island nuclear plant: radiation emissions. *Am J Epidemiol* 132:397-412 (1990).
15. Brown LM, Pattern LM, Blot WJ. Lung cancer in relation to environmental pollutants emitted from industrial sources. *Environ. Res.* 34:250-261 (1984).
16. Eifler CW, Stinnett SS, Buffer PA. Simple models for community exposure from point sources of pollution. *Environ Res* 25:139-146 (1981).
17. Wrensch M, Swan S, Murphy PJ, Liscomb J. Hydrogeologic assessment of exposure to solvent-contaminated drinking water: pregnancy outcomes in relation to exposure. *Arch Environ Health* 45:210-216 (1990).
18. Crump KS, Ng TH, Cuddihy RG. Cancer incidence patterns in the Denver metropolitan area in relation to the Rocky Flats plant. *Am J Epidemiol* 126:127-135 (1987).
19. Wertheimer N, Leeper F. Electrical wiring configuration and childhood cancer. *Am J Epidemiol* 109:273-284 (1979).
20. Milham S. Mortality from leukemia in workers exposed to electrical and magnetic fields. *N Engl J Med* 307:249 (1982).
21. Marshall JR, Graham S. Use of dual responses to increase validity of case-control studies. *J Chron Dis* 37:125-136 (1984).
22. Hui SL, Walter SD. Estimating the error rates of diagnostic tests. *Biometrics* 36:167-171 (1980).
23. Wallace LA. The total exposure assessment methodology (TEAM) study. Washington DC: U.S. Environmental Protection Agency, 1987.
24. Smith TJ. Exposure assessment for occupational epidemiology. *Am J Ind Med* 12:249-268 (1987).
25. Hulka B, Wilcosky T, Griffith J. Biological markers in epidemiology. New York: Oxford University Press, 1990.
26. Stein Z, Hatch M. Biological markers in reproductive epidemiology: prospects and precautions. *Environ Health Perspect* 74:67-75 (1987).
27. Everson RB, Randerath E, Santella RM, Cefalo RC, Avitts TM, Randerath K. Detection of smoking-related covalent DNA adducts in human placenta. *Science* 231:54-57 (1986).
28. Hatch M, Warburton D, Santella R. Polycyclic aromatic hydrocarbon-DNA adducts in spontaneously aborted fetal tissue. *Carcinogenesis* 11:1673-1675 (1990).
29. Sposto R, Stram DO, Awa AA. An investigation of random errors in the DS86 dosimetry using data on chromosome aberrations and severe epilation (RERF TR 7-90). Hiroshima: Radiation Effects Research Foundation, 1990.
30. Meyer F, Verreault R. Erythrocyte selenium and breast cancer risk: brief report. *Am J Epidemiol* 125:917-919 (1987).
31. Vineis P, Caporaso N. Applications of biochemical epidemiology in the study of human carcinogenesis. *Tumorigenesis* 74:19-26 (1988).
32. Wald N, Boreham J, Bailey A. Serum retinol and subsequent risk of cancer. *Br J Cancer* 54:957-961 (1986).
33. London SK, Thomas DC, Bowman JD, Sobel E, Peters JM. Exposure to residential electric and magnetic fields and risk of childhood leukemia. *Am J Epidemiol* 134:923-937 (1991).
34. Rubin J. Multiple imputation for non-response in surveys. New York: John Wiley, 1987.
35. Rothman K, Poole C. A strengthening program for weak associations. *Int J Epidemiol* 17 (Suppl):955-959 (1988).
36. Antonarakis SE, and the Down's Syndrome Collaborative Group. Paternal origin of the extra chromosome in trisomy 21 as indicated by analysis of DNA polymorphisms. *N Engl J Med* 324:872-876 (1991).
37. Polissar L. The effect of migration on comparison of disease rates in geographic studies in the United States. *Am J Epidemiol* 111:175-182 (1980).
38. Cohen BL. Ecological versus case-control studies for testing a linear-no threshold dose-response relationship. *Intl J Epidemiol* 19:680-684 (1980).
39. Preston-Martin S, Thomas DC, White SC, Cohen D. Prior exposure to medical and dental x-rays related to tumors of the parotid gland. *J Natl Cancer Inst* 80:943-949 (1988).
40. Lawrence CE, Taylor PR, Trock BJ, Reilly AA. Trihalomethanes in drinking water and human colorectal cancer. *J Natl Cancer Inst* 72:563-68 (1984).
41. Armentrout DW. Overview of EPA major air data bases. In: *Environmental epidemiology* (Kopfler FC, Craun GF, eds). Chelsea, MI: Lewis Publishers, 1986; 149.
42. Armstrong BG. The effects of measurement errors on relative risk regressions. *Am J Epidemiol* 132:1176-1184 (1990).
43. Greenland S. Statistical uncertainty due to misclassification: implications for validation substudies. *J Clin Epidemiol* 41:1167-1174 (1988).
44. Schoenberg J, Klotz J, Wilcox HB, Nicholls GP, Gil-del-Real MT, Stenham A, Mason TJ. Case-control study of residential radon and lung cancer among New Jersey women. *Cancer Res* 50:6520-6524 (1990).
45. Brunekreef B, Noy D, Clausen P. Variability of exposure measurements in environmental epidemiology. *Am J Epidemiol* 125:892-898 (1987).
46. Beyea J, DeCicco J. Reestimating the noble gas releases from the Three Mile Island accident. *Health Phys* (submitted).
47. Doll R. Are we winning the fight against cancer? An epidemiological assessment. *Eur J Cancer* 26:500-508 (1990).
48. Davis DL, Hoel D, Lopez A. International trends in cancer mortality in France, West Germany, Italy, Japan, England and Wales, and the USA. *Lancet* 336:474-481 (1990).
49. Sorsa M, Ojajarvi A, Salomaa S. Cytogenetic surveillance of workers exposed to genotoxic chemicals: preliminary experiences from a prospective cancer study in a cytogenetic cohort. *Teratog Carcinog Mutagen* 10:215-221 (1990).
50. Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. *Stat Med* 8:431-440 (1989).
51. Cartwright R, Ahmad RA, Barham-Hall D, Glashan RW, Rogers HJ, Higgins E, Kahn MA. Role of N-acetyltransferase phenotypes in bladder carcinogenesis: a pharmacogenetic epidemiologic approach to bladder cancer. *Lancet* 2:842-845 (1982).
52. Lewalter J, Korallus V. Blood protein conjugates and acetylation of aromatic amines. *Int Arch Occup Environ Health* 56:179-196 (1985).
53. Phillips AM, Silbergeld EK. Health effects studies of exposure from hazardous waste sites—where are we today? *Am J Ind Med* 8:1-7 (1985).
54. Greenland S. The effect of misclassification in the presence of covariates. *Am J Epidemiol* 112:564-569 (1980).
55. Kupper LL. Effects of the use of unreliable surrogate variables on the validity of epidemiology research studies. *Am J Epidemiol* 120:643-648 (1984).
56. Khoury MJ, Flanders WD, Greenland S, Adams MJ. On the measurement of susceptibility in epidemiologic studies. *Am J Epidemiol* 129:183-190 (1989).
57. Dohrenwend BP, Dohrenwend BS, Kasl SV, Warheit GJ, Bartlett GS, Chisholm RF, Goldstein RL, Goldstein K, Martin JL. Report of the Public Health and Safety Task Group on Behavioral Effects. In: *Staff reports to the President's Commission on the Accident at Three Mile Island Public Health and Safety Task Force*. No 052-003-00732-1. Washington, DC: U.S. Government Printing Office, 1979; 257-308.
58. Dickman S. Chernobyl effects not as bad as feared. *Nature* 351:335 (1991).
59. Ozonoff D, Colten ME, Cupples A, Heeren T, Schatzkin A, Mangione T, Dresner M, Colton T. Health problems reported by residents of a neighborhood contaminated by a hazardous waste facility. *Am J Ind Med* 11:581-597 (1987).
60. Roht LH, Vernon SW, Weir FW, Pier SM, Sullivan G, Reed LJ. Community exposure to hazardous waste disposal sites: assessing reporting bias. *Am J Epidemiol* 122:418-433 (1985).