Annu. Rev. Public Health. 1995. 16:43–59

EXPOSURE ASSESSMENT ERROR AND ITS HANDLING IN NUTRITIONAL EPIDEMIOLOGY

Lenore Kohlmeier
Department of Nutrition and Epidemiology, School of Public Health, The University of North Carolina, Chapel Hill, North Carolina 27599-7400

Bärbel Bellach
Department of Health Risks and Prevention, Robert Koch Institute, General Pape Strasse 64, D 12101 Berlin, Germany

KEY WORDS: nutritional epidemiology, exposure assessment, energy adjustment, calibration, measurement error

ABSTRACT

Exposure assessment is the weakest element in nutritional epidemiologic studies. In the absence of an adequate arsenal of biomarkers of intake in the United States, food frequency questionnaires are widely used to assess habitual frequency of consumption of foods. These tools need to be designed for the population under study, based on prior information on the eating behavior of the population. The questions to be addressed to insure appropriate application of these tools are presented. The influence of various sources and types of measurement error on various scientific hypotheses is addressed. In assessment of nutrient adequacy, information on intra- to interindividual variation of the nutrient or substance of interest is essential. Risk assessment requires examination of sources and extent of bias and differential and nondifferential measurement error within the study. The theory required for error correction is well developed, but rarely carried out because of lack of software and lack of information needed to calibrate the measures.

0163-7525/95/0510-0043$05.00

INTRODUCTION

Research on the etiologic role of food-related exposures (both environmental and nutritional) in humans depends on accurate assessment of the exposure, in any of its dimensions. As we understand more about the influence of various aspects of diet-related exposures on the absorption, transport, metabolism, and excretion of other substances, as well as the ability of nutrients to influence gene transcription, most, if not all, epidemiological studies are interested in assessing some dimension of diet as a primary exposure or covariate. As a result, the nutritional epidemiologist is often asked for advice on a method suitable for the assessment of diet in an epidemiological study. The constraints are then outlined as a limit on the number of questions (5–10 preferably) to be self-administered in a paper-pencil questionnaire, and to take no longer than 10 minutes. Total diet should be assessed by this tool since the exact exposure and covariates of interest have not yet been identified.[1]

The range of food-borne exposures of interest include substances carried incidentally on foods, contaminants such as pesticides or heavy metals, chemicals added intentionally as additives, preservatives or fortification, nutrients and an ever-increasing list of nonnutrients produced by the plant or animal, food temperature upon consumption, preparation prior to consumption, packaging, as well as the sum of energy contributed to the diet by all foodstuffs consumed in a defined time period. The dimensions of interest include amount, frequency, and duration of consumption. The exposure most commonly desired is the dietary intake, as an exposure rate, integrated over many years. Collecting information on habitual dietary intake of individuals has long been recognized as a challenging problem in nutrition and epidemiology (5, 13, 22). Scientists wanting to add information about dietary exposure to their studies have been faced with a choice of applying expensive records, employing experienced dietitians, administering an available food frequency questionnaire, or searching for a suitable biomarker of exposure. Each approach has its strengths and weaknesses.

FOOD-BORNE EXPOSURE ASSESSMENT TODAY

Currently, the selection of methods is driven largely by practicability, with little expectation of major improvements (57). The dissatisfaction with currently available methods was evidenced by the popularity of the First International Meeting on Dietary Assessment Methods, held in Minnesota in 1992.

[1]Our advice to nutritional epidemiologists asked to construct such a set of questions is, before they are held responsible two years later when very unusual race-, gender-, or age-specific results are found, to just say no.

The bewildering inconsistencies in the results of studies of diet and diseases, such as those on fat and breast cancer, were attributed to tools that are too imprecise to characterize habitual dietary intake of individuals (47). This paper addresses the weaknesses in exposure assessment underlying the current conduct of nutritional epidemiology, and the degree of measurement error arising from our methods of assessing exposure.

Historical and International Perspective

Historically, in the United States, records of consumption were first applied to assess dietary intakes. As food compositional data became available in the 1930s the importance of individual nutrient intake assessment increased (30). Widdowson & McCance introduced a "precise weighing method" in 1936, which has remained the standard in the United Kingdom. Shortly afterward, Burke's diet history method was introduced as a way to shortcut the hand calculation of nutrient intake then current (12). One outcome was the development of food frequency questionnaires, introduced by Stefanik & Trulson in 1962 (48). This method dominates most of nutritional epidemiology in the United States (24). In Europe the situation is different: Computerized diet histories are used in national surveys and case-control studies in Germany (23); weighed records are used in the United Kingdom; and because of strong national preferences, a potpourri of methods is used in the large cohort on diet and cancer—the EPIC study (37).

Anatomy of a Food Frequency Questionnaire

Epidemiological research in the United States relies heavily on semiquantitative food frequency questionnaires for dietary assessment. There are many reasons why this method is preferred: It is designed to allow ranking of individuals whose habitual dietary behavior is known, at a low cost. The development of up-to-date food frequency questionnaires designed specifically for the population of interest in a particular study is hampered by lack of motivation of scientists to insure that methods valid for their population are being applied and scarce resources for developing and validating new methods. Consequently, existing questionnaires are often borrowed and administered uncritically in study populations whose dietary behavior differs substantially from those of the populations for whom the instruments were designed. They are frequently described as validated, implying thereby that they are valid for the purpose at hand. This practice may be due to ignorance of the anatomy of such questionnaires.

Unlike other methods of dietary assessment, food frequency questionnaires are designed specifically to assess variance in the frequency of intake of particular foods, using a minimal number of closed questions. Design decisions are made about which foods to include, which to group, which consumption-frequency levels to allow, what to set as a usual portion size, and how individual foods should be weighted in the development of the nutrient database.

The nutrient database consists of the foods that have been determined to contribute the most to the variance in intake of the nutrient of interest in the population under study (9). These individual food items are often grouped in a much smaller set of questions such as "Please fill in your average use, during the past year, of beef, pork, lamb as a sandwich or mixed dish, e.g. stew, casserole, lasagna, etc." These grouping of foods elicit changes in responses in various ethnic and gender groups (45). The emphasis remains on the frequency of consumption. The person responding to the questionnaire generally then adds up frequencies of consumption across foods and consumption of individual foods across meals—no mean task (44).

Nutrient calculations are based on specially constructed nutrient databases that apply weighted averages of the proportions of intakes of all foods covered by the questions. Determination of average nutrient values requires up-to-date information on the relative consumption of the individual items as a proportion of the group under question in the population of interest. For example, if middle-aged African-American men eat more pork than Caucasian men of the same age, the values for thiamin for the response of daily consumption of beef, pork, or lamb will be based on a different database that weighs pork more heavily than the amount for a daily response for Caucasians. These assumptions about relative consumption need to be checked regularly and updated if necessary.

Finally, it should be kept in mind that the food frequency questionnaire restricts true variance in intake because it limits the number of foods that subjects can report, and truncates the range of quantitative intakes. Generally, 5 to 9 categories of frequency are allowed, ranging from never to a few times per day, whereas more categories seem to improve the validity of the instrument (17). This restriction is responsible for what Beaton refers to as the loss of real variance with this tool due to reductionism and summation (6).

The epidemiologist should therefore consider the following questions before applying a food frequency questionnaire designed for another population:

1. Does this tool capture 80–90% of the interpersonal variance in consumption of the food-borne exposures under study? To do so, the questionnaire must be based on or compared with a recent survey of total diet assessed independently. This information must be conducted in the age, race, gender, ethnic, or religious group under study. The categories of frequency of consumption also need to be examined for appropriateness for the ethnic group understanding.
2. Do the nutrient values attributed to each response apply in this group of people? Individuals consuming more pumpkin pie will have incorrect carotene intakes if the nutrient data are based primarily on apple pie consumption. The nutrient value assigned to the questions should be a weighed

average of up-to-date nutrient information on all consumed items that the question subsumes, because foods differ in some or all of their dietary constituents.

3. Are there systematic biases in response between groups of people of interest? Is the accuracy of information captured, for example, in 50-year-old African-American men similar to that from 50-year-old Caucasian men? Do 70-year-old men respond with a different degree of errors to the same questions than 20-year-old men do? If so, applying the same food frequency questionnaire to all people in studies spanning such different groups can result in an artifactual effect of diet.
4. Are the assumptions about portion size appropriate for the subjects under study, as well as for this gender, age group, and population? Are a significant number of elderly persons, vegetarians, children, or Asians (for whom the assumptions on portion size are inappropriate) included in the study population? Use of a single set of portion sizes could result in inaccurate over- or underestimation of intakes.
5. If this tool is being used to monitor changes in intake of specific foods or nutrients over time, how will changes in diet over time be accounted for? How will the introduction of new foods into the market, changes in price (which affect consumption through changes in portion sizes or frequency of consumption), changes in the use of specific commodities in the food products (such as the oils used in margarine) be captured? Changes in relative consumption levels will need to be accounted for in the weighed estimates of nutrient averages in the nutrient database underlying the food frequency questionnaire.
6. Should a separate dietary assessment instrument or biomarker be administered in a subsample of the population to calibrate the results from the food frequency questionnaire and adjust for errors in it? This procedure has been recommended (8) and is becoming generally recognized (34, 35).

ERROR IN MEASUREMENT OF EXPOSURE

Sources of Error

Errors in nutritional epidemiological studies are partly attributable to the use of flawed dietary intake assessment tools, based on short time spans for the estimation of habitual diet. They include errors in remembering the foods consumed: Either foods are forgotten or phantom foods injected; the frequency of food consumption is inaccurately reported; and the portions consumed are incorrectly quantified. True intraindividual daily and seasonal fluctuations in intake, both of which are highly variable and based on patterns that differ from individual to individual (52), are also components of "error" in the measurement of habitual diet.

Part of the error derives from inadequate cognitive support on the part of the subject. From what is known about memory retrieval, it is unlikely that foods are stored in memory by themselves. They need to be retrieved from long-term memory into short-term memory by reconstruction of the period of time. The time period serves as a memory cue to what was eaten (4). The essential difference between methods of dietary assessment from a psychological point of view lies in the nature of the cues provided to the respondent to elicit information about usual diet (46). In recall and history methods, the cues are the period of time. Food frequency questionnaires use specific foods as the category of "event" that the subject is asked to recall. The better method is the one that best matches the content and organization of memory and the cognitive processes operating on memory representation and response generation.

Types of Error

Error in measurement is a problem in all observational sciences. Since measurement error potentially affects all statistical analysis, considerable attention has been paid to its effect on the properties of estimates and testing procedures. Fuller (20) and Thomas (53) describe this work in detail. From the special nutritional epidemiological point of view, the problem lies in determining the extent of error (including true intraindividual variability) in measures of dietary intakes and how this error affects estimates of adequacy or risk.

If, as is usually the case, measurement errors cannot be avoided, their effect on the estimated exposure-disease relationship should be taken into account. Measurement error can attenuate the observed compared with the true exposure-disease relationship; it can distort associations and interactions between covariates and outcomes; and variance can also be altered. Quantitative assessment of the effects of measurement error requires the structure and distribution of measurement error be known.

Epidemiological publications usually classify measurement errors into "random" and "systematic." Since systematic errors can be randomly distributed, we avoid this terminology and discriminate between unbiased and biased measurement methods and differential and nondifferential errors.

Let X' be an erroneous measurement of the exposure X, for instance, the average daily fat consumption as approximated by a 7-day-weighed dietary protocol. Measuring the daily dietary intake as a surrogate for usual intake results in an error term ε, which is a mixture of methodological error and true individual variation in intake from day to day.

X' is defined as an unbiased measurement of X if the average measurement (7-day-protocol results for fat intake) approaches the true measure as the number of samples increases (habitual fat intake for the individual). Unbiased measurements result for

$$X' = X + \varepsilon, \quad 1.$$

where ε is a random error variable with expectation 0. Any tendency to over- or underreport the intake of some foods should be adjusted for in each group in which it exists separately. The occurrence of underreporting of alcohol intake or the overestimation of fruit consumption indicates a biased method.

X' is a biased measurement when the average measurement does not approximate the true intake. The simplest case is when the expectation of the error variable ε from equation 1 does not equal 0.

Both biased and unbiased errors can be either differential or nondifferential. Measurement error is nondifferential as long as the error distribution is identical for all individuals of a study or for each subgroup of a population. This would occur in equation 1, when the distribution of the error variable ε would be the same for every individual under study. Measurement errors are differential if the participants of a study react differently to a measurement method that is used within a study. The distribution of the error term ε in equation 1 in this case differs between population groups under study. Take, for example, the case in which hospitalized subjects report a lower variance in their diets than population controls. This discrepancy may be reflective of their current diet, but not the true variance of their habitual diets, The variance and therefore the errors in this case are differential, even if the measurements are unbiased. Biased differential error occurs when obese individuals underestimate or underreport their fat intakes, whereas lean individuals report accurately or overreport.

The distinction between differential and nondifferential measurement error is important for error assessment, adjustment, and correction strategies. For nondifferential errors, the direction of influence on the estimated exposure-disease relationship is presumed to be biased toward zero. With some further information, often the bias can be corrected. Differential errors influence the estimated exposure-disease relations in ways that can be predicted only if there is information about the error in all subgroups under study. Thus, in situations where there are different (but unknown) exposure variances between cases and controls, for whatever reason, the odds ratios from logistic regression may give the appearance of a quadratic relationship between exposure and disease. Robertson et al (40) show an example in which it is impossible to distinguish between differential error and a true curved relationship. Further illustrations of the devastating effects of differential bias are given later.[1]

In summary, dietary intake is rarely, if ever, measured without error; errors can be biased or unbiased; both kinds can be differential or nondifferential. It is important to know the nature of the errors in order to correct the resulting parameter estimates and assess the confidence levels for study conclusions.

[1]Energy Adjustment Does Not Correct for Underreporting, B Bellach, L Kohlmeier, submitted.

Validation and Calibration

Validation of the ability of a method to capture habitual intake accurately is not possible. The gold standard for habitual consumption does not yet exist. Because most studies have no basis in "truth," they are therefore calibration and not validation studies. Calibration studies describe the difference between two methods, and provide information on the covariance structure of the errors between methods to allow correction. Furthermore, they are undertaken in order to adjust results between methods or with a single method used over time to a common baseline analogous to calibration in the clinical laboratory.

Assessment of the extent of measurement error inherent in a method is generally presented in validation papers as correlation coefficients between two methods, although this assessment is disputed as a valuable measure (7). Unadjusted correlation coefficients reported in several validation studies for a number of nutrients under various methods are presented in Table 1. They range from 0.21 to 0.74, when food frequency responses are compared with methods considered to be more reliable. Most of the calculated correlation coefficients are under 0.5 and are remarkably consistent for the same nutrient between studies; this consistency may be due to the extent of correlated errors between methods and the similarities in the methods being applied.

Other indices of "validity" include the kappa statistic, Kendall's Tau, and

Table 1 Correlation coefficients for food frequency questionnaire as compared to reference methods

Nutrients	Rimm[a] 1992	Posner[b] 1992	Pietinen[c] 1988	Longnecker[d] 1993	Willett[e] 1985	Block[f] 1990
Calories	.27	.42	.45	.45	.51	51
Fat	.42	.42	.33	.41	.51	.60
PuFA	.29	.29	.38	.60	.34	.48
MuFA	.46	.29	.30		.50	.59
Vitamin A	.45			.40	.35	.47
Iron	.32				.46	.47
Cholesterol		.30	.35	.35	.39	.55

[a] Unadjusted Pearson correlation coefficients between semiquantitative food frequency questionnaires and two one-week records in 157 men from Boston (39).

[b] Spearman rank correlation coefficients between food frequency questionnaires and three-day records in 77 men and 73 women of the Framingham Offspring (34).

[c] Unadjusted Pearson correlation coefficients between food frequency questionnaires and 24 days of food-consumption records (33) in 190 Finnish men.

[d] Unadjusted Pearson correlation coefficients for food frequency questionnaires and 6 days of diet records in 138 adults (28).

[e] Unadjusted Pearson correlation coefficients comparing food frequency results with the means of four one-week diet record (56).

[f] Unadjusted Pearson correlations of the food frequency responses as compared with means of three four-day records in 277 women (10). In this study the authors transformed the data to reduce skewness, which may explain the higher correlation coefficients.

Spearman's rank correlation coefficient. The combination of these three indices gives a better picture of the method's ability to accurately assess intake and to rank individuals (7).

Correlation coefficients as well as many other indices serve only as good summary measures of the performance within a population, since they depend on the range of intakes observed in these groups. For example, the lower correlation coefficients in women than in men can be interpreted either as a lack of precision in women or the result of their smaller range of intake (16). Correlation coefficients only reflect the degree of linear relationship between two methods. Results of regression analysis (estimated regression coefficients and their standard deviation) reflect much more the quantitative differences between methods (8).

To avoid misinterpretation of correlation coefficients and to enhance comparability of different validation or calibration studies, Delcourt et al (16) propose that the standard deviation of the differences between method measurements and the mean difference be used as an indicator of absolute agreement. They also suggest that standard deviation of the average be presented as an indicator of the range of intakes.

Deattenuation of correlation coefficients and energy adjustment of nutrient values are often undertaken in validation reports. Both of these manipulations tend to increase the strength of the relationship. It has been argued that the crude findings underestimate validity, because the method used as a standard inaccurately represents the truth. Careful studies by Plummer & Clayton (34, 35) reveal the existence of large correlated error between various dietary assessment methods and demonstrate that comparison studies that apply methods with correlated errors (where the subjects under- or overestimate with both methods) actually overestimate the validity of the method under study. These authors warn that subjects under close scrutiny may become rehearsed in their responses; a common component of correlated error can occur and thus affect all attempts to adjust for measurement error.

The evaluation of the validity of methods of dietary exposure assessment remains a problem. Since the methods are imperfect, judgment of the degree of validity that a method offers is flawed. Determination of the degree of validity that the measure applied should have to test specific hypotheses remains uncertain. The relationships between validity, error, efficiency, and study size and the effects on estimates are discussed later.

THE INFLUENCE OF MEASUREMENT ERROR

Questions posed by nutritional epidemiologists generally include some of the following (21):

1. How well nourished is a population? What is the nutritional adequacy of a population or subpopulation with regard to a specific nutrient?
2. Is there a relationship between a food-borne exposure and risk of a specific disease?
3. What is the quantitative relationship between unit of consumption of a food or food component under study, be it a biomarker or reported intake, and the outcome of interest?

From a statistical point of view, each of these questions demands different sets of information, sets different demands on the quality of the information, and requires different approaches to measurement and adjustment. We discuss these three components individually in the following sections.

Assessing Nutrient Adequacy or Toxicity

Many nutritional epidemiologists are primarily concerned with the evaluation of the status of a population, with regard to their dietary intakes and requirements. This evaluation involves (*a*) comparison of true exposures with some reference values to determine whether the population average is indicative of low risk in general, and (*b*) examination of the population distribution of exposure to determine the extent and nature of the part of the population at high risk (of either deficiency or excessive exposure). This process is analogous to toxicological risk assessment, and requires stable quantitative information on the true absolute levels of intake. Intraindividual variance is not relevant in the estimation of means or group averages. However, to estimate the prevalence of deficiency, intra- to interindividual variances need to be considered. The ratio of intra- to interindividual variance is not only nutrient or substance specific, but can differ by gender, age, ethnic group, and country. This ratio is also determined by the variety of foods available to the population. These considerations apply not only to dietary assessment methods, but also to biomarkers of exposure. For comparison sake, the extreme differences in intra- to interindividual ratios for different nutrients (21, 29, 55) are shown in Table 2.

These variances, and the level of efficiency that is acceptable (55), determine the number of repeated measures or days of intake to be included in the sampling plan. The desirable number of repeated measures is averaged per person and treated as the best individual estimate of exposure (intake). This estimate is used as the individual value in the population distribution of intakes. This new distribution is then used as a basis to estimate the size of the population at risk of high or low intakes. The probability approach (32) is better for determining the size of a population at risk than is using the percentage of individuals below a cutoff level. This approach uses the probability of deficiency at a given level of intake, and sums up the probabilities across

Table 2 Within- to between-person variance ratios

	24 days of records [a]	3 days of records [b]	7 days of records [c]
Energy	1.5	0.9	1.7
Protein	2.2	1.5	
Total fat	1.9	1.3	2.0
Saturated fat (S)	1.4	1.7	
Monosaturated fat	2.6		
Polyunsaturated fat (P)	1.5	2.6	
P/S quotient	1.0		2.6
Cholesterol	3.2	4.2	4.0
Vitamin A	4.6		
Vitamin C	2.9		
Vitamin E	1.6		
Calcium	1.5	.09	

[a] Ref. (21) [b] Ref. (55) [c] Ref. (29)

the population distribution. Knowledge of or assumptions about the risk of deficiency or excess at various levels of intakes are needed.

Assessment of Dietary Contributions to Risk of Disease

Whereas nutritionists are traditionally interested in adequacy, epidemiologists are traditionally interested in determining the presence of risk related to exposure. Accurate information on absolute intake levels is therefore not essential. The overriding need is to categorize individuals into groups of exposure with minimal misclassification. A clear differentiation of individuals exposed to high and low levels of the dietary factor of interest at the relevant time period is critical. It is only when the results of studies are being compared or pooled that the accuracy and quantification of exposure become important.

Measurement error, independently of its type, may lead to misclassification and attenuation in this type of analysis (6). Flegal showed that if the underlying hypothesis of a nutrient effect is true, unbiased, nondifferential measurement error of a metrical dietary exposure can result in differential misclassification of individuals into quintiles of this exposure (18). This misclassification results in an unpredictable attenuation of a corresponding estimation of risk. Brenner & Loomis (11) expand on Flegal's work by assessing the direction and magnitude of the resulting misclassification bias under several assumptions about the underlying nondifferential measurement error. Their simulation studies confirm that unbiased nondifferential measurement error leads to bias toward the null of the estimated exposure-disease relationship. They note further that this bias is less than the expected bias with nondifferential misclassification.

Biased nondifferential and nonrandom error (systematic over- or underestimation of exposure) biases the exposure-disease relationship either toward the null or away from the null. The direction depends on the underlying distribution of exposure, the true exposure-disease relationship, and the cut points used for categorization. These results stress the need for careful evaluation of possible effects of categorizing a continuous exposure variable that is erroneously measured.

Dietary assessment methods employed in epidemiological risk assessment should assess the exposure level or frequency for those foods contributing most to the explanation of variance in the outcome of interest in the population under study. If a method assesses this level well, misclassification of exposure will be minimal. The degrees of misclassification resulting from food frequency questionnaire responses for extreme quintiles have been reported in a few validation papers (38, 58). Misclassification of the extreme quintiles by one quintile ranges from 47 to 59%, depending on the nutrient of interest.

Walker & Blettner evaluated the degree of misclassification associated with various levels of correlation coefficients (54). Under the assumption that *M*, the method being evaluated, and *R*, the reference method, follow a bivariate normal distribution with a correlation coefficient ρ, they calculated that when ρ falls below 0.8, half of the population will be misclassified. At ρ equal to 0.5, two thirds of the population are misclassified. The loss of power to detect a difference can be compensated to some degree by increasing the sample size. Compensation for error by increasing size may allow detection of statistical significance, but does reduce the estimate of risk (54). Large increases in sample size are required to compensate in cohort studies for loss of power due to misclassification of disease risk. A 30-fold increase is needed to allow detection of a trend in risk across quintiles with correlation coefficients of 0.2 between *M* and *R* (54). If the results for validity of food frequency questionnaires presented in Table 1 are taken into account, the high degree of misclassification potential with these methods becomes obvious.

Quantitative Risk Assessments

Although most nutritional epidemiological analyses are currently based on risk detection, optimal levels or toxic levels need to be determined, and the increase or decrease in risk per unit of food or nutrient consumed needs to be quantified. Absolute intake levels are required to conduct such quantitative analyses of risk, just as with determination of adequacy. For some hypotheses, systematic error may not be relevant. One such example is when risk per increment of dietary substance is considered, e.g. the increase in risk of hypertension per mg of sodium consumed, without concern about the baseline. In other cases, both the intercept and the slope are important, which places the greatest challenge on the instrument of measurement. However, strategies to correct

measurement error are available for all types of quantitative epidemiological analyses, provided the requisite information on the nature of the error is available.

In nutritional epidemiology, exposure variables X considered in equation 1 are components of dietary intake. This intake can be the long-term, the average true intake, or the actual intake of some nutrient. The most frequently used mathematical model for the relationship between some disease measure d, exposures **X** and some other factors **Y** (confounders, effect modifiers) is

$$d = a + b\mathbf{X} + c\mathbf{Y}, \text{ (X and Y can be vectors).} \qquad 2.$$

For the sake of simplicity the relationship is assumed to be linear. The expectation d of a continuous variable in the classical linear regression model is related to a linear exposure expression. The logistic regression model assumes a linear relationship between the odds of disease occurrence and exposure for a binary disease indicator. In a Cox regression model the logarithm of the hazard function $h(t)$ is approximated by a linear function of exposure variables. The statistical theory of error correction applicable to each of these models has become highly developed in recent years. Because all these models fit into the theoretical concept of generalized linear models, the corrections for measurement errors based on these models can be applied (3, 14, 15, 26, 27, 31, 51, 50).

Armstrong (2) provides a simple illustration of procedures to correct measurement error in those models: It was assumed that the error in measurement error equation 1 is normally distributed with expectation 0 and variance σ_ε, independently of the true value of **X** (i.e. nondifferential unbiased measurement error). Under the further assumption that the true exposure **X** has a normal distribution in the observed population, he used the statistical result that in equation 2, parameter estimates of b are attenuated toward 0, and that the attenuation factor depends only on the variances σ_x of **X** and σ_ε of ε. Therefore, correcting for attenuation due to measurement error requires knowledge of the variances, which usually must be estimated. σ_ε can be estimated by a validation or calibration study, while the estimation of σ_x comes from the main study.

Unfortunately, the simple model for measurement error described above rarely applies exactly to data arising from epidemiological studies. Nevertheless, generalizations can be used to models with one or more erroneously measured exposure(s). Rosner et al (41) present a correction method for logistic regression when there is error (possibly correlated) in one or more covariates, and data are available from both a main study and a validation substudy. The assumption of multivariate normal distribution of all covariates as well as of the error variable is again needed. Because no gold standard exists for many exposures, these authors also propose (42) correction procedures whereby a reproducibility study instead of a validation study is used to obtain data to

estimate the error variable variance. In this case, the average of a large number of individual measurements by one method is considered to be the gold standard for an individual's true mean. The assumption here too is that error has to be nondifferential and unbiased.

A growing number of papers deal with measurement error correction in nutritional epidemiology. Some of the models used assume normal distributed errors (43, 49, 50), and some only the knowledge about the true covariance structure of the errors (26, 31). Each paper proposes a statistical procedure to correct the bias that occurs in estimating the parameter vectors **a, b, c** in equation 2 if nondifferential or known differential measurement errors exist. Unfortunately, these procedures have not yet been implemented into widely available statistical software.

There a few examples of the applications of these methods in the nutritional epidemiological substance area. Armstrong et al (1) studied the risk of colon cancer in relation to calories, protein, and fat intake. Using the same approach as Rosner et al (41), they adjusted for measurement error in the estimation of calories, protein, and fat. Estimation of error structure and distribution of the dietary assessment method came from a validation substudy (19). Adjustment for measurement error in each measure singly did not have a great impact on the estimates. However, the simultaneous corrections for variance in trivariate models resulted in strong enhancement of the estimate of risk induced by fat (odds ratio from 1.07 to 1.20) and the protective effect of protein (odds ratio from 0.80 to 0.47).

Rosner et al corrected risk estimates within the Framingham Heart Study, where the incidence of coronary heart diseases is related to several risk factors (42). Reproducibility data (e.g. blood pressure, serum cholesterol, serum glucose) were obtained from a subgroup of people seen at examination two or three times. After correcting for measurement error, estimated odds ratios comparing extreme quintiles of risk factors increased considerably (serum cholesterol 2.2 vs 2.9, serum glucose 1.3 vs 1.5, systolic blood pressure 2.8 vs 3.8).

Thomas et al (53) used a different approach to study the relationships between body mass index and intakes of total energy and fat in American men; in this study, 24-h recall data were available from all participants, and 7-day records were available from a subgroup. Instead of separately estimating uncorrected risks, estimating covariance structures, and recalculating a corrected risk estimation, they used the structural equation approach, fitting the model in only a single step. This approach requires a normal distribution of all variables and the linear relationship between them (53). The authors show that the model fit was much improved by incorporating the covariance of the measurements from the 24-h recalls and the 7-day records into the multivariate normal likelihood function to be maximized (53, p. 82). The usage of these

calibration data changes the maximum likelihood estimations of coefficient for energy intake from 0.518 to 0.267. The slope for the impact of total energy on BMI changes from 0.121 to –0.0120.

RECOMMENDATIONS

Measurement error is important in nutritional epidemiology. The goal of a study will dictate the required accuracy and thus the sample size and other aspects of the design. Since some data are virtually guaranteed to contain errors of measurement, the study design should attend to estimating the error distribution. The increasing number of procedures to correct for these errors means that these estimated error distributions can be used to make the estimated parameters more precise and increase confidence in the conclusions. As a consequence, epidemiological studies of food-borne exposures should incorporate calibration studies. The commission of the Federal Health Office in Germany of Nutritional Epidemiology (8) offers good recommendations for the design and analysis of nutritional epidemiological studies.

Literature Cited

1. Armstrong BG, Whittemore AS, Howe GR. 1989. Analysis of case-control data with covariate measurement error: application to diet and colon cancer. *Stat. Med.* 8:1151–63
2. Armstrong BG. 1990. The effects of measurement errors on relative risk regressions. *Am. J. Epidemiol.* 132:1176–84
3. Armstrong BG. 1985. Measurement error in the generalized linear model. *Community Stat.* 14:529–44
4. Baranowski T, Domel SB. 1994. A cognitive model of children's reporting of food intake. *Am. J. Clin. Nutr.* 59:212S–7S (Suppl.)
5. Barrett-Connor E. 1991. Nutrition epidemiology: how do we know what they ate? *Am. J. Clin. Nutr.* 54:182S–7S
6. Beaton GH. 1994. Approaches to analysis of dietary data: relationship between planned analyses and choice of methodology. *Am. J. Clin. Nutr.* 59:253S–61S (Suppl.)
7. Bellach B. 1993. Remarks on the use of Pearson's correlation coefficient and other association measures in assessing validity and reliability of dietary assessment methods. *Eur. J. Clin. Nutr.* 47: S42–S45 (Suppl. 2)
8. BGA Commission on Nutritional Epidemiology. 1993. Recommendations for the design and analysis of nutritional epidemiologic studies with measurement errors in the exposure variables. *Eur. J. Clin. Nutr.* 47:S53–S57
9. Block G, Hartman AM, Dresser CM, Carroll MD, Gannon J, Gardner L. 1986. A data-based approach to diet questionnaire design and testing. *Am. J. Epidemiol.* 124:453–69
10. Block G, Woods M, Potosky A, Clifford C. 1990. Validation of a self-administered diet history questionnaire using multiple diet records. *J. Clin. Epidemiol.* 43:1327–35
11. Brown CC, Kipnis V, Freedman LS, Hartman AM, Schatzkin A, Wacholder S. 1994. Energy adjustment methods for nutritional epidemiology: the effect of categorization. *Am. J. Epidemiol.* 139 (3):323–38
12. Burke BS. 1947. The dietary history as

a tool in research. *Can. J. Public Health* 23:1041
13. Cameron ME, van Staveren WA, eds. 1988. *Manual on Methodology for Food Consumption Studies.* New York: Oxford Univ. Press
14. Carroll RJ. 1989. Covariance analysis in generalized linear measurement error models. *Stat. Med.* 8:1075–93
15. Carroll RJ, Wand MP. 1991. Semiparametric estimation in logistic measurement error models. *J. R. Statist. Soc. B* 53:573–85
16. Delcourt C, Cubeau J, Balkau B, Papoz L. 1994. Limitations of the correlation coefficient in the validation of diet assessment methods. *Epidemiology* 5:518–24
17. Feskanich D, Marshall J, Rimm EB, Litin LB, Willett WC. 1994. Simulated validation of a brief food frequency questionnaire. *Ann. Epidemiol.* 4:181–87
18. Flegal KM, Keyl PM, Nieto FJ. 1991. Differential misclassification arising from nondifferential errors in exposure measurement. *Am. J. Epidemiol.* 134: 1233–44
19. Friedenreich CM, Howe GR, Miller AB. 1991. An investigation of recall bias in the reporting of past food intake among breast cancer cases and controls. *Ann. Epidemiol.* 1:439–53
20. Fuller WA, ed. 1987. *Measurement Error Models.* New York: Wiley
21. Hartman AM, Brown CC, Palmgren J, Pietinen P, Verkasalo M, Myer D, et al. 1990. Variabiliy in nutrient and food intakes among older middle-aged men: implications for design of epidemiologic and validation studies using food recording. *Am. J. Epidemiol.* 132:999–1012
22. Kohlmeier L. 1988. Analytical problems in nutritional epidemiology. In *Epidemiol. Nutr. Health Proc. Berlin Meet. Nutr. Epidemiol., 1st,* ed. L Kholmeier, E Helsing, pp. 9–18
23. Kohlmeier L. 1994. Gaps in dietary assessment methodology: meal- vs list-based methods. *Am. J. Clin. Nutr.* 59: 175S–9S (Suppl.)
24. Kohlmeier L, ed. 1994. *Dietary Assessment Resource Manual.* J. Nutr. 124: 2245S–317S
25. Kohlmeier L, Mensink GBM, Rehm J, Hermann-Kunz E, Haussler A, Maierhofer B. 1990. Development of a fully automated tool for nutrition surveys: diet history. *Annu. Rep. Fed. Health Off. (Bundesgesundheitsamt), Prog. Rep.,* pp. 226–28. Munich: MMV Med. Publ. (In German)
26. Liang K-Y, Liu XH. 1991. Estimating equations in generalized linear models with measurement error. In *Estimating Functions,* ed. VP Godambe, pp. 47–64. Oxford: Clarendon
27. Liu X, Liang K-Y. 1991. Adjustment for non-differential misclassification error in the generalized linear model. *Stat. Med.* 10:1197–211
28. Longnecker MP, Lissner L, Holden JM, Flack VF, Taylor PR, et al. 1993. The reproducibility and validity of a self-administered semiquantitative food frequency questionnaire in subjects from South Dakota and Wyoming. *Epidemiology* 4:356–65
29. Marr JW, Heady JA. 1986. Within- and between-person variation in dietary surveys: number of days needed to classify individuals. *Hum. Nutr. Appl. Nutr.* 40A: 347–64
30. Medlin C, Skinner JD. 1988. Individual dietary intake methodology: a 50-year review of progress. *J. Am. Diet Assoc.* 88:1250–56
31. Nakamura T. 1990. Corrected score function for errors-in-variables models: methodology and application to generalized linear models. *Biometrics* 77: 127–37
32. National Research Council Subcommittee on Criteria for Dietary Evaluation. 1986. *Nutrient adequacy: assessment using food consumtion surveys.* Washington, DC: Natl. Acad. Press
33. Pietinen P, Hartman AM, Haapa E, Räsänen L, Haapakoski J, et al. 1988. Reproducibility and validity of dietary assessment instruments. II. A qualitative food frequency questionnaire. *Am. J. Epidemiol.* 128:667–76
34. Plummer M, Clayton D. 1993. Measurement error in dietary assessment: an investigation using covariance structure models. Part 1. *Stat. Med.* 12:925–35
35. Plummer M, Clayton D. 1993. Measurement error in dietary assessment: an investigation using covariance structure models. Part 11. *Stat. Med.* 12:937–48
36. Posner BM, Martin-Munley SS, Smigelski C, Cupples LA, Cobb JL, et al. 1992. Comparison of techniques for estimating nutrient intake: the Framingham Study. *Epidemiology* 3:171–77
37. Riboli E. 1992. Nutrition and cancer: background and rationale of the European prospective investigation into cancer and nutrition (EPIC). *Ann. Oncol.* 3:783–91
38. Rimm EB, Giovannucci EL, Stampfer MJ, Colditz GA, Litin LB, Willett WC. 1992. Authors' response to "Invited commentary: some limitations of semi-

quantitative food frequency questionnaires". *Am. J. Epidemiol.* 135:1133–36
39. Rimm EB, Giovannucci EL, Stampfer MJ, Colditz GA, Litin LB, Willett WC. 1992. Reproducibility and validity of an expanded self-administered semi-quantitative food frequency questionnaire among male health professionals. *Am. J. Epidemiol.* 135:1114–26
40. Robertson C, Boyle P, Hsieh C-C, Macfarlane GJ, Maisonneuve P. 1994. Some statistical considerations in the analysis of case-control studies when the exposure variables are continuous measurements. *Epidemiology* 5:164–70
41. Rosner B, Spiegelmann D, Willett WC. 1990. Correction of logistic relative risk estimates and confidence intervals for random within-person measurement error: The case of multiple covariates measured with error. *Am. J. Epidemiol.* 132:734–45
42. Rosner B, Spiegelmann D, Willett WC. 1992. Correction of logistic relative risk estimates and confidence intervals for random within-person measurement error. *Am. J. Epidemiol.* 136:1400-13
43. Schafer DW. 1987. Covariate measurement error in generalized linear models. *Biometrics* 74:385–91
44. Sempos CT, Briefel RR, Flegal KM, Johnson CL, Murphy RS, Woteki CE. 1992. Factors involved in selecting a dietary survey methodology for national nutrition surveys. *Austr. J. Nutr. Diet* 49:96–104
45. Serdula MK, Byers T, Coates R, Mokdad A, Simoes EJ, Eldridge L. 1992. Assessing consumption of high-fat foods: the effect of grouping foods into single questions. *Epidemiology* 3: 503–8
46. Smith AF. 1991. Cognitive processes in long-term dietary recall. In *Natl. Cent. Health Stat. Vital Health Stat.* Hyattsville MD: US Dep. Health Hum. Serv. 6, 4:1–14
47. Stamler J. 1994. Assessing diets to improve world health: nutritional research on disease causation in populations. *Am. J. Clin. Nutr. Suppl.* 59:146S–56S
48. Stefanik PA, Trulson MF. 1962. Determining the frequency intakes of foods in large group studies. *Am. J. Clin. Nutr.* 11:335
49. Stefanski LA. 1985. The effects of measurement error on parameter estimation. *Biometrics* 72:583–92
50. Stefanski LA. 1989. Correcting data for measurement error in generalized linear models. *Comm. Stat.* 18(A):1715–34
51. Stefanski LA, Carroll RJ. 1985. Covariate measurement error in logistic regression. *Ann. Stat.* 13:1335–51
52. Tarasuk V, Beaton GH. 1991. Day-to-day variation in energy and nutrient intake: evidence of individuality in eating behaviour. *Appetite* 18:43–54
53. Thomas D, Stram D, Dwyer J. 1993. Exposure measurement error: influence on exposure-disease—relationships and methods of correction. *Annu. Rev. Public Health* 14:69–93
54. Walker AM, Blettner M. 1985. Comparing imperfect measures of exposure. *Am. J. Epidemiol.* 121:783–90
55. Wassertheil-Smoller S, Davis BR, Breuer B, Change CJ, Oberman A, Blaufox MD. 1993. Differences in precision of dietary estimates among different population subgroups. *Ann. Epidemiol.* 3:619–28
56. Willett WC. 1987. Implications of total energy intake for epidemiologic studies of breast and large-bowel cancer. *Am. J. Clin. Nutr.* 45:354–60
57. Willett WC. 1994. Future directions in the development of food-frequency questionnaires. *Am. J. Clin. Nutr.* 59: 171S–74S (Suppl.)
58. Willett WC, Sampson L, Stampfer MJ, Rosner B, Bain C, et al. 1985. Reproducibility and validity of a semiquantitative food frequency questionnaire. *Am. J. Epidemiol.* 122:51–65