Is It Necessary to Correct for Measurement Error in Nutritional Epidemiology?

Relationships between diet and chronic disease have become the focus of many analytic studies in nutritional epidemiology over the past several decades. Given often relatively limited variation in dietary intake within study populations, the results of observational studies critically depend on accurate assessment of the dietary exposure and other adjustment covariates, especially in view of the moderate-sized relative risks seen in many nutrition studies. Error in measuring exposure leads to a biased and inefficient estimate of the relationship of the exposure to disease (1). In this issue, Beulens and colleagues (2) present a prospective analysis of the relationship between alcohol intake, which participants self-reported using a food-frequency questionnaire, and cardiovascular events among men with hypertension in the Health Professionals Follow-Up Study. The authors report the findings before and after correction for measurement error, which provides a welcome opportunity to comment on the importance of measurement error in nutritional epidemiology.

Measurement error in covariate assessment of prospective studies is generally assumed to be nondifferential with respect to disease, which means that the relationship between true and reported diet is the same for persons who develop the disease as for those who do not. If this assumption is true, most authors in epidemiologic literature assume that measurement error attenuates the true relative risk—that is, it biases its estimate toward no effect. Consequently, authors may interpret null results as perhaps being due to exposure measurement error. In a similar manner, because of measurement error, the true relative risk in a statistically significant association should, if anything, be larger than that reported. This attenuation of effect always occurs when only the main exposure is measured with error, and this error follows the “classical model”: It is additive; it is independent of the size of the true exposure; and it has a mean of zero and a variance that does not change with the size of the true exposure (1).

Measurement error that occurs when assessing a person’s usual diet with a food-frequency questionnaire does not follow the classical model. Instead, it generally involves bias related to true intake in addition to random variation (3). Bias related to true intake often occurs as the “flattened-slope phenomenon” (consumers with high levels of intake tend to underreport and low-level consumers tend to over-report), which biases relative risk upward (4). On the other hand, random between-person and within-person variation biases the relative risk downward. The effect of random variation is more important in most situations, with the overall effect leading to an attenuated relative risk when a single error-prone exposure is present (5). In this situation, the nominal significance level of statistical tests of the exposure effect remains unaffected by the measurement error so that the usual significance tests of the relative risk are valid tests of the exposure–disease relationship. Nevertheless, measurement error causes statistical power to fall by a factor that increases as the amount of measurement error increases (6). This means that studies become less able to detect exposure–disease relationships as measurement error increases.

The effect of measurement error becomes much more complex if the disease model includes more than 1 exposure measured with error. This factor is important to consider because most analyses of diet–disease relationships, as in Beulens and colleagues’ study (2), include adjustment for energy intake (7) and other potentially error-prone adjustment covariates. In such cases, error in the main exposure attenuates the observed association with disease, and error in the adjustment covariates “contaminates” the observed association because of residual confounding (3). As a result, the bias in the estimated effect could be in either direction and not simply toward the null. Moreover, uncorrected statistical tests no longer accurately measure the degree of significance, making it impossible to reliably interpret the observed diet–disease association. Adjustment for multivariate measurement error is therefore necessary not only for unbiased estimation of the true diet–disease relationship (for example, the relative risk), but also for accurately testing whether the relative risk statistically significantly differs from 1.0.

To adjust for measurement error, Beulens and colleagues (2) applied the linear regression calibration approach, which is a state-of-the-art method in nutritional epidemiology (8, 9). In principle, this method requires performing accurate measures of nutritional intake in a calibration substudy and fitting multiple regressions of each true (that is, accurately measured in the calibration subsample) exposure variable on the error-prone and precisely measured covariates included in the disease model. The slopes in these multiple regressions are then applied to correct the observed relative risks in the main study. Because usual dietary intakes are impossible to measure accurately in free-living populations, the regression slopes in the calibration subsample may still be estimated by using reference measurements of usual intakes in place of their exact values. These reference measurements may contain errors of their own, but those errors should ideally be purely random, independent of the true covariates and errors in the reported intakes on the food-frequency questionnaire (3). As we explain below, the availability of such reference measurements in nutritional studies is limited.

To date, epidemiologic studies have used more intensive self-report methods as the reference instrument for the...
calibration subsample, for example, multiple-day food records, as in the Health Professionals Follow-Up Study (2), or several recalls of food intake in the previous 24 hours. Cumulative evidence from studies using “recovery” biomarkers, such as doubly-labeled water, as a reference standard for total energy intake (recovery biomarkers do meet requirements for a reference instrument [10]) shows that these intensive self-report reference methods can have systematic errors that are correlated with food-frequency questionnaire errors, therefore violating a key requirement of reference instruments (11). Unfortunately, reliable recovery biomarkers are available for only a few dietary factors—in addition to doubly labeled water. Urinary nitrogen is a marker for protein intake, and urinary potassium may be a marker for dietary potassium intake (10). Therefore, in most studies, the use of self-report instruments in a calibration subsample as the reference standard is currently our best, although imperfect, approach to adjustment for measurement error.

In their uncorrected analysis of the association between alcohol consumption and myocardial infarction, Beulens and colleagues (2) found a statistically significant relative risk of 0.85 ($P < 0.001$). After adjustment for measurement error in most dietary covariates and body mass index, the relative risk was 0.68 but with only borderline statistical significance ($P = 0.051$). We expect confidence intervals to widen after measurement error correction, partly because of added uncertainty in the estimated regression calibration slopes as a result of taking measurement error into account (9). Had the authors considered all error-prone variables (including trans and $\omega$-3 fatty acids and physical activity) in the same analysis, the estimated confidence intervals would have been even wider (12). Larger calibration subsudies, such as in the European Prospective Investigation into Cancer and Nutrition (EPIC) (13) and the National Institutes of Health–AARP [American Association of Retired Persons] Diet and Health Study (14), allow a more accurate calibration and, together with selection of more parsimonious diet–disease models, help limit the extent to which confidence intervals become wider after correction for measurement error.

Furthermore, the linear regression calibration method assumes that regressions relating reference measurements of intake to reported intakes on the food-frequency questionnaire are linear and homoscedastic. Meeting this condition often requires transformation of the involved variables (for example, log transformation). Moreover, we need a new method for foods that may be episodically or never consumed by a substantial proportion of individuals, alcohol being an important example (15). Beulens and colleagues do not discuss how they dealt with reports of no alcohol consumption and skewed distributions of positive alcohol intake and intakes of other dietary factors (for example, energy) in the disease model (2), which could affect the adjusted relative risks for cardiovascular outcomes and their corresponding confidence intervals.

Although current execution of regression calibration may not fully correct for measurement error, it does express some of the uncertainty arising from measurement error and therefore remains an important practice. To improve our understanding of dietary measurement error, we need to identify new recovery biomarkers and to investigate “concentration” biomarkers (for example, serum carotenoids), which do not meet the requirements for reference measurements but do correlate with intake (10, 16). We need new methods for combining concentration markers with dietary-assessment methods (17).

Our discussion of measurement error in nutrition studies should convince the reader to exercise caution when interpreting relationships between diet and clinical outcomes. Owing to the current limitations of available procedures and reference instruments, we cannot assume that corrected estimates of diet–disease associations in any single study are definitive. To firmly establish a hypothesis, we need carefully conducted studies in diverse populations with different dietary patterns and ranges of intake, incidence rates, and sociocultural histories. In addition, we must use different dietary assessment and reference instruments, which may result in different measurement error structures and different methods of error correction.

The uncertainties surrounding measurement error should send a strong message to those who formulate recommendations about nutrition. With unexplained inconsistencies in results, the discretion of silence may be preferable to the valor of setting recommendations. Conversely, when several studies of the same topic have consistent, robust results, we can be confident that the association in question is strong enough to withstand the impact of measurement error. Results that satisfy these requirements can be a basis for future dietary guidelines.

Anne C.M. Thiébaut, PhD
National Cancer Institute
Bethesda, MD 20892-7232

Laurence S. Freedman, PhD
Gertner Institute for Epidemiology and Health Policy Research
Tel Hashomer 52161, Israel

Raymond J. Carroll, PhD
Texas A&M University
College Station, TX 77843-3143

Victor Kipnis, PhD
National Cancer Institute
Bethesda, MD 20892-7354

Grant Support: Dr. Carroll’s research was supported in part by a grant from the National Cancer Institute (CA-57030) and the Texas A&M Center for Environmental and Rural Health via a grant from the National Institute of Environmental Health Sciences (P30-ES09106).

Potential Financial Conflicts of Interest: None disclosed.
Requests for Single Reprints: Victor Kipnis, PhD, National Cancer Institute, Biometry Research Group, Division of Cancer Prevention, Bethesda, MD 20892-7354; e-mail, kipnisv@mail.nih.gov.

Current author addresses are available at www.annals.org.

References
Current Author Addresses: Dr. Thiébaut: National Cancer Institute, Nutritional Epidemiology Branch, Division of Cancer Epidemiology and Genetics, Bethesda, MD 20892-7232.
Dr. Freedman: Gertner Institute for Epidemiology and Health Policy Research, Biostatistics Unit, Sheba Medical Center Tel Hashomer 52161, Israel.
Dr. Carroll: Texas A&M University, Department of Statistics, Blocker Building Room 447; College Station, TX 77843-3143.
Dr. Kipnis: National Cancer Institute, Biometry Research Group, Division of Cancer Prevention, Bethesda, MD 20892-7354.