Impact of Exposure Measurement Error in Nutritional Epidemiology

Victor Kipnis, Laurence S. Freedman

For the past several decades, observational studies of diet and cancer have yielded many inconsistent results (1). Given the limited variation in dietary intakes within many study populations and the seemingly modest diet–cancer associations, results of such studies depend critically on an accurate assessment of the dietary exposure (2). Measurement error in exposure leads to seriously biased relative risks of cancer for dietary intakes and substantially reduces the statistical power to detect existing relationships (2). The international consortium of cohort studies known as the Pooling Project (3) aims to gain precision in relative risk estimates and overcome this loss of statistical power by combining analyses of individual data from multiple studies that examine associations between diet and cancer. In this issue of the Journal, Lee et al. (4) report the results of an analysis of fat, protein, and meat consumption in relation to the risk of renal cell cancer in 13 cohorts within the Pooling Project. Using multivariable models, they obtained null results for all of the associations they considered. In this editorial, we comment on the impact of exposure measurement error on the estimated associations between diet and disease in this and other studies.

Exposure measurement error in prospective studies can usually be assumed to be nondifferential with respect to disease, that is, the relationship between the reported diet and the true diet is the same for the case subjects and the control subjects. When the exposure, but none of the confounders, is measured with error, this error attenuates the relative risk (i.e., biases the relative risk toward no effect). Although statistical tests for the association remain valid, that is, they preserve the nominal statistical significance level (5), the measurement error reduces the statistical power to detect the association. Consequently, the number of cases required to maintain the nominal statistical power is increased by a factor equal to the inverse-squared correlation between the reported exposure and the true exposure (6). Moreover, although adjustment for measurement error can remove existing bias in estimated relative risks, it cannot compensate for the loss of statistical power.

Within the Pooling Project, the main dietary assessment instruments were study-specific food-frequency questionnaires (FFQs) and closely related diet histories. Most studies also included validation subsamples with multiple food records or 24-hour dietary recalls as more precise “reference” instruments. Lee et al. (4) report that estimated correlations between FFQ-reported intakes and true intakes for fat, fatty acids, and protein were 0.4 or greater when those reference instruments were used. However, the use of self-reported reference instruments in validation studies often overestimates such correlations (7). But even if the true values of correlations between the reported and the true intakes were indeed 0.4, the loss of statistical power would be equivalent to an effective reduction in the number of renal cell cancer cases by a factor of 1/(0.4)^2 or 6.25, or, from a total of 1478 cases to 236 cases. This effective number of cases would give a statistical power of 90% to detect a true relative risk of 1.7 or more between the highest and lowest quintiles of intake. More moderate but nevertheless important associations could have been missed.

Lee et al. (4) also reported that some statistically significant associations in the age-adjusted models became weaker and non–statistically significant in models that controlled for potential confounders, especially body mass index (BMI), fruit and vegetable intake, and alcohol intake. It is interesting that none of these three potential confounders are ordinary adjusting covariates. BMI may mediate the relationship of fat or protein intake and renal cell cancer, and adjusting for a mediator is not recommended (8). Intakes of alcohol and fruits and vegetables are also susceptible to measurement error, and adjusting for such error-prone confounders introduces additional problems in the analysis of the association between the main exposure and disease, as we explain in the next paragraph.

The effects of measurement error increase in complexity when confounders are also measured imprecisely. Although estimated relative risks are still attenuated by error in the main exposure, they are further “contaminated” by errors in the confounders (i.e., residual confounding) (5,9). As a consequence, estimated relative risks can be attenuated or inflated by any magnitude. Moreover, unadjusted statistical tests no longer preserve the nominal statistical significance level, so that the observed diet–disease association cannot be reliably interpreted. Adjusting for multivariable measurement error becomes necessary not only to remove bias from estimated relative risks but also to allow proper testing of their statistical significance.

Lee et al. (4) adjusted for measurement error by applying linear regression calibration (10,11). This method gives valid results if the reference instrument used in the validation study is error free or has random errors that are independent of the true exposure, the confounders, and errors in the FFQ (9). However, studies (7,12) that have used biomarkers, which themselves meet the above requirements (the so-called “recovery” biomarkers) (13), show that
food records and 24-hour recalls, such as those used as reference instruments in the Pooling Project, can have errors that are correlated with FFQ errors. Because conventional reference instruments violate a key requirement of reference instruments, their use in regression calibration may not fully correct for measurement error.

Unfortunately, in most cases, recovery biomarkers cannot replace conventional reference measurements because they are currently available for only a few dietary intakes—that is, doubly labeled water for total energy intake, urinary nitrogen for protein intake, and perhaps urinary potassium for potassium intake (13). Lee et al. (4) cite the Observing Protein and Energy Nutrition (OPEN) study (7) as supporting their claim that using 24-hour recalls as a reference instrument for energy-adjusted nutrients produces results similar to those using recovery biomarkers. However, this result was observed in the OPEN study for men only; attenuation for women was underestimated by 59% using 24-hour recalls (7). Also, the OPEN results pertained to a simple energy-adjusted analysis and may not hold for a multivariable analysis of measurement error. Nevertheless, using imperfect reference instruments to adjust for measurement error often represents the best available option. This adjustment conveys at least some of the extra uncertainty caused by the error and should therefore remain an important practice for understanding true diet–disease associations. However, to improve this practice, we need to take several additional steps.

First, regression calibration requirements should be fully met. The multivariable disease model used by Lee et al. included at least seven confounders (six for men), three of which were dietary intakes (total energy, alcohol, and fruit and vegetable) (4). Only age, BMI, energy intake, and alcohol intake were included in their regression calibration models. This abbreviated calibration could introduce further bias.

Second, linear regression calibration requires that reference measurements be linearly related to FFQ-reported intakes and that the residuals in these regressions have constant variance. Often, it is necessary to transform the variables to meet these conditions. Moreover, for food items that are generally not consumed every day—for example, alcohol—a completely new methodology is needed (14). Lee et al. do not report how they dealt with reports of zero alcohol consumption or with skewed distributions of total energy; such decisions could affect the adjusted relative risks.

Third, a regression calibration that is based on an imperfect reference instrument should be accompanied by a sensitivity analysis to evaluate the robustness of the results to the possible correlation between reference instrument errors and FFQ errors. Results from recovery biomarker studies may suggest plausible correlations to be investigated in such analyses.

Fourth, many studies include “concentration” biomarker measurements (eg, serum concentrations of carotenoids) that cannot be used as references because they do not provide a direct measure of intake, but are nonetheless correlated with the intake of interest (13,15). We need new methodologies for combining these measurements with FFQs to improve dietary assessment (16).

In summary, continued caution is required when interpreting associations, or the lack thereof, between dietary factors and disease. Despite the considerable efforts of Lee et al. (4) to account for measurement error, this caution also applies to their study. Given the current limitations of available self-report instruments used for measuring diet or as references and the paucity of recovery biomarkers, it will be rare when we will be able to regard the results of individual nutritional epidemiology studies, even after correction for measurement error, as definitive. Only when multiple studies with different designs in diverse populations produce consistent, robust results, can the evidence regarding an association be sufficiently persuasive.

References