Invited Commentary: OPEN Questions

Walter Willett1,2,3

1 Department of Nutrition, Harvard School of Public Health, Boston, MA.
2 Department of Epidemiology, Harvard School of Public Health, Boston, MA.
3 The Channing Laboratory, Department of Medicine, Harvard Medical School, Boston, MA.

Received for publication March 14, 2003; accepted for publication March 20, 2003.

Abbreviation: OPEN, Observing Protein and Energy Nutrition.

The papers by Subar et al. (1) and Kipnis et al. (2) in this issue of the Journal provide a unique set of data and a useful contribution to nutritional epidemiology. The Observing Protein and Energy Nutrition (OPEN) Study they report on was based on a reasonable sample of participants, had a remarkably high completion rate, and is the largest known study to use doubly labeled water as the standard for comparison. In addition, the measurements of doubly labeled water, which can be highly error prone (3), appear to have been made with a high degree of technical precision. The study also has several limitations inherent to using this approach, including that it enabled evaluation of only total energy and one specific nutrient, protein. Furthermore, the results cannot be readily generalized to other food frequency questionnaires because the values for total energy intake estimated by use of their questionnaire were substantially lower than typically provided by other full-length questionnaires in current use. In addition, although great expense was devoted to biomarker measurements, only 2 days of 24-hour diet recalls were used, many fewer days of assessment than are included in usual questionnaire validation studies.

A more serious flaw in the design of the OPEN Study (1, 2) is failure to include a realistic measure of the within-person variability of their standard, doubly labeled water plus urinary nitrogen. This measure should have been obtained for a sample of participants by collecting a replicate measure a number of months apart, say roughly 6 to 12, which would be consistent with the 1-year time frame of their questionnaire. In their study, the only measure of reproducibility was for 25 participants, without any interval between the replicates; the authors report a coefficient of variation of 5.1 percent total energy expenditure, suggesting little within-person variation. However, this is likely to underestimate biologic variability substantially in relation to the time frame of the dietary questionnaire. Increasing variability given a greater time interval would be expected, for example, because persons are likely to change their physical activity with time because of differences in seasons and other reasons (the fact that the authors note no average seasonal variation in total energy expenditure is not relevant to individual variability). This phenomenon is well documented in a quantitative review of the reproducibility of total energy expenditure measured by using doubly labeled water in 25 studies with repeated measurements (4). In this report, Black and Cole found that, overall, the within-person coefficient of variation was of similar magnitude to the between-person variation whether or not the studies involved an intervention, and it clearly increased with increasing interval between repeated measurements. They estimated that the within-person coefficient of variation was approximately twice as great at an interval of 1 year compared with having no interval between replicates.

The lack of a realistic measure of variability in their standard method (1, 2), which is surprising for a study focusing on error, seriously limits the value of the findings and is likely to have led to misleading conclusions. As one example, Subar et al. (1) emphasize the importance of “intake-related bias” associated with their dietary questionnaire or 24-hour recalls, meaning that the ratio of intake measures divided by biomarker measures is inversely correlated with the biomarker measurements. Because any unaccounted-for variation (random within-person error) in the biomarker would appear in the denominator of the ratio and the numerator of the biomarker variable, the inverse correlation could just be all or part artifact.

Effects of adjusting for energy intake

The authors (1, 2) appropriately provide data for energy-adjusted protein intake (as protein density = protein intake/
total energy expenditure) in addition to absolute intakes, which is an advance beyond previous studies using urinary nitrogen excretion to estimate protein intake. In nutritional epidemiology, the focus is primarily on the composition of diets, as represented by energy-adjusted nutrient intakes; individual persons or populations can change their intakes of specific nutrients appreciably only by changing the composition of their diet.

In the absence of changes in physical activity, even small long-term changes (a few percentage points) in the total quantity of food consumed, as represented by total energy intake, would have important effects on body weight (and both physical activity and body weight have major health effects themselves). Total energy itself is thus of little interest in nutritional epidemiology because, to be useful as a measure of energy balance, it would need to be measured with unattainably extreme precision. Even highly precise measurements would not be useful for this purpose unless energy expenditure were measured with similarly high precision. In part because total energy is the only nutrient for which intake is tightly regulated and thus essentially out of individual control independent of weight and physical activity, the range of true intakes will be highly constrained, which will further add to difficulties in measurement. Because food frequency questionnaires primarily assess the relative mix of foods consumed, they are not likely to provide a good estimate of total energy intake. Nevertheless, total energy intake assessed by dietary questionnaires can still be useful as an indicator of over- or underreporting food intake because, in the calculation of energy-adjusted intakes, to the extent that the errors for energy and nutrient intakes are proportional or correlated, they will “cancel” each other (5).

In reality, errors in estimates of energy and specific nutrients tend to be highly correlated (as high as 0.9) because intakes are estimated from the same foods; high between-nutrient correlations in errors were also seen in the OPEN Study (refer to panels c and f of figures 2 and 3 in the paper by Subar et al. (1)). Many studies have provided evidence that adjustment for total energy can reduce errors; adjustment usually increases the strength of associations between intakes calculated from food frequency questionnaires and intakes from diet records, blood biomarkers, and disease outcomes (6, 7). Probably the most important contribution of the OPEN Study has been to provide direct confirmation of this fortuitous side benefit of adjusting for total energy; in this study, this adjustment both eliminated underestimation and greatly improved the correlation coefficient between dietary intake and the biomarkers.

**Correlated errors**

Another finding from the OPEN Study (1, 2) with potential implications for nutritional epidemiology is that errors in measuring nutrient intakes from 24-hour recalls can be correlated with errors in the same nutrient measured by using a food frequency questionnaire. From the beginning of our work in the 1970s, we have been concerned that this type of error might exist because both methods depend on many of the same cognitive processes, including memory and perception of serving sizes (7, 8). The result of this between-method correlation of error would be to overestimate the validity of food frequency questionnaires. For this reason, we have used weighed dietary records as the primary comparison method in questionnaire validation studies because this method is open ended, does not depend on memory or perception of portion sizes, and enables details of food to be recorded directly.

It is of interest to compare the results of the OPEN Study (1, 2) with those of a recent validation study by some of the same authors (9), in which they compared their questionnaire with four 24-hour recalls. The findings from the OPEN Study support the suspicion raised at that time (10) that the correlations for absolute intake measured by their questionnaire were likely to have been exaggerated by correlated errors between methods.

In the OPEN Study (1, 2), correlated errors were also suggested for protein density, and the authors claim that validity of their questionnaire for protein density assessed by comparison with 24-hour recalls was overestimated by up to 60 percent. However, this degree of overstatement applied only to women, and confidence intervals should have been provided, as they would have readily included zero. In addition, the OPEN Study data for women are almost certainly a fluke, possibly because of the use of only two 24-hour recalls or their model. The correlation coefficient for validity of protein density estimated by diet recalls was 0.79, substantially higher than reported in previous validation studies (typically 0.4–0.5), the recent report by Subar et al. (9) \((r = 0.60)\), or for men in the OPEN Study \((r = 0.50)\).

The authors of the OPEN Study (1, 2) primarily used the attenuation factor (lambda) to assess validity, but doing so can give a misleading, gloomy perspective of error from the standpoint of nutritional epidemiology because lambda reflects both the correlation between methods and differences in scaling. The correlation coefficient is more informative because it does not reflect differences in scale and is directly related to misclassification among categories such as quintiles. Unfortunately, failure to obtain a realistic estimate of within-person variability in the OPEN Study means that the correlation (although not the attenuation factor) between biomarkers and the food frequency questionnaire was underestimated. Despite this limitation, for protein density, the OPEN Study values for \( r \) and lambda using the biomarkers were not remarkably lower than those we obtained for energy-adjusted protein intake from our validation studies that used diet records as the standard (for \( r \) and lambda, the OPEN values were 0.43 and 0.40 for men and 0.35 and 0.32 for women; for men in the Health Professionals Follow-up Study, they were 0.44 and 0.21, respectively) (11), and, for women in the Nurses’ Health Study, the values were 0.50 and 0.38, respectively (unpublished data).

Thus, the findings for energy-adjusted protein intake in the OPEN Study do not support earlier speculation by Kipnis et al. (12) or their implication in the present papers that the validity of food frequency questionnaires was seriously overestimated in previous validation studies. Notably, in those previous studies, correlations for energy-adjusted protein intake tended to be lower than for other nutrients.
probably related to the low between-person variation in this nutrient.

Kipnis et al. make the sweeping conclusion from their observations that “the interpretation of findings from FFQ [food frequency questionnaire]-based epidemiologic studies of diet-disease associations needs to be reevaluated” (2, p. 14). Their findings actually suggest the contrary because the OPEN Study supports previous work on the effects of energy adjustment and the validity of energy-adjusted nutrients. We should be grateful to Subar et al. (1) and Kipnis et al. (2) for the effort they have given to this substantial project. Because this methodology is extremely expensive (and sufficient doubly labeled water to replicate this study is not even available at present) and provides such limited information in relation to the range of nutrients of interest in epidemiologic studies, the cost of replicating this study would be difficult to justify. However, it is still not too late to obtain a replicate measure of their biomarkers in a sample of their study population, because variability over an interval of several years would be germane to etiologic studies of diet and chronic disease. I hope they will do so.

REFERENCES