

Not the Time to Abandon the Food Frequency Questionnaire: Counterpoint

Alan R. Kristal and John D. Potter

Cancer Prevention Program, Fred Hutchinson Cancer Research Center, Seattle, Washington

Virtually all epidemiologists (including us) who have investigated relationships between diet and chronic disease in the past 20 years have used food frequency questionnaires (FFQ) to assess diet and will continue to publish findings based on FFQ-derived data. However, in our editorial "Is it Time to Abandon the Food Frequency Questionnaire," we addressed how we should conduct epidemiologic studies in the future. Our response to Drs. Willett and Hu is organized around the question of validity. In *Psychometric Theory* (1), Jum Nunnally and Ira Bernstein state that "Validity usually is a matter of degree rather than an all-or-none property, and validation is an unending process." Let us look afresh at three explicit standards by which the validity of the FFQ can be judged: face validity, construct validity, and predictive validity.

Face validity can be described simply as judgments by experts about the ability of a measuring instrument to function as intended. Applied to an FFQ, we can frame this question as, "Could the perceived frequency of consuming 100-125 foods or food groups over the past year, with some additional questions about food preparation and purchasing, accurately measure usual nutrient intake from the foods that were actually consumed?" Lacking data from a survey of experts, we must consider this question from our own perspective and that of our immediate colleagues. First, let us consider the face validity of an FFQ at the individual-item level. It seems unlikely that many nutritionists would use the estimated frequency of consuming "Beef, pork, or lamb as a sandwich or mixed dish, e.g., stew, casserole, lasagna, etc." to estimate an individual's nutrient intake from these foods. Indeed, how could one sensibly assign a single nutrient composition to this item given the enormous number of nutritionally dissimilar foods commonly consumed in the United States, Europe, or Asia which could be included? Consider such basic, broad-scale dissimilarities as the meats themselves; the relative amounts of meat, vegetables, cheese, and other ingredients used in mixed dishes; and the use of fats during preparation. Many FFQ items, including even relatively simple items such as "Pizza" and "Tortillas," raise similar issues. Second, if we consider face validity at the level of the entire FFQ, it seems unlikely that many epidemiologists or nutritionists would, after completing an FFQ, judge that their answers accurately reflected their usual dietary patterns or could be used to measure their own energy and nutrient intake. Many epidemiologists would probably agree that, in specific contexts and for specific purposes, FFQs can do a reasonable job in characterizing individuals on very broad dietary patterns. But there is little reason to believe that the FFQ could accurately capture the diversity and variability of foods actually consumed or measure an individual's usual nutrient intake from these foods. From the perspective of face validity, it seems far more

likely that dietary assessment approaches (e.g., food records) based on foods that are actually consumed will more accurately measure true nutrient intake.

Construct validity can be described as the degree to which the data collected reflect or measure the variable of interest. Applied to an FFQ, we can frame this question as the magnitude of the variance shared between nutrient intake estimated from the FFQ and true nutrient intake. We do have data to evaluate this question, although obtaining them requires use of recovery biomarkers to measure true nutrient intake, which are limited currently to doubly-labeled water for total energy and 24-hour urinary nitrogen excretion for total protein. The OPEN study showed clearly that the correlations of FFQ-derived data with these recovery biomarkers are poor (2). As noted by Willett and Hu, the magnitude of these correlations is limited by the reliability of the recovery biomarkers. If we assume a conservative 0.70 reliability for these biomarkers and inflate the values reported in the OPEN study accordingly, the maximum possible correlation of an FFQ with true energy intake would be 0.14 for women and 0.28 for men, and for protein, corresponding values would be 0.43 for men and 0.46 for women. The construct validity of FFQ-based assessments of other nutrients can be less rigorously evaluated using concentration biomarkers (e.g., serum selenium or phospholipid ω -3 fatty acids). These biomarkers reflect diet, but because they are also affected by other environmental and metabolic factors, it is difficult to interpret construct validity based on the range of correlations (0.2-0.6) generally observed between concentration biomarkers and FFQ-derived data. In general, we know that FFQ-derived data correlate with several biomarkers, but the amount of shared variance between measures ranges from 4% to 36%.

From our perspective, the most important standard for validity is predictive validity, which can be described as the ability of an instrument to predict functional relationships between the construct measured and an observable outcome. Applied to an FFQ, this can be framed as the ability to detect associations between a dietary exposure and disease outcome. As noted by Willett and Hu, there are several outcomes that can be predicted from FFQ-derived data; for example, high consumption of fruits and vegetables predicts reduced risk of cardiovascular disease. But many associations that we would expect to observe between diet and cancer, based on an understanding of cancer biology, animal experiments, and ecologic studies, have not been detected using FFQ-derived data. One way to evaluate predictive validity is to compare, in an observational study, results using FFQ-derived data to results using data from a dietary assessment method based on capturing foods actually consumed. A second approach, based on an intervention trial, is to compare the observed experimental outcomes with those predicted by measurement of dietary change derived from alternative dietary assessment methods. We now have sufficient data to examine both of these approaches in the context of the long-hypothesized relationship between fat intake and breast cancer risk. In 2003, Bingham published results of a study of dietary fat and breast

cancer risk, comparing an FFQ and a 7-day record as dietary assessment methods (3). In this study, based on only 168 cases, the relative risks for total fat, contrasting the highest to lowest quintiles, were 1.79 for the 7-day record ($P_{\text{trend}} = 0.05$) and 1.35 for the FFQ ($P_{\text{trend}} = 0.23$). Results from a similar study in the Women's Health Initiative were reported at the April 2006 International Conference on Dietary Assessment and made available as an advance publication on May 3, 2006 (4), which compared the baseline FFQ and 4-day diet records from women in the control arm of the Dietary Modification trial. In this study, based on 603 cancer cases, the adjusted relative risks for total fat, contrasting the lowest to highest quintiles, were 2.09 (95% confidence interval, 1.31-3.61) for the food record ($P_{\text{trend}} = 0.0008$) and 1.71 (95% confidence interval, 0.70-4.18) for the FFQ ($P_{\text{trend}} = 0.18$) or, in an alternative analysis excluding participants with missing covariate data, 2.54 for the food record ($P_{\text{trend}} = 0.006$) and 1.24 for the FFQ ($P_{\text{trend}} = 0.41$). Willett and Hu dismiss the Bingham finding, stating that "if the UK findings were correct the WHI trial should have stopped years ago." In fact, the 9.1% reduced breast cancer risk observed in the intervention arm of the WHI is remarkably close to the 8.8% reduction in risk predicted using the results given above from the observational study in the control arm (see ref. 4 for a details about this analysis). These results suggest that, at least for studies of diet and breast cancer risk, the predictive validity of a food record, based on as few as 4 days of actual intake, is superior to the predictive validity of the FFQ. It is indeed difficult to explain the consistent lack of associations in FFQ-based studies on current fat intake and breast cancer; however, the consistency of results based on diet records and a large clinical trial cannot be easily dismissed as chance findings.

Why then do FFQ-derived data predict risk of cardiovascular disease, but not cancer? We speculate that the true associations of diet with cardiovascular disease are much

stronger than observed, and that the FFQ can detect these associations despite its mediocre construct validity. Each of the cancers is probably more complex than coronary heart disease, at least at the molecular level and from the perspective of dietary influences; thus, more precise dietary assessment is needed to detect associations between diet and cancer risk. Lacking experimental studies or studies using alternative dietary assessment methods, we should not presume that findings of no association between diet and cancer risk based on FFQs alone are strong evidence of the absence of such associations. Indeed, it is possible that, when using FFQs, we are unable to detect important associations of diet with other diseases as well. In epidemiologic research, our approach has long been to focus on increasingly large study sizes to compensate for poor dietary exposure assessment. However, in some settings, even with very large participant samples, the error and bias in FFQ-based dietary assessment may obscure modest associations. Thus, we also must focus, improving the validity of dietary exposure assessment. It is clear to us that it is well past the time to abandon the FFQ for future studies of diet and cancer risk and to implement dietary assessment methods that capture data on foods actually consumed.

References

1. Nunnally JC, Bernstein IH. Psychometric theory. New York: McGraw-Hill; 1994.
2. Kipnis V, Subar AF, Midthune D, et al. Structure of dietary measurement error: results of the OPEN biomarker study. *Am J Epidemiol* 2003;158:14-21.
3. Bingham S, Luben R, Welch A, Wareham N, Khaw KT, Day N. Are imprecise methods obscuring a relation between fat and breast cancer? *Lancet* 2003;362:212-4.
4. Freedman LS, Potischman N, Kipnis V, et al. A comparison of two dietary instruments for evaluating the fat-breast cancer relationship. *Int J Epidemiol* 2006;35:1011-21.