

Is It Time to Abandon the Food Frequency Questionnaire?

Alan R. Kristal, Ulrike Peters, and John D. Potter

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

Although painful to admit, it is possible that epidemiologists have been deluded in their acceptance of food frequency questionnaires (FFQ) as the standard tool for dietary assessment in large studies of diet and cancer. The substantial limitations of FFQs have been known for some time (1) and published studies based on FFQ-derived data have long included in their discussion sections a litany of weaknesses due to suboptimal dietary assessment. However, few of us expected the astonishingly poor measurement characteristics of FFQs when compared with doubly labeled water (a gold standard for energy intake; ref. 2), nor had we expected to learn that diet and cancer associations detected when dietary assessment is based on dietary biomarkers (e.g., ref. 3) or food records (4) are undetectable when based on FFQs. We are facing a crisis: hundreds of millions of dollars and many scientists' careers have been invested in studies using only FFQs to measure diet, but it is possible that these studies have not been, and will not be, able to answer many if not most questions about diet and cancer risk.

This commentary has two broad functions. First, it describes how FFQs were developed and why they became the standard measure of diet for epidemiologic research. Second, it suggests several directions for future research and practice to improve dietary assessment in studies of diet and cancer.

Development and Acceptance of the Food Frequency Questionnaire for Studies of Diet and Cancer

In 1981, Doll and Peto (5) published a landmark report in which they estimated that 35% (with a range from 10% to 70%) of cancer was attributable to diet. The human evidence for this conclusion was largely based on international ecological studies in which cancer incidence rates were correlated with statistics on per-capita food disappearance. Given the obvious limitations of these studies, case-control or, more optimally, large cohort studies were needed to test more refined hypotheses about diet and/or nutrients and disease risk. Almost all early case-control studies used a technique called "diet history" to assess either current or past diet. Diet histories were lengthy, open-ended, and unstandardized interviews administered and analyzed by nutritionists, which attempted to characterize what a study participant "usually" ate (6). When used in case-control studies, diet histories are highly subject to bias, given the difficulty and rarity of interviewer (and, of course, participant) blinding. And because the interview could take 90 minutes, it was unsuitable for large cohort studies. Thus, out of necessity was born a simplified, self-administered, and inexpensive form of the diet history, the food frequency questionnaire (7).

Much research has been focused on both evaluating and improving FFQs, including studies of cognitive processes in food recall, inter- and intra-method reliability, and associations with "objective" measures such as body mass index and serum

micronutrient concentrations. We learned that when individuals are asked to recall diet in the very recent past, their episodic memory is reasonably accurate; however, after only a few days, episodic memory of diet erodes and recall of past diet is constructed from general knowledge about foods, most probably based on beliefs (even hopes) about one's usual or characteristic diet (8). Thus, with the exception of ceremonial foods (e.g., turkey at Thanksgiving and caviar at New Year's) or foods never eaten, recall of usual past diet (e.g., "How often did you eat a 1/2 cup serving of broccoli over the past year?") is unlikely to be quantitatively precise. We learned that FFQs are reproducible, with test-retest correlations for most nutrients in the range of 0.5 to 0.9. However, validity (intermethod reliability), based on comparisons with multiple-day diet records or 24-hour recalls, is generally not good. Correlations between FFQ- and recall-derived nutrients are often <0.4 and rarely >0.6. Associations of FFQs with both anthropometric measures and dietary biomarkers are generally weaker, although this might be expected to be due to the complex and indirect relationships between diet and dietary biomarkers. Findings of poor validity usually look better following statistical adjustment for total energy or day-to-day variability, but this does not change the fact that shared variance between an FFQ and a criterion measure ranges from 1% to 40% and that much of this shared variance may simply be correlated error (9). Although this level of agreement is objectively dismal, as long as a "validation study" is completed, the FFQ is termed "validated" and thus acceptable to study sections and manuscript reviewers.

Although limitations in the ability of FFQs to measure diet accurately were either suspected or known, most scientists believed that there were really no alternatives. Dietary measures based on actual food consumption, 24-hour dietary recalls, and food records could not be used in case-control studies because the exposure of interest was diet before disease diagnosis. Nor could measures based on actual food consumption be used in large cohort studies, in large part due to costs. For example, as part of the design of the Women's Health Initiative, alternatives to FFQs were considered. The estimated costs of dietary assessment at baseline alone (for 160,000 women) were \$1.2 million for FFQs, \$23.2 million for 3-day food records, and \$25.0 million for three 24-hour dietary recalls. Although we may wish to mount a large cohort study using multiple 24-hour recalls (considered a "gold standard" but certainly not 24-carat), the likelihood of obtaining funding for such a study is bleak.

Three relatively recent developments have focused bright lights on the limitations of FFQs. The first is the growing lack of consistency both within and across studies examining diet and cancer risk. For example, early findings for large cohorts are not confirmed with longer term follow-up (10), and many of the strong findings on diet and cancer risk from case-control studies could not be replicated in cohort studies (11) or in clinical trials (12, 13). Perhaps, due to low participation rates among eligible controls, case-control studies were detecting predictors of study participation rather than dietary risk factors for disease. The second development was the publication of the Observing Protein and Energy Nutrition study (2), which compared results from a well-designed FFQ to two

gold-standard criterion measures: urinary nitrogen excretion to measure protein intake and doubly labeled water to measure energy intake. The correlations for energy were 0.1 for women and 0.2 for men; for protein, the correlations were 0.3 for both men and women. These results imply that a study using an FFQ would observe a true relative risk of 2.0 as 1.06 for energy and 1.11 for protein. Whereas some nutrients, such as carotenoids or calcium, are probably better measured than protein or energy, these results suggest that even the largest cohort studies are unlikely to detect modest associations when using an FFQ for dietary assessment. The third development was the publication from a cohort study in which both food records and FFQs were collected (4). In this study, there was a statistically significant association of dietary fat with breast cancer risk based on the food records, but not based on the FFQ. A soon-to-be published article from a second study will confirm this important finding. The evidence is mounting that much of the inconsistency, and some of the null results, in studies of diet and cancer are due to poor dietary assessment. One incontrovertible conclusion is that we need new strategies for dietary assessment that can be practically incorporated into large cohort studies.

Four Proposals for Improving Dietary Assessment in Large Cohort Studies

Some strategies for new research or new epidemiologic practice take advantage of recently available technologies or are based on new thinking about nutritional assessment dogma. These strategies are by no means exhaustive. We hope they illustrate that there are many opportunities to improve dietary assessment and that these opportunities will grow as computerized systems for capturing information become smaller, cheaper, more powerful, and easily integrated into wireless communication networks.

Improve Food Frequency-Type Measures. Food frequencies are limited in their ability to collect complex information due to practical restrictions inherent in printed questionnaire formats. Current FFQs require categorized responses which may not capture important variability in use of energy- or nutrient-dense foods. For example, carbonated beverage portion sizes of "small," "medium," and "large" do not capture the current marketplace range of 8 to 36 oz. servings. Complex skip algorithms are also not feasible for printed questionnaires but are logical when multiple details about food purchasing and preparation are required to properly characterize a food. One example is the use and composition of multiple mixed dishes, characteristic of Asian dietary patterns, which cannot be readily captured using current FFQs. Finally, although it makes sense that pictures of foods would make it easier to report portion sizes, it is not feasible to embed multiple printed portion-size pictures into the response options for each food. Computer-administered questionnaires, delivered via internet or on touch-screen tablet computers, can address each of these problems. We do not know if using new technology to design and deliver FFQs would improve the validity of dietary assessment, but it would be worthwhile to find out.

Measure Dietary Behavior, Not Just Nutrients. In addition to focusing attention on trying to measure nutrients, we could also formulate hypotheses in terms of dietary behaviors. Questions about usual dietary practices (e.g., "When you ate bread, how often was it whole wheat or other whole grain bread?") may be more easily and accurately recalled than the frequencies and portion sizes of a long list of foods. Using this approach alone to measuring diet, study hypotheses requiring information on nutrient intake would be limited to those that could be assessed using an objective biomarker.

Collect Real-time Food-Use Information Using Computer-Aided Technologies. There are many opportunities for technological solutions to assist both in capturing and analyzing information on current, actual food use. Study participants could use a digital phone with an embedded camera to transmit pictures and descriptions of foods eaten on a meal-by-meal basis. A computer-administered 24-hour recall could be delivered over the internet or on a pocket PC. Distributing small, inexpensive computers to study participants would be far less costly and carry less participant burden than administering repeated 24-hour recalls.

Collect Multiple-Day Food Records but Analyze Them as for a Case-Cohort or Nested Case-Control Study. Current dogma is that study participants must be trained to complete food records and that records must be documented or reviewed with the participant after completion to ensure that descriptions of each food are complete. It may be, however, that an undocumented food record is good enough and that, by following a set of coding rules for missing information, even an imperfectly maintained food record could be analyzed for nutrient intake. If this were true, then participants could complete multiple-day food records, which could then be stored for later retrieval and analysis. We examined this hypothesis in a pilot study and found that correlations between documented and undocumented 3-day food records ranged between 0.87 and 1.0 (14). The high cost of food records is attributable to the documentation process, coding, and data entry for analysis. By limiting the number of records analyzed to those that are informative for a case-cohort or nested case-control study, food records would cost little more than FFQs. Research on ways to enhance the quality of undocumented records would be well motivated.

Concluding Remarks

We should be very circumspect about analyses of current studies that have used FFQs for dietary assessment. Analyses of these studies will no doubt continue as planned and efforts such as the Pooling Project of Diet and Cancer (15) may yield important findings. However, we are not likely to learn much more about diet and cancer risk by continuing to use standard food frequency questionnaires. We need, once again, to adopt a curious and exploratory attitude about dietary assessment. Large cohort studies are being initiated in Asian countries and they have a unique opportunity to develop and evaluate different approaches for dietary assessment. Cohort studies currently under way in the United States and Europe could change their methods for dietary assessment when next surveying their cohorts to update exposure information. When it comes to dietary assessment, we need more thought for food.

References

1. Byers T. Food frequency dietary assessment: how bad is good enough? *Am J Epidemiol* 2001;154:1087-8.
2. Schatzkin A, Kipnis V, Carroll RJ, et al. A comparison of a food frequency questionnaire with a 24-hour recall for use in an epidemiological cohort study: results from the biomarker-based Observing Protein and Energy Nutrition (OPEN) Study. *Int J Epidemiol* 2003;32:1054-62.
3. Fowke JH, Chung F-L, Jin F, et al. Urinary isothiocyanate levels, brassica, and human breast cancer. *Cancer Res* 2003;63:3980-6.
4. 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.
5. Doll R, Peto R. The causes of cancer: quantitative estimates of avoidable risks of cancer in the United States today. *J Natl Cancer Inst* 1981;66:1193-308.
6. Burke B. The dietary history as a tool in research. *J Am Diet Assoc* 1947;23:1041-6.
7. Stefanik P, Trulsson M. Determining the frequency of foods in large group studies. *Am J Clin Nutr* 1962;11:335-43.

8. Smith AF, Jobe JB, Mingay DL. Retrieval from memory of dietary information. *Appl Cogn Psychol* 1991;5:269–96.
9. Kipnis V, Carroll RJ, Freedman LS, Li L. Implications of a new dietary measurement error model for estimation of relative risk: application to four calibration studies. *Am J Epidemiol* 1999;150:642–51.
10. Potter JD. Vegetables, fruit, and cancer. *Lancet* 2005;366:527–10.
11. Voorrips LE, Goldbohm RA, van Poppel G, Sturmans F, Hermus RJJ, van den Brandt PA. Vegetable and fruit consumption and risks of colon and rectal cancer in a prospective cohort study. *Am J Epidemiol* 2000;152:1081–92.
12. Schatzkin A, Lanza E, Corle D, et al. Lack of effect of a low-fat, high-fiber diet on the recurrence of colorectal adenomas. *N Engl J Med* 2000;342:1149–55.
13. Alberts DS, Martinez ME, Roe DJ, et al. Lack of effect of a high-fiber cereal supplement on the recurrence of colorectal adenomas. *N Engl J Med* 2000;342:156–62.
14. Kolar AS, Patterson RE, Neuhauser ML, et al. A practical method for collecting 3-day food records in a large cohort. *Epidemiology* 2005;16:579–83.
15. Hunter CP, Spiegelman D, Adami H, et al. Non-dietary factors as risk factors from breast cancer, and as effect modifiers of the association of fat intake and risk of breast cancer. *Cancer Causes Control* 1997;8:49–56.