Cognitive research enhances accuracy of food frequency questionnaire reports: results of an experimental validation study

ABSTRACT

Objective To test whether changes to a food frequency questionnaire (FFQ) based on cognitive theory and testing result in greater accuracy. Accuracy was examined for 4 issues: 1) Grouping: asking about foods in single versus separate questions; 2) Different forms of food: asking consumption frequency of several forms of a food (eg, skim, 2%, whole milk) versus a nesting approach involving frequency of the main food (eg, milk) and proportion of times each form was consumed; 3) Additions (eg, sugar to coffee): asking independent of the main food versus nested under the main food; 4) Units: asking frequency and portion size versus frequency of units (eg, number of eggs).

Design Participants in 2 randomly assigned groups completed 30 consecutive daily food reports (DFRs), followed by 1 of 2 FFQs that asked about foods consumed in the past month. One was a new cognitively-based National Cancer Institute (NCI) Diet History Questionnaire (test-DHQ); the other was the 1992 NCI-Block Health Habits and History Questionnaire (HHHQ).

Subjects/setting 623 participants, 25-70 years, from metropolitan Washington, DC

Statistical analyses performed DFR and FFQ responses were compared using categorical (percent agreement) and continuous (rank order correlation, discrepancy scores) agreement statistics.

Results Grouping: accuracy was greater using separate questions. Different forms of food: accuracy was greater using nesting. Additions: for mayonnaise and sugar and milk added to coffee, accuracy was greater using independent questions; for milk on cereal and margarine on bread, accuracy was greater using nesting. Units: neither approach was consistently superior.

Conclusions Accuracy of FFQ reporting can be improved by restructuring questions based on cognitive theory and testing.

Cognitive research enhances accuracy of food frequency questionnaire reports: results of an experimental validation study

Food frequency questionnaires (FFQs) are used commonly to assess dietary exposures in epidemiologic studies of chronic diseases. Concern about measurement error has stimulated numerous validation and calibration studies comparing dietary intakes estimated from FFQs to those estimated by other instruments (1). Although some research has addressed the relationship of FFQ design elements to response accuracy (2), more such research is needed in order to improve the instruments, not just to calibrate them.

A typical FFQ asks the respondent to report about the frequency with which many food and beverage items are consumed over some time period; many also ask about typical portion sizes. Thus, accurate reporting on an FFQ requires a person to engage in a variety of cognitive processes, such as long-term recall, magnitude estimation of both frequency and quantity, and aggregation of frequency and quantity information (2).

In an attempt to identify ways to ease the task of completing an FFQ while enhancing the accuracy of responses, Subar et al. conducted cognitive evaluation of various approaches to collecting self-reports of usual food intake (3). First, questions on the National Cancer Institute (NCI)-Block Health Habits and History Questionnaire (HHHQ) that appeared especially challenging to respondents were identified, and alternative formats for collecting that information were devised. These HHHQ questions and the alternative formats were studied in cognitive think-aloud interviewing (4). Redesigned questions that reflected findings were then incorporated into a new FFQ—a test version of the NCI Diet History Questionnaire (test-DHQ).

Our objective was to test the effectiveness of selected major design innovations in the test-DHQ by comparing the accuracy of alternative querying approaches for 4 question-design issues:

1) <u>Grouping</u>: asking about many related foods in a single question versus in separate questions.

2) <u>Different forms of food</u>: for nutritionally different forms (eg, whole, lowfat, nonfat milks) of a main food (eg, milk), asking about the frequency and portion size of each different form versus the frequency and portion size of the main food followed by the proportion of times subordinate forms are consumed.

3) <u>Additions</u>: for foods added to other foods (eg, the addition of sugar to coffee or tea), asking about the addition independent from asking about the main food versus the addition relative to the main food.

4) <u>Units</u>: for foods conceptualized in units (eg, eggs), asking about the frequency and usual portion size versus the frequency with which units are consumed.

METHODS AND MATERIALS

Experimental Design

After stratified random assignment into 2 groups, study participants completed 30 days of criterion information, and then completed their respective FFQ's for the reference period of the previous month. Table 1 shows the items used to evaluate each question-design issue. Additional items were queried similarly on both FFQs ("control" items).

Instruments

The Daily Food Report (DFR), is a 1-page (2-sided) machine-scannable list of 88 food and drink items. Each participant filled in bubbles next to each item to indicate how many times the item was consumed that day. Items were placed on the DFR: 1) if they were relevant to the 4 question-design issues under study; and 2) if, based on other food intake data (USDA's Continuing Survey of Food Intakes of Individuals 1994-96), it was expected that they would be consumed sufficiently frequently within 30 days to allow comparisons between questioning approaches with power of 0.80, given $\alpha = 0.05$.

The machine-scannable 1992 version of the HHHQ, which asks frequency and portion size questions about 97 individual food and drink items, was used (5). Although later versions of the HHHQ are available, the 1992 version was used because the approaches developed by Subar et al. to be tested were modifications of that instrument (3). The third instrument used was a test version of the DHQ that consists of frequency and portion size questions about 195 individual food and drink items and additional questions about the specific forms in which foods are consumed (6).

Sample and Study Procedures

Reliably prompt mail was required to monitor daily completion of DFRs, so participants were sampled from a restricted geographical area. The sampling frame consisted of households with addresses listed in residential telephone directories in Washington, DC; Alexandria, Arlington, and Fairfax Counties, Virginia; and Montgomery, Prince George's, and Frederick Counties (excluding Frederick City), Maryland.

In August 1996, postcards were sent to potential participants that provided brief information about the study and indicated that the household would soon be contacted by a telephone interviewer. Upon contact, the interviewer briefly described the study purposes and incentives, and administered a short screener questionnaire to assess the individual's eligibility for and interest in the study. To be eligible, a prospective participant had to be between 25 and 70 years old; be able to speak and read English; reside in the United States; be non-institutionalized; have telephone access; have Monday through Saturday home or business mail pick-up; and be available during the data collection period. Participation was limited to 1 person per household.

Recruitment telephone calls were attempted for 4,632 telephone lines to achieve the target of 1,400 successfully screened households. Successful telephone contact was made with 2,871 households (62.0%). Of these, the respondent was ineligible in 560 (19.5%); declined to participate in 903 (31.5%); and completed the screener and was eligible to participate in 1,408 (49.0%) households.

A personalized letter describing the study in detail, a sample DFR, and a consent postcard were mailed by overnight delivery to all eligible respondents. They were requested to complete and return the consent postcard within 7 days. Of the 1,408 potential participants, 743 (52.7%)

returned signed postcards.

From the pool of returned postcards, pairs of individuals were matched, to the greatest extent possible, on 4 characteristics—gender, race/ethnicity, educational level, and age group. From each pair, 1 member was assigned randomly to the HHHQ group and the other to the test-DHQ group. Because not all participants returning postcards could be enrolled in the study, participant pairs were selected to maximize diversity; 650 individuals were enrolled in the study and received study materials.

Participants were asked to complete and mail a DFR each day for 30 consecutive days. Within 7 days of when a participant was due to mail the last DFR, 1 of the 2 FFQs and instructions for completing this instrument were mailed to the participant. Methods used to sustain participation over the data collection period included frequent contact through mail and telephone; reminder telephone calls for missing DFRs; toll-free access to research staff to answer questions; payments of $10 after 14 DFRs had been received and $15 after the final FFQ had been completed; and gifts (10 $100 and 2 $500 gifts) to randomly selected participants who had completed the study. Of the 650 participants enrolled in the study, 623 (95.8%) returned at least 26 DFRs and the FFQ. Of these, 64% were female; 76% were white, 14% African-American, 4% Latino, 6% other or unknown; 31% were ages 25 to 39, 43% ages 40 to 54, 26% ages 55 to 70; 1% had less than high school education, 12% no higher than high school education, 86% higher than high school education; and 13% resided in Washington, DC, 31% Virginia, 10% rural Frederick, MD, 44% other Maryland.

Analytical Procedures

On the DFRs, participants reported the number of times each item was consumed each day. On the FFQs, participants reported a rate of intake (ie, number of times per time period). Thus, the DFR information was continuous, whereas the FFQ information was categorical. To compare the 2 types of information required constructing new variables—category-converted DFR variables and continuous-converted FFQ variables.

Variables

DFR. For each food, daily reported frequencies of intake on the DFR were summed across all days reported, and these sums were standardized to a 30-day period. To construct the categorical DFR variables, the value of each standardized continuous variable was classified into the appropriate frequency category for that item on the relevant FFQ.

FFQs. The values of the FFQ categorical variables were the responses themselves (eg, "3-4 times per week"). To construct the continuous FFQ variables, the midpoints of frequency categories were used to convert categorical responses to monthly frequencies, standardized to a 30-day period.

Statistical Analyses

With the categorical variables, 3 statistics were calculated to assess the accuracy of FFQ responses—the percentage of individuals whose FFQ responses agreed exactly with their category-converted DFR information; the percentage of those who never recorded consuming an item (on the DFR) who reported never eating that item (on the FFQ); and the percentage of those who ever recorded consuming the food (on the DFR) who reported consuming the food (on the FFQ). The continuity-adjusted chi-square test or the Fishers' Exact Test was used for statistical inferences (7).

With the continuous variables, 3 statistics were calculated to assess FFQ response accuracy. For each item, the Spearman correlation between the DFR-recorded frequency and the continuous-converted FFQ-reported frequency was calculated. The Spearman correlation was chosen because it does not assume normality in the underlying distributions; dietary data are generally not normally distributed. In addition, actual discrepancies and absolute discrepancies of reported (FFQ) from recorded (DFR) frequency were computed. The actual discrepancies reflect both the amount by which frequency reports differed from the criterion measure and the direction of the difference; in averaging over individuals for any item, overestimates and underestimates offset each other. The absolute discrepancies reflect only the amount of the discrepancy; the mean indicates the average magnitude of reporting errors relative to the criterion measure.

Statistical inferences were based on 2-tailed 1- and 2-sample t-tests; alpha was set at 0.05. Cases with missing data on a particular FFQ food item were excluded from analysis of that DFR-FFQ comparison.

RESULTS

Comparability of HHHQ and Test-DHQ Samples

To examine comparability between participants completing the HHHQ and the test-DHQ, the demographic characteristics, FFQ completion rate, dietary intake recorded on the DFR instrument, and accuracy of FFQ reports for similarly worded foods were compared.

The HHHQ and test-DHQ groups were not significantly different in gender, age, race/ethnicity, educational level, or residence. Of the 623 participants in the study, 314 completed the HHHQ and 309 the test-DHQ, indicating that the greater length of the test-DHQ did not adversely affect completion in this study.

For each dietary item examined, the proportion of each FFQ group that reported any consumption of that item on the DFR and, for those individuals who did, the mean standardized frequency of consumption (for 30 days) were computed (data not shown). The proportion of consumers and mean consumption frequency for most foods were quite similar for the 2 FFQ samples. Of the 29 foods examined, the proportion of consumers differed significantly between the 2 samples for 6 foods; the mean frequency differed between the 2 samples for 3 foods.

The reporting accuracy by participants in the 2 FFQ groups was compared for the 4 items that are asked similarly on the 2 FFQs (Table 1). Statistical comparisons were made for overall percent agreement, percent agreement for consumers, percent agreement for non-consumers, mean actual discrepancy, and mean absolute discrepancy (data not shown). Of the 20 comparisons, there was only 1 statistically significant difference in reporting accuracy between the 2 groups.

The results indicate that the participants in the 2 groups were comparable demographically, in their completion rates of the FFQ, in their diets, and in their degree of accuracy in responding to FFQs.

Food Grouping: Single versus Separate Questions

Response accuracy of frequency reports for foods (Table 1) asked about in a single question versus asked about in separate questions were compared. In most but not all comparisons, the test-DHQ separated food question approach was superior to the HHHQ grouped food single question approach (Table 2). Although there were no statistically significant differences between the 2 approaches in agreement for those respondents who did not consume the item, agreement for those who did consume the item was higher for the test-DHQ's separated food question approach. For consumers of the examined foods, there were statistically significant biases in the HHHQ for all 4 foods and in the test-DHQ for 2 foods. In addition, the magnitude of the bias was consistently lower for the test-DHQ than for the HHHQ. The HHHQ's grouped food single question approach was associated with statistically significant underreporting for all 4 foods examined. For example, for doughnuts, cookies, cake, and pastry, asking a single question on the HHHQ led to underreporting of about 6 times per month, whereas asking about the same foods in 3 questions on the test-DHQ led to underreporting of about 4 times per month.

Different Forms of Food: Multiple Separate Questions versus Nesting

The HHHQ asks separate questions about frequency of consumption and portion size for different forms of certain foods. The test-DHQ uses a nesting approach in which questions about the frequency of consumption and portion size for the main food are asked followed by questions about the relative proportions of time subordinate forms are consumed (Table 1).

In general, agreement statistics for the total samples indicate that frequency estimates for the main food were better with the nesting approach than with the multiple separate question approach (Table 3). Agreement for those never consuming as well as those consuming was generally high for all 4 foods examined, and similar for the 2 FFQ approaches. For consumers, there was statistically significant bias in the HHHQ for all 5 foods examined and in the DHQ for 4 foods. The magnitude of the bias was significantly smaller for the test-DHQ nesting approach than for the HHHQ multiple separate question approach for all foods. For example, milk to drink was overreported by about 11 times per month with the HHHQ multiple separate question approach, but by only 3 times per month with the test-DHQ nesting approach.

A cold cereal adjustment question on the HHHQ presents a commonly used method of adjusting the total frequency of cold cereal consumption. The question asks the respondent to report how many servings of cold cereal were consumed during the reference period. When the total of the 3 individual cereal items differs from the response to the summary question by 20% or more, an analytical adjustment is made to the "Other cold cereals" line item. The response accuracy of this adjustment procedure was compared to that of the test-DHQ nesting approach. There was little change in the various agreement statistics for the HHHQ; the test-DHQ nesting approach continued to be superior (Table 3).

Reporting frequency of intake accurately is likely to be more difficult when more than 1 form of the food is consumed, as more judgments are required. For the 4 foods examined, mean absolute discrepancy scores were higher (ie, accuracy was lower) for individuals who consumed multiple forms of the foods than for those who consumed only 1 form (data not shown). One would expect that the test-DHQ nesting approach would be especially advantageous for individuals who consume multiple forms of the food. Mean bias and mean absolute discrepancy scores were lower (ie, accuracy was greater) with the test-DHQ than with the HHHQ in each subgroup (ie, 1- and multiple-form consumers) for every food. For 3 of the 4 foods examined, the test-DHQ advantage over the HHHQ in mean bias was greater in subgroups that had consumed multiple types of a food than in subgroups that had consumed single forms of the food. For example, among test-DHQ milk drinkers, those consuming only 1 type of milk overreported by about 2 times per month; those consuming multiple types of milk overreported by about 6 times per month. For HHHQ milk drinkers, those consuming only 1 type of milk overreported by about 6 times per month; those consuming multiple types of milk overreported by 18 times per month.

Information about intake of individual forms of foods was generally more accurate with the test-DHQ approach than with the HHHQ approach (Table 4). Of the 7 foods examined, there was a statistically significant bias for 6 foods using the HHHQ approach and for 2 foods using the test-DHQ approach. When considering absolute discrepancies, the test-DHQ responses were more accurate than the HHHQ responses for 5 foods, statistically significantly different for 2. The HHHQ approach was significantly more accurate than the test-DHQ approach for 1 food.

Additions to Foods: Single Independent Question versus Nesting

The third design area examined--whether to ask about the use of an addition to many food items in a single independent question (the HHHQ approach) or to nest questions about the use of the addition underneath each main food (the test-DHQ approach) (Table 1)---produced inconsistent results across the examined foods (Table 5). Sugar in coffee/tea was reported somewhat more accurately with the HHHQ single independent question approach than with the test-DHQ nesting approach. For milk on cereal, the test-DHQ nesting approach was somewhat better than the HHHQ single independent question approach. For salad dressing and mayonnaise, no differences between approaches were statistically significant; the test-DHQ was somewhat better in correctly classifying non-consumers and consumers, and the HHHQ was better in correctly estimating the degree of use among consumers. For margarine on bread, the HHHQ was somewhat better overall than the test-DHQ; however, the test-DHQ was superior in estimating the degree of consumption among consumers. For butter, the approaches were comparable overall.

Foods Conceptualized in Units: Frequency and Portion Size versus Frequency of Units

The last design issue examined was the utility of asking about foods in their usual serving units compared to asking about frequency and portion size (Table 1). Accuracy statistics for the 2 FFQs were generally very similar (data not shown). Accuracy was relatively high for both instruments and for all items for non-consumers and consumers. However, for consumers, systematic bias in reporting the frequency of consumption was evident. Eggs were underreported on the HHHQ by 1 a month and overreported on the test-DHQ by 2 a month; the HHHQ approach was associated with significantly less bias. However, the test-DHQ approach was associated with less absolute error. Coffee and tea consumption were overestimated by both FFQs by 6 to 8 cups a month; neither approach was significantly superior.

DISCUSSION

It is likely that many characteristics of FFQs affect reporting accuracy, including the number and nature of line items, wording, formatting, and response categories. The results of our study illustrate the complexity in the interplay of these characteristics.

The first question-design issue investigated in this study was the effect on accuracy of grouping many food items into single questions. In closed-ended dietary questionnaires like FFQs, it is impossible to ask about consumption of the entire food supply. FFQ designers must balance the desire for completeness with the response burden of asking about more line items. For example, Krebs-Smith et al. showed that for total fruit and vegetable consumption, FFQs with more line items were associated with higher estimates of intake; however, they did not have reference data to assess response accuracy (8). An often used strategy to reduce the number of line items without sacrificing completeness is to group many foods into one line item. Serdula et al. established that grouping foods into single questions affects the reported frequency of use (9,10). The percentage of non-consumption was significantly lower when reporting about separated foods than when reporting about grouped foods; reported consumption by those consuming the foods was significantly higher for the separated foods than for the grouped foods for doughnuts/cookies/cake/pie/pastries; potato chips/popcorn/salty snacks; and mayonnaise/salad dressing (9). However, no reference data were available in that study either. In our study, the accuracy of grouped versus separated foods for 4 different food groupings, including the above 3, was examined. Reported frequency of consumption was higher when foods were separated (test-DHQ) than when foods were grouped (HHHQ), consistent with Serdula et al. Accuracy was greater for the separated foods questions than for the grouped foods questions.

Although an appropriate level of grouping is somewhat specific to the particular foods and the eating patterns of the particular population, these results suggest some factors to consider when designing FFQs. The foods asked together in a single line item should be perceived and used by the respondents as substitutes. Thus, tomatoes and tomato juice, variants of the same food, are likely not used as substitutes. Another consideration is the breadth of the candidate foods and their uses. For example, doughnuts, cookies, cake, and pastry include many individual items; in addition, these foods are used in many different situations, for example, at multiple meals and snacks, and at non-routine celebratory occasions. These factors increase the difficulty of the respondent's task of combining frequency of use over all individual foods and uses implied by the question.

A second question-design issue examined was whether nutritionally different forms of a food (eg, milk with different fat contents) would be reported more accurately when asked as separate items (as on the HHHQ) or using a nesting approach that asks first about the main food and then asks the respondent to apportion this overall consumption across the forms (as on the test-DHQ). The test-DHQ's nesting approach is based on the idea that respondents store and process information in an hierarchical manner, and can more easily access basic rather than subordinate information (11). According to this hypothesis, a judgment is made most easily about the main food (eg, milk); this judgment can then be refined into judgments about subordinate forms of that food (eg, lowfat milk). This hierarchical structuring is consistent with the evolution of the food supply in the U.S.; over time, more forms of main foods have become available. The nesting approach appears to have been effective in enhancing the accuracy of reported frequency of consumption of such main foods as bread, milk to drink, cold cereals, and soups. For participants who provided information about the particular subordinate forms of foods that were consumed, accuracy with the nesting approach was generally equal or superior to that with the HHHQ's separated foods approach.

Nesting will likely be most advantageous for collecting information about main foods when all the following conditions apply: a) the main food is perceived by the particular population group as a clear and distinct entity; b) the food is available in more than 1 form; and c) the forms differ in nutritionally important ways. In addition, our results suggest that the nesting approach may be particularly useful when individuals consume more than 1 of the subordinate forms of a food. The nesting approach may become all the more useful as the availability of modified forms of foods in our marketplace increases.

The results concerning the first 2 design issues may seem inconsistent: The first result shows that separating foods into separate questions elicits more accurate responses than combining foods into a single question, whereas the second result shows that separate questions elicit less accurate responses than a global question about the different forms of a food.

The crucial distinction concerns the levels of a conceptual hierarchy at which aggregation is made. Cognitive psychologists (eg, Rosch et al. (11)) have distinguished between basic level concepts (eg, milk), subordinate concepts (eg, whole milk), and superordinate concepts (eg, dairy beverages). Our results indicate that combining multiple basic-level concepts into a single question yields less accurate responding than does asking them as separate questions. But asking single questions about basic-level items, followed by nested questions about the subordinate forms, yields more accurate responding than does asking separate questions about the subordinate forms.

The third question-design issue was whether in asking about additions to foods, accuracy is greater when asked as independent line items (as on the HHHQ) or using a nesting approach that asks about the main food followed by the proportion of time the addition is used on that main food (as on the test-DHQ). The accuracy of the nesting approach depends to a large extent on the accuracy of frequency information given about the main food. Frequency reports for cold cereal and total bread were superior with the test-DHQ nesting approach, and accuracy for additions to these foods was also superior for the test-DHQ. Another factor may be the number of items individuals have to consider simultaneously. For additions to coffee and tea, for which the frequency reports between test-DHQ and HHHQ were similarly accurate, the HHHQ single independent question approach was somewhat better than the test-DHQ nesting approach. For "salad dressing, including on sandwiches or on potato salad, etc.," however, the two approaches were comparable. Asking about additions to a longer list of main foods in a single question, untested in this study, might lead to lowered accuracy.

The last question-design issue examined was the utility of asking about foods usually consumed in standard units (eg eggs) in terms of the frequency of units consumed rather than asking the frequency and portion size. Although Subar et al. noted that judgments about such foods appeared easier when they were asked in their usual serving units (3), the empirical results showed no comparative advantage for either approach.

The question-design issues examined in this validation study are not mutually exclusive: for example, nested additions on the test-DHQ are less grouped than comparable HHHQ items. The results are highly specific to the particular comparison being made; no finding was universal across all the foods in any given issue. In addition, there are various ways to operationalize these issues in a particular FFQ. Furthermore, our study examined accuracy over a 1-month period; many FFQs ask about food use over a longer time period, such as a year.

A major strength of this study is its design. The reference instrument was a pre-coded recording form composed of a targeted list of foods to be completed daily for 30 days; an FFQ then asked participants about that time period. With this design, frequency estimates made on the FFQ were compared to previously reported behavior for the entire time period asked. Although a similar "checklist" approach has been used previously (12-13), our study is unique in that the sample was large, non-institutionalized, and drawn from a diverse population, encompassing urban, suburban, and rural areas, and whites, African-Americans, and Latinos. The high completion rate for participants in this study demonstrates that the checklist method is useful for validation studies.

APPLICATIONS/CONCLUSIONS

- Cognitive research can reveal superior ways to ask about particular foods on an FFQ and other dietary assessment instruments.
- Dietitians should be aware of these cognitive issues when selecting an FFQ for a particular target audience.
- The findings of this research were used to update and complete a version of the DHQ for public use: www-dccps.ims.nci.nih.gov/ARP/DHQ

References

1. Dietary assessment calibration/validation register. Available at: http://www-dacv.ims.nci.nih.gov. 2001.

2. Smith AF. Cognitive processes in long-term dietary recall. Vital and Health Statistics 1991;6:1-34.

3. Subar AF, Thompson FE, Smith AF, Jobe JB, Ziegler RG, Potischman N, Schatzkin A, Hartman A, Swanson C, Kruse L, Hayes R, Lewis DR, Harlan LC. Improving food frequency questionnaires: A qualitative approach using cognitive interviewing. J Am Diet Assoc 1995;95:781-8.

4. Willis GB, Royston P, Bercini D. The use of verbal report methods in the development and the testing of survey questionnaires. Appl Cognit Psychol 1991;5:251-67.

5. Block G, Hartman AM, Dresser CM, Carroll MD, Gannon J, Gardner L. A data-based approach to diet questionnaire design and testing. Am J Epidemiol 1986;124:453-69.

6. Subar AF, Midthune D, Kulldorff M, Brown CC, Thompson FE, Kipnis V, Schatzkin A. An evaluation of alternative approaches to assign nutrient values to food groups in food frequency questionnaires. Am J Epidemiol 2000;152:279-86.

7. Fleiss, J. L. (1981). Statistical Methods for Rates and Proportions, 2nd ed. New York: Wiley.

8. Krebs-Smith SM, Heimendinger J, Subar AF, Patterson BH, Pivonka E. Using food frequency questionnaires to estimate fruit and vegetable intake: Association between the number of questions and total intakes. J Nutr Ed 1995;27:80-85.

9. Serdula M, Byers T, Coates R, Mokdad A, Simoes EJ, Eldridge L. Assessing consumption of high-fat foods: the effect of grouping foods into single questions. Epid. 1992;3:503-8.

10. Serdula MK, Mokdad AH, Byers T, Siegel PZ. Assessing alcohol consumption: beverage-specific versus grouped-beverage questions. J Stud Alcohol 1999;60:99-102.

11. Rosch E, Mervis CB, Gray W, Johnson D, Boyes-Braem P. Basic objects in natural categories. Cognit Psychol 1976;8:382-439.

12. Johnson NE, Sempos CT, Elmer PJ, Allington JK, Matthews ME. Development of a dietary intake monitoring system for nursing homes. J Am Diet Assoc 1982;80:549-57.

13. Hammond J, Nelson M, Chinn S, Rona RJ. Validation of a food frequency questionnaire for assessing dietary intake in a study of coronary heart disease risk factors in children. Eur J Clin Nutr 1993;47:242-250.

Table 1
Differences in how questions about the test food(s) are asked on the HHHQ and the test-DHQ by question design issue

How food is asked on HHHQ	How food is asked on test-DHQ
Foods that are identical or similar	
Chocolate candy	Chocolate candy
Bananas	Bananas
Orange juice or grapefruit juice	Orange juice or grapefruit juice
Green salad	Lettuce salads (with or without other vegetables
Foods grouped as a single item on the HHHQ versus as separate items on the test-DHQ	
Doughnuts, cookies, cake, pastry	Doughnuts, sweet rolls, or Danish Cookies or brownies (including low-fat) Cake (including low-fat)
Salty snacks, such as potato chips, corn chips, popcorn	Potato chips, tortilla chips, or corn chips (including low-fat or low-salt) Pretzels
Tomatoes, tomato juice	Tomato juice or vegetable juice Tomatoes, fresh (including in salads)
Hamburgers, cheeseburgers, meatloaf, beef burritos, tacos	Beef hamburgers or cheeseburgers Ground beef in mixtures, such as tacos, burritos, meatballs, casseroles, chili, meatloaf

Asking about differing forms of food as multiple questions on the HHHQ versus a nesting approach on the test-DHQ

Frequency of consumption	**Frequency of consumption**[1]
Whole milk and beverages with whole milk (not including on cereal) 2% milk and beverages with 2% milk (not including on cereal) Skim milk, 1% milk or buttermilk (not including on cereal)	Milk as a drink (NOT in coffee, NOT on cereal) Whole milk (4% fat)[1] 1% or 2% fat milk[1] Skim, nonfat, or 1/2% milk[1]
High fiber, bran or granola cereals, shredded wheat Highly fortified cereals, such as Total, Just Right or Product 19 Other cold cereals such as corn flakes, Rice Krispies	Cold cereal Total or Product 19[1] High-fiber cereals such as Fiber One, All Bran, or 100% Bran[1] Other fiber cereals, such as Cheerios, Shredded Wheat, Raisin Bran, Bran Flakes, Granola[1] Any other cold cereal[1]
White bread (including sandwiches, bagels, burger rolls, French or Italian bread Dark bread, such as wheat, rye, pumpernickel, (including sandwiches)	Bagels, English muffins Breads or rolls FOR SANDWICHES (including burger or hot dog rolls) White bread[1] Dark bread[1] Breads or dinner rolls, NOT INCLUDING ON SANDWICHES White bread[1] Dark bread[1]
Vegetable and tomato soups, including vegetable beef, minestrone Other soups	Soups Bean-based soups[1] Cream soups, including chowders[1] Tomato or vegetable soups[1] Broth soups with or without noodles or rice[1]

Asking about additions to foods: single question on HHHQ versus nesting approach on test-DHQ

Frequency of consumption	Frequency of consumption[1]
Sugar in coffee or tea	Coffee Sugar or honey[1] Iced tea Sugar or honey[1] Hot tea Sugar or honey[1]
Milk in coffee or tea	Coffee Whole milk (4% fat)[1] 1% or 2% fat milk[1] Skim, nonfat, or 1/2% milk[1]
Cream (real) or half-and-half in coffee or tea	Coffee Cream or Half-and-Half[1]
Non-dairy creamer in coffee or tea	Coffee Low-fat, non-dairy creamer[1] Regular non-dairy creamer[1]
Milk on cereal	Cold cereal Whole milk (4% fat)[1] 1% or 2% fat milk[1] Skim, nonfat, or 1/2% milk[1]
Regular salad dressing & mayonnaise, including on sandwiches or on potato salad, etc.	Salad dressing for lettuce salads or vegetables (including low-fat) Breads...FOR SANDWICHES... Mayonnaise or mayonnaise-type dressing (including low-fat)[1] Tuna (canned) including in salads, sandwiches, or casseroles Tuna prepared with mayonnaise or other dressing (including low-fat)[1]

Margarine on bread or rolls	Breads...FOR SANDWICHES... Margarine (including low-fat)[1] Breads...NOT...ON SANDWICHES Margarine (including low-fat)[1]
Butter on bread or rolls	Breads...FOR SANDWICHES... Butter[1] Breads...NOT...ON SANDWICHES Butter[1]
Asking frequency and portion size on the HHHQ versus frequency of units on the test-DHQ for unit-specific foods[2]	
Eggs[3]	How many eggs, egg whites, or egg substitutes (NOT counting eggs in baked goods and desserts)...
Coffee, regular or decaf[3]	How many cups of coffee, caffeinated or decaffeinated,...
Tea (hot or iced)[3]	How many glasses of ICED tea, caffeinated or decaffeinated,... How many cups of HOT tea, caffeinated or decaffeinated,...

[1]Proportion of time food was consumed was asked for foods noted. Asked as: "How often is the [main food] eaten in this form?" Response categories: Almost never or never; About 1/4 of the time; About ½ of the time; About 3/4 of the time; Almost always or always

[2]See (3) for visual depiction of the format of the test-DHQ question.

[3]Portion size options are S, M, L.
Eggs: S = 1 egg
M = 2 eggs
Coffee: M = 1 medium cup
Tea: M = 1 medium cup

Table 2

Agreement between Daily Food Report and FFQ (HHHQ and test-DHQ) for single versus separate questions

Food	Total sample: Percent exact agree[a,b]		Total sample: Spearman's correlation[c]		Non-consumers: Percent agree[b,d]		Consumers: Percent agree[b,e]		Consumers: Mean bias[f]		Consumers: Mean absolute discrepancy[g]	
	HHHQ	DHQ	HHHQ	DHQ	HHHQ	DHQ	HHHQ	DHQ	HHHQ	DHQ	HHHQ	DHQ
Doughnuts, cookies, cake, pastry	16.7 *	30.5	0.67	0.78	71.4	100.0	89.8 *	97.7	<u>-6.3</u> *	<u>-4.3</u>	9.4 *	7.0
(n)	(312)	(305)	(312)	(305)	(7)	(3)	(305)	(302)	(305)	(302)	(305)	(302)
Salty snacks	27.1 *	36.1	0.65	0.77	58.8	48.0	90.9 *	97.5	<u>-1.8</u>	0.7	6.4	5.3
(n)	(314)	(302)	(314)	(302)	(17)	(25)	(297)	(277)	(297)	(277)	(297)	(277)
Tomatoes, tomato juice	30.8	37.0	0.68	0.76	69.7	66.7	91.8	95.7	<u>-1.4</u>	<u>1.1</u>	5.6	5.3
(n)	(312)	(300)	(312)	(300)	(33)	(24)	(279)	(276)	(279)	(276)	(279)	(276)
Ground beef plain, mixed	36.2	36.9	0.69	0.69	78.0	67.6	88.9 *	97.0	<u>-2.0</u> *	0.9	3.5	3.9
(n)	(312)	(301)	(312)	(301)	(50)	(37)	(262)	(264)	(262)	(264)	(262)	(264)

[a]Total sample percent exact agree is the proportion of respondents whose category-converted Daily Food Report frequency is identical to their reported FFQ frequency category.

[b]Asterisks indicate that proportions are statistically significantly ($p < 0.05$) different.

[c]Spearman's rank order correlation is the agreement between the respondent's Daily Food Report frequency and his/her categorical FFQ response after conversion into a continuous frequency.

[d]Non-consumer percent agree is the percent of those reporting no consumption of the food on the Daily Food Report who also reported not consuming it on the FFQ.

[e]Consumer percent agree is the percent of those who reported consuming the food on the Daily Food Report who also reported consuming it on the FFQ.

[f]Mean bias is FFQ - Daily Food Report. Underline indicates that the mean bias of that FFQ is statistically significantly ($p < 0.05$) different

from 0. Asterisks indicate that the magnitude of the bias in each FFQ is statistically signficantly ($p < 0.05$) different from each other.

[g]Mean absolute discrepancy is |FFQ - Daily Food Report|. Asterisks indicate that the magnitude of the bias in each FFQ is statistically significantly ($p < 0.05$) different from each other.

Table 3

Agreement between Daily Food Report and FFQ (HHHQ and test-DHQ) for different forms of foods: multiple separate questions versus nesting

Food	Total sample				Non-consumers		Consumers		Consumers			
	Percent exact agree[a,b]		Spearman's correlation[c]		Percent agree[b,d]		Percent agree[b,e]		Mean bias[f]		Mean absolute discrepancy[g]	
	HHHQ	DHQ	HHHQ	DHQ	HHHQ	DHQ	HHHQ	DHQ	HHHQ	DHQ	HHHQ	DHQ
Milk to drink	45.3 *	54.6	0.74	0.87	77.5	87.5	88.5	84.5	<u>10.6</u> *	<u>3.4</u>	12.8 *	8.0
(n)	(289)	(304)	(289)	(304)	(89)	(104)	(200)	(200)	(200)	(200)	(200)	(200)
Bread	27.9 *	41.6	0.51	0.64	-	100.0	99.7	100.0	<u>-8.8</u> *	<u>-3.3</u>	14.6	12.8
(n)	(308)	(305)	(308)	(305)	(0)	(1)	(308)	(304)	(308)	(304)	(308)	(304)
Soups	37.3	45.3	0.76	0.77	63.0	81.4	93.6	97.4	<u>1.7</u> *	-0.3	3.9 *	2.9
(n)	(306)	(309)	(306)	(309)	(27)	(43)	(279)	(266)	(279)	(266)	(279)	(266)
Cold cereals	41.2	45.6	0.85	0.91	85.1	82.8	95.7	97.5	<u>3.8</u> *	<u>1.5</u>	7.5 *	5.5
(n)	(306)	(305)	(306)	(305)	(74)	(64)	(232)	(241)	(232)	(241)	(232)	(241)
Cold cereals, adj. on HHHQ (n)	43.8 (306)	45.6 (305)	0.88 (306)	0.91 (305)	85.1 (74)	82.8 (64)	95.7 (232)	97.5 (241)	<u>4.0</u> * (232)	<u>1.5</u> (241)	7.1 * (232)	5.5 (241)

[a]Total sample percent exact agree is the proportion of respondents whose category-converted Daily Food Report frequency is identical to their reported FFQ frequency category.

[b]Asterisks indicate that proportions are statistically significantly ($p < 0.05$) different.

[c]Spearman's rank order correlation is the agreement between the respondent's Daily Food Report frequency and his/her categorical FFQ response after conversion into a continuous frequency.

[d]Non-consumer percent agree is the percent of those reporting no consumption of the food on the Daily Food Report who also reported not

consuming it on the FFQ. A dash indicates that there are no individuals reporting no consumption of this food on the Daily Food Report.

[e]Consumer percent agree is the percent of those who reported consuming the food on the Daily Food Report who also reported consuming it on the FFQ.

[f]Mean bias is FFQ - Daily Food Report. Underline indicates that the mean bias of that FFQ is statistically significantly ($p < 0.05$) different from 0. Asterisks indicate that the magnitude of the bias in each FFQ is statistically significantly ($p < 0.05$) different from each other.

[g]Mean absolute discrepancy is |FFQ - Daily Food Report|. Asterisks indicate that the magnitude of the bias in each FFQ is statistically significantly ($p < 0.05$) different from each other.

Table 4

Agreement between Daily Food Report and FFQ (HHHQ and test-DHQ) for differing forms of milk, cereal, bread and soup: multiple separate versus nested querying

Food	Total sample: Percent exact agree[a,b]		Total sample: Spearman's correlation[c]		Non-consumers: Percent agree[b,d]		Consumers: Percent agree[b,e]		Consumers: Mean bias[f]		Consumers: Mean absolute discrepancy[g]	
	HHHQ	DHQ	HHHQ	DHQ	HHHQ	DHQ	HHHQ	DHQ	HHHQ	DHQ	HHHQ	DHQ
Whole milk to drink	84.1	81.0	0.65	0.59	96.0	99.5	59.6 *	35.4	1.0	-1.3	3.9	4.5
(n)	(301)	(231)	(301)	(231)	(249)	(183)	(52)	(48)	(52)	(48)	(52)	(48)
Highly fortified cereal	77.6 *	91.8	0.38	0.35	82.1 *	95.0	69.6	55.6	<u>6.0</u>	2.7	7.3	4.2
(n)	(308)	(267)	(308)	(267)	(285)	(258)	(23)	(9)	(23)	(9)	(23)	(9)
High fiber cereal	56.1 *	92.2	0.34	0.21	59.3 *	98.4	88.0 *	16.7	<u>9.4</u>	-4.3	10.2	6.8
(n)	(310)	(268)	(310)	(268)	(285)	(250)	(25)	(18)	(25)	(18)	(25)	(18)
White bread	15.7	18.8	0.54	0.65	0.0	100.0	92.3 *	83.0	<u>-11.3</u> *	<u>-7.0</u>	14.0 *	10.1
(n)	(312)	(277)	(312)	(277)	(1)	(1)	(311)	(276)	(311)	(276)	(311)	(276)
Dark bread	35.5	33.0	0.81	0.67	65.2	74.5	88.3	82.1	<u>2.8</u>	<u>3.1</u>	6.6 *	8.2
(n)	(310)	(276)	(310)	(276)	(46)	(47)	(264)	(229)	(264)	(229)	(264)	(229)
Vegetable/ tomato soups (n)	36.9	43.1	0.58	0.70	54.5	64.5	84.6	85.6	<u>1.4</u> *	-0.1	3.2 *	2.2
	(309)	(295)	(309)	(295)	(101)	(107)	(208)	(188)	(208)	(188)	(208)	(188)
Other soups	34.4 *	42.8	0.59	0.55	48.7	48.8	86.3	94.0	<u>1.2</u>	0.6	2.9	2.3
(n)	(311)	(269)	(311)	(269)	(78)	(86)	(233)	(183)	(233)	(183)	(233)	(183)

[a]Total sample percent exact agree is the proportion of respondents whose category-converted Daily Food Report frequency is identical to their

reported FFQ frequency category.

[b]Asterisks indicate that proportions are statistically significantly ($p < 0.05$) different.

[c]Spearman's rank order correlation is the agreement between the respondent's Daily Food Report frequency and his/her categorical FFQ response after conversion into a continuous frequency.

[d]Non-consumer percent agree is the percent of those reporting no consumption of the food on the Daily Food Report who also reported not consuming it on the FFQ.

[e]Consumer percent agree is the percent of those who reported consuming the food on the Daily Food Report who also reported consuming it on the FFQ.

[f]Mean bias is FFQ - Daily Food Report. Underline indicates that the mean bias of that FFQ is statistically significantly ($p < 0.05$) different from 0. Asterisks indicate that the magnitude of the bias in each FFQ is statistically significantly ($p < 0.05$) different from each other.

[g]Mean absolute discrepancy is |FFQ - Daily Food Report|. Asterisks indicate that the magnitude of the bias in each FFQ is statistically significantly ($p < 0.05$) different from each other.

Table 5
Agreement between Daily Food Report and FFQ (HHHQ and test-DHQ) for additions to foods: independent versus nested querying

Food	Total sample				Non-consumers		Consumers		Consumers			
	Percent agree[a,b]		Spearman's correlation[c]		Percent agree[b,d]		Percent agree[b,e]		Mean bias[f]		Mean absolute discrepancy[g]	
	HHHQ	DHQ	HHHQ	DHQ	HHHQ	DHQ	HHHQ	DHQ	HHHQ	DHQ	HHHQ	DHQ
Milk on cereal	43.7 *	52.9	0.86	0.89	85.1	94.7	91.7	87.8	<u>2.9</u> *	0.8	5.5	4.5
(n)	(309)	(204)	(309)	(204)	(67)	(57)	(242)	(147)	(242)	(147)	(242)	(147)
Salad dressing	28.9	31.5	0.60	0.69	63.6	64.5	82.8	89.1	<u>-1.7</u>	<u>3.0</u>	5.9	6.1
(n)	(312)	(260)	(312)	(260)	(22)	(31)	(290)	(229)	(290)	(229)	(290)	(229)
Coffee/tea additions:												
Sugar	60.5	52.6	0.85	0.86	95.4 *	86.9	78.2	80.7	2.9	<u>8.8</u>	16.3	15.4
(n)	(309)	(234)	(309)	(234)	(130)	(84)	(179)	(150)	(179)	(150)	(179)	(150)
Milk	60.9	65.0	0.82	0.82	93.5	94.9	71.2	69.0	1.6	<u>6.0</u>	12.9	12.7
(n)	(307)	(234)	(307)	(234)	(123)	(118)	(184)	(116)	(184)	(116)	(184)	(116)
Cream	71.1	71.3	0.71	0.64	95.4	97.5	56.5	48.7	1.8	-0.5	8.2	11.7
(n)	(304)	(240)	(304)	(240)	(196)	(162)	(108)	(78)	(108)	(78)	(108)	(78)
Creamer	72.0	75.9	0.84	0.86	97.5	98.0	69.4	75.6	<u>10.3</u>	<u>10.3</u>	13.6	13.4
(n)	(307)	(236)	(307)	(236)	(196)	(150)	(111)	(86)	(111)	(86)	(111)	(86)
Bread additions:												
Butter	34.0	38.5	0.59	0.61	78.0 *	90.5	60.8 *	47.8	1.1	-1.0	4.8	5.0
(n)	(309)	(273)	(309)	(273)	(106)	(95)	(209)	(178)	(209)	(178)	(209)	(178)
Margarine	42.4	40.1	0.80	0.70	86.4	84.7	81.7	74.6	<u>2.4</u> *	-0.2	6.5	6.3
(n)	(311)	(279)	(311)	(279)	(103)	(98)	(208)	(181)	(208)	(181)	(208)	(181)

[a]Total sample percent exact agree is the proportion of respondents whose category-converted Daily Food Report frequency is identical to their reported FFQ frequency category.
[b]Asterisks indicate that proportions are statistically significantly ($p < 0.05$) different.
[c]Spearman's rank order correlation is the agreement between the respondent's Daily Food Report frequency and his/her categorical FFQ

response after conversion into a continuous frequency.

[d]Non-consumer percent agree is the percent of those reporting no consumption of the food on the Daily Food Report who also reported not consuming it on the FFQ.

[e]Consumer percent agree is the percent of those who reported consuming the food on the Daily Food Report who also reported consuming it on the FFQ.

[f]Mean bias is FFQ - Daily Food Report. Underline indicates that the mean bias of that FFQ is statistically significantly ($p < 0.05$) different from 0. Asterisks indicate that the magnitude of the bias in each FFQ is statistically significantly ($p < 0.05$) different from each other.

[g]Mean absolute discrepancy is |FFQ - Daily Food Report|. There were no statistically significant differences between the FFQ's in the magnitude of the bias ($p <0.05$).

Acknowledgments

The authors thank Anna G. McIntosh for supervising the interviewers and Anne Brown Rodgers for invaluable editorial assistance.