American Journal of Epidemiology
Vol. 154, No. 12
Printed in U.S.A.

Comparative Validation of the Block, Willett, and National Cancer Institute Food Frequency Questionnaires

The Eating at America's Table Study

Amy F. Subar,[1] Frances E. Thompson,[1] Victor Kipnis,[2] Douglas Midthune,[2] Paul Hurwitz,[3] Suzanne McNutt,[3] Anna McIntosh,[3] and Simon Rosenfeld[2]

Researchers at the National Cancer Institute developed a new cognitively based food frequency questionnaire (FFQ), the Diet History Questionnaire (DHQ). The Eating at America's Table Study sought to validate and compare the DHQ with the Block and Willett FFQs. Of 1,640 men and women recruited to participate from a nationally representative sample in 1997, 1,301 completed four telephone 24-hour recalls, one in each season. Participants were randomized to receive either a DHQ and Block FFQ or a DHQ and Willett FFQ. With a standard measurement error model, correlations for energy between estimated truth and the DHQ, Block FFQ, and Willett FFQ, respectively, were 0.48, 0.45, and 0.18 for women and 0.49, 0.45, and 0.21 for men. For 26 nutrients, correlations and attenuation coefficients were somewhat higher for the DHQ versus the Block FFQ, and both were better than the Willett FFQ in models unadjusted for energy. Energy adjustment increased correlations and attenuation coefficients for the Willett FFQ dramatically and for the DHQ and Block FFQ instruments modestly. The DHQ performed best overall. These data show that the DHQ and the Block FFQ are better at estimating absolute intakes than is the Willett FFQ but that, after energy adjustment, all three are more comparable for purposes of assessing diet-disease risk. *Am J Epidemiol* 2001;154:1089–99.

diet; epidemiologic methods; food habits; nutrition assessment; nutrition surveys; questionnaires

Editor's note: *Invited commentaries on this article appear on pages 1100 and 1103, and the authors' response appears on page 1105.*

Nutritional exposures are considered among the major modifiable risk factors for several major cancers (1, 2). The most practical and economical method for collection of comprehensive dietary data in large epidemiologic studies is the food frequency questionnaire (FFQ). Improved methods in this area are essential to providing accurate estimates of dietary intake for observational epidemiologic studies and clinical trials, thereby enhancing our understanding of the role of diet in the etiology and prevention of chronic diseases. In addition, it is estimated that, if FFQs could be improved such that correlations between the measured and true exposure of percentage of energy from fat increased from 0.60 to 0.70, sample size requirements for epidemiologic studies would be reduced by 50 percent (3), leading to substantial reductions in study costs.

FFQs ask respondents to report their usual frequency of consumption of each food from a list of foods for a specific time period. Compared with other approaches, such as 24-hour dietary recalls and food records, the FFQ generally collects less detail regarding the foods consumed, cooking methods, and portion size. Therefore, the quantification of intake is not considered as accurate. However, unlike records or recalls, FFQs are designed to capture usual dietary intake. Most are completed independently by a respondent and are relatively inexpensive. Therefore, the FFQ is usually the method of choice in large-scale epidemiologic studies.

FFQs require validation prior to or as part of dietary research. The approach taken in most studies is to examine the concordance of food frequency responses with reference instruments such as multiple 24-hour dietary recalls or diet records using measurement error models to estimate correlations between nutrient intakes measured by FFQs and truth. For most foods and nutrients, such correlations are in the range of 0.40–0.70 (4).

Because of the continuing need for improved measurement of usual dietary intake and of epidemiologic questionnaires in general (5, 6), investigators at the National Cancer Institute developed a new, cognitively based FFQ, the Diet History Questionnaire (DHQ). The DHQ, which inquires

Received for publication January 12, 2001, and accepted for publication June 22, 2001.

Abbreviations: CSFII, Continuing Survey of Food Intakes by Individuals; DHQ, Diet History Questionnaire; EATS, Eating at America's Table Study; FFQ, food frequency questionnaire; FIAS 3.0, Food Intake Analysis System, version 3.0.

[1] Applied Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, MD.
[2] Biometry Research Group, Division of Cancer Prevention, National Cancer Institute, Bethesda, MD.
[3] Westat, Rockville, MD.
Reprint requests to Dr. Amy F. Subar, 6130 Executive Boulevard, MSC 7344, EPN 4005, Rockville, MD 20852 (e-mail: amy_subar@nih.gov).

about usual intake over the past year, extends previous National Cancer Institute research that led to development of the National Cancer Institute-Block Health Habits and History Questionnaire (Block) (7). Improvements in the questionnaire incorporated changes in three major areas.

First, the questionnaire was refined based on results from intensive cognitive interviewing in over 75 persons, 50–70 years of age, varying in income, education, and ethnicity. Numerous cognitive issues in FFQs related to comprehension, order of food items, intake of seasonal foods, intake averages from multiple food items, and format were found and addressed in the DHQ. Details regarding this work have been summarized elsewhere (8). Many of the cognitive changes in the DHQ improved the validity of frequency estimates (9).

Second, the list of foods and the portion size ranges for the DHQ were developed from analyses of the most recent available nationally representative dietary data, the US Department of Agriculture's 1994–1996 Continuing Survey of Food Intakes by Individuals (CSFII) (10). Particular attention was given to the increased presence of low-fat choices in the food supply and to the use of fats in food preparation. To develop the food list, the CSFII data were also examined with respect to the food sources of the nutrients of interest to chronic disease etiology or prevention: energy, fiber, carotenoids, and vitamins E, C, and A (11).

Third, a new method was developed to convert FFQ responses into daily nutrient intake estimates (12). These conversions have been made in various ways. The Block FFQ was a major advance in that food use data from a representative sample of the US population were used to derive both nutrient and portion size values (7). The analytical approach developed for the DHQ is a refinement of the Block method and represents refinement in using national dietary data to develop a FFQ nutrient database.

Among the FFQs in use in the United States over the past decade, the Block and Willett FFQs or modifications of either are among the most widely used in epidemiologic research. The purpose of the Eating at America's Table Study (EATS) was to determine correlations between nutrient intakes estimated from each of the three FFQs (DHQ, Block, and Willett) and "truth" and to compare the correlations obtained from each of the instruments. "True" intakes were estimated using a measurement error model based on repeat 24-hour recalls collected over the course of 1 year as the reference (validation) instrument.

MATERIALS AND METHODS

Sample and study design

EATS, which began in August 1997, incorporated nationally representative sampling and random digit dialing techniques to obtain 12,615 telephone numbers. The intent was to include participants 20–70 years of age, balanced by gender, and to exclude those on liquid diets or who self-reported poor English reading skills. After excluding nonworking, nonresidential, or unanswered numbers and ineligible participants, 3,590 eligible persons were identified; 1,640 (46 percent) were willing to participate. At the time of the initial telephone screening, eligible participants were administered a brief questionnaire regarding demographics, body weight, smoking history, and physical activity. This was followed by four telephone-administered, nonconsecutive, 24-hour recalls, timed to occur throughout 1 calendar year, with one recall per season. The timing and number of recalls provided data to assess seasonal variations in intake and enough days to estimate the intraindividual variability associated with a range of nutrients with precision (13). Following the year in which the 24-hour dietary recalls were collected, participants were randomized into two groups, each completing two FFQs, 1 month apart: One group completed the DHQ and Block FFQ, and the other completed the DHQ and Willett FFQ. Within groups, the administration order was randomized. All respondents received the FFQs by mail and were instructed to complete the questionnaires independently and return them in postage-paid return envelopes. This study was approved by the institutional review boards at the National Cancer Institute and Westat. Informed consent was obtained via respondents' voluntary willingness to participate in telephone interviews and to complete questionnaires.

Twenty-four-hour dietary recalls

The four 24-hour recalls, scheduled 3 months apart, were collected from September 1997 to August 1998. Participants were required to complete a previous recall to be eligible to stay in the study. Specific goals for the weekday-weekend pattern of collected intakes were established to ensure that the intake data better reflected participants' usual dietary intake patterns. The study set goals for the weekday-weekend pattern of collected intakes to be four weekday intakes on 25 percent of the participants, three weekday/one weekend intakes on 50 percent, and two weekday/two weekend intakes on 25 percent to allow for the practical aspects of data collection as well as to ensure variability in weekend versus weekday distributions across participants. These goals were met by more than 99 percent of the sample. Eighty-four percent of the participants reported using the measuring guides (measuring cups and spoons, ruler, and two-dimensional pictures), and the remainder used their own guides or were interviewed without guides.

Interviewers for the recalls were required to have at least a Bachelor's degree in health, nutrition, or home economics and to participate in a 32-hour interviewer training. Their work was closely monitored for quality control among senior staff nutritionists by reviewing 100 percent of the collected data and by listening to 10 percent of the interviews. The 24-hour dietary recalls were collected using the multiple-pass methodology developed for the 1994–1996 CSFII (10). The Food Instruction Booklet, used in that survey and adapted for EATS, contained probes specific to each category of foods to ensure standardized interviews.

The 24-hour dietary recall data were coded using the Food Intake Analysis System, version 3.0 (FIAS 3.0), developed at the University of Texas. At the time of this study, FIAS 3.0 was up-to-date with the 1994 and 1995 CSFII. Foods not found in the database were added on the basis of released CSFII data files through 1997.

Food frequency questionnaires

The DHQ used in EATS was a 36-page booklet. The DHQ queries the frequency of intake for 124 separate food items and asks portion size for most of these by providing a choice of three ranges. For 44 of the 124 foods, from one to seven additional embedded questions are asked about related factors such as seasonal intake, food type (e.g., low fat, lean, diet, caffeine free), and/or fat uses or additions. The DHQ also includes six additional questions about the use of low-fat foods, four summary questions, and ten dietary supplement questions. A copy of the instrument used in EATS is available at the following site: http://appliedresearch.cancer.gov/DHQ/index.html.

The 1995 Block and the Willett (purple form) instruments were the versions used in EATS, as these were the most common FFQs being used at the time the study began. The Block FFQ is eight pages and queries 106 foods. It asks usual portion size as "small," "medium," or "large," providing reference medium portion sizes. In addition, it includes 13 dietary supplement questions, six questions on restaurant eating, five summary questions, eight questions on fat use or low-fat foods, and seven demographic/health-related questions. National dietary data were used to construct the food list, portion sizes, and nutrient database (7). For this study, scanned data for the Block instrument were processed at the National Cancer Institute using available software. Standard software settings were used except that "Fruit-Adjust," "Veg-Adjust," and "Recalc" were turned off.

The four-page Willett FFQ queries 126 foods. It does not include separate portion size questions but asks respondents to report their frequency of a given reference portion size. It includes ten dietary supplement questions and ten questions primarily regarding fat intake. No information is published about how the nutrient database for the Willett instrument is constructed. For this study, the Willett instrument was scanned and processed at Harvard.

Thirty-five respondents were excluded from DHQ analyses because they had skipped two facing pages of the 36-page booklet. For all three FFQs, we excluded questionnaire data that suggested biologically implausible daily energy intakes of <800 or >4,200 kcal for men or <600 or >3,500 kcal for women. This excluded 7 percent, 7 percent, and 4 percent of women and 9 percent, 4 percent, and 4 percent of men for the DHQ, Block FFQ, and Willett FFQ, respectively.

Analyses

Descriptive statistics were carried out to assess the demographics, response rates, and median nutrient intakes excluding supplements. There were 26 nutrients or dietary constituents (including the percentage of macronutrients from energy) common to the recall and the three FFQ nutrient databases for which analyses across all instruments could be performed. In each study group, nutrient correlations were estimated for each FFQ and truth with a measurement error model and 24-hour dietary recalls as the reference instrument (14) using all available recall and FFQ data. The model assumes that error in a test instrument may include systematic bias as well as within-person variation, but error in the reference instrument (here, 24-hour recalls)

TABLE 1. Demographic characteristics of respondents who completed one recall, four recalls, and at least one food frequency questionnaire, Eating at America's Table Study, 1997–1998

Demographic characteristics	Respondents who completed one recall (*n* = 1,500) (%)	Respondents who completed four recalls and were randomized to receive a FFQ* (*n* = 1,301) (%)		Respondents who completed at least one FFQ (*n* = 1,061) (%)	
		DHQ*/Block (*n* = 650)	DHQ/Willett (*n* = 650)	DHQ/Block (*n* = 514)	DHQ/Willett (*n* = 547)
Gender					
Men	50.8	50.7	49.3	48.1†	46.4†
Women	49.2	49.2	50.7	52.0	53.6
Age group (years)					
20–39	47.5	44.5	44.6	42.4	41.1†
40–59	42.6	44.8	44.6	45.9	47.4
≥60	9.9	10.6	10.9	11.7	11.5
Race/ethnicity					
White	79.0	81.5	81.9	84.1†	84.3†
African American	10.3	9.2	9.4	8.0	8.2
Latinos	5.3	4.5	4.6	3.7	3.8
Other/unknown	5.3	4.8	4.2	4.3	3.7
Education (years)					
<12	6.1	4.8	4.8	4.7	4.9
12	18.7	19.4	18.9	18.3	17.7
>12	75.1	75.9	76.3	77.0	77.3

* FFQ, food frequency questionnaire; DHQ, Diet History Questionnaire.
† Proportions for demographic group are significantly ($p \leq 0.05$) different from those completing one recall by chi-squared test.

includes only within-person random error uncorrelated with error in any FFQ. The measurement error model is written as follows:

$$Q_i = \beta_0 + \beta_1 T_i + e_i$$

$$F_{ij} = T_i + u_{ij},$$

where Q_i is the FFQ value for the ith person, T_i is the true usual intake of a given nutrient for the ith person, β_0 is the intercept and β_1 is the slope of the linear regression of Q_i on T_i, e_i is random within-person error, F_{ij} is the jth repeat reference measurement (24-hour recalls) for person i, and u_{ij} is within-person random error for person i, repeat measurement j, assumed to be independent of T_i and e_i. Essentially, the estimated correlation coefficients between the FFQ and true intakes are equivalent to deattenuated crude correlations between the FFQ and reference instruments. Because, in this case, the repeat recalls were conducted about 3 months apart, it is further assumed that random error u_{ij} and $u_{ij'}$ corresponding to four repeat measurements of the same person are independent. Except for energy and the percentage of energy from macronutrients, the measurement error models were applied both with and without energy adjustment using the residual method (13).

There are many parameters of interest when evaluating how each FFQ performs relative to true dietary intake. For each FFQ comparison, we estimated the correlation coefficient between the FFQ and true intake (ρ) and the attenuation coefficient (λ). ρ reflects validity, which is whether the FFQs measure what they are supposed to measure. The attenuation coefficient is the amount by which the log relative risk between the exposure and some disease would be distorted because of measurement error in the FFQ. Although, in general, the λ can take on any value, it typically ranges from 0 to 1 for dietary variables. A λ close to 1 indicated minimum attenuation, whereas a λ close to 0 indicates maximum attenuation.

Generally, nutrient intakes were not normally distributed for any of the instruments. Therefore, prior to modeling, by nutrient and gender, we calculated a Box-Cox power trans-

TABLE 2. Response rates for each food frequency questionnaire, Eating at America's Table Study, 1997–1998

Population	DHQ* (%)	Block (%)	Willett (%)
Total	77	76	82
Women	82	80	86
Men	72	72	78

* DHQ, Diet History Questionnaire.

TABLE 3. Median intakes for women completing one or two food frequency questionnaires and four recalls by study group and questionnaire, Eating at America's Table Study, 1997–1998

Nutrient/dietary constituent	DHQ*/Willett group				DHQ/Block group			
	Recalls (n = 254)	DHQ (n = 254)	Recalls (n = 272)	Willett (n = 272)	Recalls (n = 229)	DHQ (n = 229)	Recalls (n = 238)	Block (n = 238)
Energy (kcal)	1,687	1,630	1,693	1,682	1,694	1,555	1,651	1,541
Protein (g)	63.6	60.8	63.2	75.1†	63.7	58.2	63.3	61.1
Protein (% kcal)	15.2	15.4	15.1	17.5†	15.5	15.6	15.6	16.4
Carbohydrate (g)	213.0	214.6	213.8	213.9	216.1	202.5	210.8	179.8
Carbohydrate (% kcal)	50.9	52.5	51.1	50.7	51.1	53.5	51.1	48.0
Fat (g)	61.3	56.9	61.1	57.4	58.8	52.2	58.8	59.9
Fat (% kcal)	34.0	32.1	33.7	31.7	32.0	31.4	32.3	34.3
Saturated fat (g)	19.8	18.4	19.6	20.2	19.5	16.7	19.4	20.9
Monounsaturated fat (g)	23.6	21.0	23.2	21.4	22.1	19.4	22.2	22.7
Polyunsaturated fat (g)	12.6	12.5	12.3	10.1†	12.0	11.5	11.7	10.7
Cholesterol (mg)	188	161	181	219†	194	160†	185	191
Dietary Fiber (g)	13.6	14.3	13.3	15.1	13.0	14.0	12.7	11.9
Vitamin A (μg RE*)	803	966†	803	1,068†	857	917	819	975†
Vitamin E (mg ATE*)	7.0	7.6	6.9	8.8†	6.9	7.2	6.7	8.1†
Vitamin C (mg)	77	106†	74	113†	79	109†	79	94†
Thiamin (mg)	1.30	1.25	1.30	1.20	1.29	1.21	1.28	1.17
Riboflavin (mg)	1.55	1.55	1.53	1.54	1.56	1.51	1.52	1.59
Niacin (mg)	18.8	18.5	18.8	21.2	19.3	17.5	18.9	16.8
Vitamin B_6 (mg)	1.46	1.62	1.46	1.85†	1.47	1.54	1.44	1.46
Calcium (mg)	620	648	614	619	651	635	642	687
Iron (mg)	11.9	12.2	11.9	12.2	12.2	12.1	11.9	11.1
Magnesium (mg)	233	272†	233	267	239	250	235	232
Phosphorus (mg)	1,039	1,030	1,031	1,170	1,079	1,006	1,045	1,009
Zinc (mg)	8.8	9.0	8.7	10.3†	8.8	8.7	8.6	8.6
Potassium (mg)	2,312	2,732†	2,294	2,728†	2,337	2,644	2,312	2,304
Sodium (mg)	3,054	2,560†	3,054	1,686†	2,994	2,460†	2,961	2,287†

* DHQ, Diet History Questionnaire; RE, retinol equivalent; ATE, alpha-tocopherol equivalent.
† The food frequency questionnaire value is >15% different from the value for the recalls.

formation (15) that maximized the Shapiro-Wilk test statistic (16, 17) for the average of the four recalls. This power transformation was then applied to all instruments. After transformations, nutrient outliers were excluded from the modeling if intakes were less or greater than three times the interquartile range between the first and third quartiles for that nutrient. Exclusions across all nutrients and instruments ranged from zero persons for calcium to 14 for vitamin A. On average, three men and three women were excluded in the modeling for any given nutrient.

RESULTS

Of the 1,640 respondents (829 men and 811 women) who agreed to participate in the screening interview, 1,500, 1,418, 1,358, and 1,301 completed the first, second, third, and fourth 24-hour dietary recall, respectively. These 1,301 participants were randomized into the two study groups. Of these, 1,024 returned the first FFQ mailing, 999 completed the second, and 961 completed both FFQs. Table 1 presents the demographic characteristics of participants as the study progressed from completion of one recall, to randomization after completion of four recalls, to completion of at least one FFQ. These data show that, compared with those completing all four recalls, those completing at least one FFQ were significantly more likely to be women, older, and of White race/ethnicity. No educational differences were seen.

Table 2 presents the response rates for the instruments by gender. Response rates ranged from 72 percent for the Block FFQ and DHQ among men to 86 percent for the Willett FFQ among women.

Tables 3 and 4 show, for women and men, respectively, median intakes for 26 nutrients/dietary constituents for the recalls and FFQs by study group among respondents completing all four recalls and at least one FFQ. The values marked with a single-dagger footnote in the tables indicate FFQ values greater than 15 percent different from the recall median (after averaging the four recalls). All three FFQs estimated nutrient intakes that were closer to recalls for women than for men. For women, the DHQ- and Block FFQ-estimated nutrient intakes were fairly comparable with the recalls, though the DHQ overestimated vitamin C intake by about 38 percent. The Willett instrument's nutrient estimates tended to overestimate intakes for women and were more than 25 percent greater than the recalls for vitamins A, E, and C; vitamin B_6; and sodium. For men, underestimation was common for all FFQs, as compared with recalls, but was most common and of greater magnitude for the Willett instrument. Fats and cholesterol were among the most commonly underestimated nutrients for men across all three FFQs.

TABLE 4. Median intakes for men completing one or two food frequency questionnaires and four recalls by study group and questionnaire, Eating at America's Table Study, 1997–1998

Nutrient/dietary constituent	DHQ*/Willett group				DHQ/Block group			
	Recalls (*n* = 202)	DHQ (*n* = 202)	Recalls (*n* = 238)	Willett (*n* = 238)	Recalls (*n* = 201)	DHQ (*n* = 201)	Recalls (*n* = 226)	Block (*n* = 226)
Energy (kcal)	2,447	2,084	2,455	1,865†	2,521	2,118†	2,491	2,048†
Protein (g)	92.0	77.9†	92.6	77.4†	97.7	80.0†	97.4	84.3
Protein (% kcal)	15.1	15.3	15.2	16.6	15.9	15.1	15.9	16.6
Carbohydrate (g)	291.2	257.2	291.2	226.9†	307.5	272.6	299.4	238.7†
Carbohydrate (% kcal)	49.3	51.3	49.0	50.0	49.5	52.0	49.2	47.2
Fat (g)	94.1	69.3†	93.7	66.1†	89.6	71.7†	89.9	76.6
Fat (% kcal)	34.0	32.6	34.3	31.7	33.4	31.0	33.5	34.0
Saturated fat (g)	30.6	23.3†	30.4	23.9†	29.2	23.8†	29.3	26.4
Monounsaturated fat (g)	36.1	26.9†	36.0	25.4†	34.5	27.2†	34.8	29.9
Polyunsaturated fat (g)	18.6	15.6†	18.6	10.7†	18.8	14.8†	18.6	14.2†
Cholesterol (mg)	275	206†	281	236†	296	216†	300	278
Dietary Fiber (g)	19.6	17.3	19.1	15.9†	19.5	17.1	18.7	14.3†
Vitamin A (µg RE*)	1,075	973	1,047	982	1,145	1,073	1,134	1,170
Vitamin E (mg ATE*)	9.6	9.1	9.6	9.2	9.7	8.8	9.6	9.0
Vitamin C (mg)	106	110	106	103	106	118	95	101
Thiamin (mg)	1.92	1.58†	1.90	1.27†	1.93	1.70	1.90	1.52†
Riboflavin (mg)	2.17	1.88	2.16	1.62†	2.26	2.11	2.25	2.06
Niacin (mg)	28.3	24.8	28.3	22.3†	28.8	25.2	28.7	24.0†
Vitamin B_6 (mg)	2.14	1.97	2.14	1.90	2.30	2.14	2.28	1.94
Calcium (mg)	856	765	855	638†	883	866	864	779
Iron (mg)	18.6	16.0	18.5	12.3†	18.6	16.0	18.1	14.3†
Magnesium (mg)	344	339	349	278†	355	349	350	282†
Phosphorus (mg)	1,472	1,273	1,467	1,179†	1,549	1,395	1,519	1,315
Zinc (mg)	13.3	11.9	13.3	11.3	13.9	12.2	13.5	11.4†
Potassium (mg)	3,188	3,334	3,179	2,784	3,500	3,437	3,429	2,859†
Sodium (mg)	4,402	3.295†	4,405	1,753†	4,412	3,387†	4,382	2,924†

* DHQ, Diet History Questionnaire; RE, retinol equivalent; ATE, alpha-tocopherol equivalent.
† The food frequency questionnaire value is >15% different from the value for the recalls.

TABLE 5. Deattenuated correlations (*p*) between the food frequency questionnaires and truth using a measurement error model, unadjusted and adjusted for energy intake, Eating at America's Table Study, 1997–1998

Nutrient	Women			Men		
	DHQ*	Block	Willett	DHQ	Block	Willett
Energy						
Unadjusted	0.48	0.45	0.18	0.49	0.45	0.20
Protein						
Unadjusted	0.46	0.43	0.14	0.47	0.47	0.30
Adjusted	0.60	0.53	0.54	0.57	0.61	0.58
% kcal protein	0.61	0.50	0.53	0.55	0.56	0.55
Carbohydrate						
Unadjusted	0.50	0.49	0.33	0.53	0.46	0.40
Adjusted	0.69	0.66	0.63	0.63	0.64	0.67
% kcal carbohydrate	0.69	0.65	0.65	0.68	0.70	0.71
Fat						
Unadjusted	0.55	0.53	0.30	0.52	0.53	0.25
Adjusted	0.66	0.67	0.65	0.62	0.55	0.60
% kcal fat	0.67	0.66	0.64	0.66	0.60	0.65
Saturated fat						
Unadjusted	0.60	0.56	0.41	0.57	0.58	0.34
Adjusted	0.66	0.65	0.66	0.68	0.67	0.66
Monounsaturated fat						
Unadjusted	0.56	0.55	0.31	0.51	0.56	0.27
Adjusted	0.62	0.60	0.64	0.60	0.54	0.63
Polyunsaturated fat						
Unadjusted	0.48	0.43	0.19	0.52	0.41	0.24
Adjusted	0.64	0.48	0.47	0.61	0.33	0.52
Cholesterol						
Unadjusted	0.61	0.55	0.52	0.48	0.57	0.52
Adjusted	0.66	0.57	0.68	0.64	0.67	0.70
Fiber						
Unadjusted	0.60	0.62	0.41	0.62	0.62	0.44
Adjusted	0.77	0.80	0.68	0.80	0.77	0.73
Vitamin A						
Unadjusted	0.58	0.47	0.43	0.60	0.59	0.43
Adjusted	0.62	0.50	0.54	0.69	0.65	0.51
Vitamin E						
Unadjusted	0.43	0.28	0.17	0.55	0.39	0.23
Adjusted	0.51	0.19	0.46	0.57	0.24	0.48
Vitamin C						
Unadjusted	0.61	0.65	0.54	0.66	0.60	0.61
Adjusted	0.68	0.72	0.71	0.74	0.70	0.78
Thiamin						
Unadjusted	0.54	0.45	0.33	0.61	0.56	0.54
Adjusted	0.65	0.57	0.57	0.83	0.72	0.76

Table continues

Table 5 shows the estimated correlations between truth and the FFQs for the transformed nutrients with and without energy adjustment, by gender. Correlations of energy with truth for both genders for the DHQ and Block questionnaire were close to 0.5; for the Willett instrument, correlations were about 0.2. Generally, correlations for the Willett instrument were substantially lower than either the DHQ or Block FFQ before energy adjustment. After energy adjustment, however, these correlations improved dramatically for the Willett FFQ, while for the DHQ and Block FFQ, they also improved, but less so. The net effect was that, after energy adjustment, the correlations for all three instruments became more similar to one another. Overall, the DHQ and Block FFQ performed most similarly to one another, and both per-

TABLE 5. Continued

Nutrient	Women			Men		
	DHQ	Block	Willett	DHQ	Block	Willett
Riboflavin						
Unadjusted	0.58	0.53	0.49	0.63	0.60	0.55
Adjusted	0.66	0.59	0.69	0.78	0.68	0.79
Niacin						
Unadjusted	0.48	0.34	0.25	0.55	0.56	0.40
Adjusted	0.64	0.46	0.53	0.65	0.59	0.57
Vitamin B_6						
Unadjusted	0.54	0.48	0.33	0.65	0.63	0.52
Adjusted	0.65	0.57	0.61	0.79	0.64	0.70
Calcium						
Unadjusted	0.66	0.62	0.55	0.69	0.62	0.57
Adjusted	0.73	0.66	0.67	0.81	0.72	0.79
Iron						
Unadjusted	0.49	0.42	0.30	0.59	0.55	0.46
Adjusted	0.59	0.57	0.52	0.71	0.71	0.67
Magnesium						
Unadjusted	0.57	0.60	0.44	0.61	0.64	0.43
Adjusted	0.78	0.81	0.83	0.79	0.76	0.82
Phosphorus						
Unadjusted	0.55	0.55	0.33	0.58	0.56	0.39
Adjusted	0.69	0.62	0.65	0.73	0.68	0.77
Zinc						
Unadjusted	0.46	0.48	0.28	0.42	0.44	0.37
Adjusted	0.51	0.58	0.43	0.49	0.42	0.66
Potassium						
Unadjusted	0.59	0.59	0.38	0.58	0.60	0.43
Adjusted	0.76	0.75	0.75	0.76	0.72	0.79
Sodium						
Unadjusted	0.45	0.44	0.07	0.36	0.43	0.23
Adjusted	0.53	0.44	0.28	0.41	0.28	0.30

* DHQ, Diet History Questionnaire.

formed differently from the Willett FFQ; the DHQ correlations were generally as good as or better than those for either the Block FFQ or the Willett FFQ after adjustment.

Table 6 shows selected attenuation coefficients for nutrients. The findings are similar to those shown in table 5, with improvement from unadjusted to adjusted greatest for the Willett instrument. Overall, the DHQ had the highest coefficients (smallest attenuation) for both men and women.

Given the interpretation considerations related to multiple testing, we chose not to test for significant differences in correlations between instruments. Table 7 summarizes, through counts and means, information from tables 3–6 for purposes of descriptive comparison of the FFQs. With respect to how closely each instrument estimated nutrient intakes compared with the recalls, the DHQ was superior for men. The Willett instrument performed the least favorably with up to 65 percent of the nutrient values for men 15 percent less or greater than the median recall values. For correlations, two comparisons are shown. The first is a count of how many times the correlations were highest for each FFQ, and the second is the mean of the correlations across all nutrients. The DHQ correlations were highest for most nutrients in all analyses. The mean correlation coefficients were also higher for the DHQ, though never by more than 0.07 for the energy-adjusted values. The DHQ was highest with respect to attenuation except for unadjusted values for men, though the mean coefficients were similar for all instruments in energy-adjusted models.

DISCUSSION

These data indicate that the DHQ and Block instruments tended to track together in the magnitude of correlations before energy adjustment and to increase similarly after adjustment compared with the Willett FFQ. This is not surprising. These two instruments rely on national dietary data to develop the food lists, portion sizes, and nutrient databases (7, 12), ask specific portion size questions, and assign different portion size values to men and women. The Willett

TABLE 6. Selected attenuation coefficients (λ) between the food frequency questionnaire and truth using a measurement error model, unadjusted and adjusted for energy intake, Eating at America's Table Study, 1997–1998

Nutrient	Women			Men		
	DHQ*	Block	Willett	DHQ	Block	Willett
Energy						
Unadjusted	0.39	0.37	0.16	0.40	0.41	0.18
Protein						
Unadjusted	0.33	0.32	0.10	0.36	0.37	0.24
Adjusted	0.40	0.35	0.31	0.46	0.44	0.44
Carbohydrate						
Unadjusted	0.42	0.41	0.26	0.46	0.43	0.34
Adjusted	0.53	0.50	0.44	0.57	0.56	0.60
Fat						
Unadjusted	0.44	0.42	0.27	0.42	0.46	0.23
Adjusted	0.47	0.51	0.49	0.47	0.41	0.45
Saturated fat						
Unadjusted	0.48	0.46	0.37	0.47	0.52	0.32
Adjusted	0.49	0.58	0.52	0.55	0.55	0.53
Monounsaturated fat						
Unadjusted	0.44	0.42	0.27	0.41	0.49	0.24
Adjusted	0.44	0.42	0.45	0.45	0.42	0.46
Polyunsaturated fat						
Unadjusted	0.37	0.31	0.18	0.40	0.33	0.21
Adjusted	0.46	0.33	0.38	0.42	0.21	0.42
Fiber						
Unadjusted	0.49	0.59	0.33	0.53	0.63	0.38
Adjusted	0.66	0.82	0.56	0.71	0.81	0.69
Vitamin A						
Unadjusted	0.48	0.44	0.30	0.56	0.59	0.37
Adjusted	0.47	0.47	0.33	0.64	0.64	0.41
Vitamin E						
Unadjusted	0.31	0.17	0.14	0.39	0.25	0.19
Adjusted	0.31	0.10	0.33	0.30	0.11	0.33
Vitamin C						
Unadjusted	0.53	0.64	0.48	0.62	0.67	0.60
Adjusted	0.58	0.70	0.64	0.67	0.73	0.76

* DHQ, Diet History Questionnaire.

instrument does not specifically ask portion size, and the analytical software used at Harvard assigns identical portion sizes for men and women.

These validation data are generally comparable with those found from previous studies of various versions of the Block FFQ and the Willett FFQ (18–28). Direct comparisons of the present data with this previous research are difficult because of the variations in analytical methods used for assessing correlation coefficients (crude vs. deattenuated) and adjusting for energy intake (none vs. nutrient density or residual methods). In addition, some of the previous studies use different reference instruments (records or recalls of varying numbers of days), have varying sample sizes, and consist of distinctly specific population groups (female nurses, male health professionals, prepregnant/pregnant women, etc.). Because this study was not embedded in a larger epidemiologic study, sampling could occur randomly nationwide, across a wide age range. Though, no doubt, willingness to participate leads to some self-selection, these data represent a more general population than most other validations.

An unexpected finding was the generally lower deattenuated correlations for the Willett FFQ and the dramatic impact of energy adjustment in improving these values compared with the Block and DHQ instruments. This suggests that, for the purposes of assessing absolute intakes, the Block FFQ and DHQ are substantially better choices than the Willett FFQ and that energy adjustment must be used when using the Willett instrument. Generally, the correlations for all three instruments are higher and more comparable after energy adjustment, indicating that this may reduce error in FFQ reporting.

Two important questions are why the Willett instrument produced lower correlations for absolute intakes than the

TABLE 7. Summary of correlation and attenuation data by gender and food frequency questionnaire, Eating at America's Table Study, 1997–1998

	% of times medians were < or > 15% of recall values*	No. of times correlations were hightest†	Mean of correlation coefficients	No. of times attenuation coefficients were highest†,‡	Mean of attenuation coefficients†,‡
			Women		
Diet History Questionnaire	15				
Unadjusted		20	0.52	18	0.41
Adjusted		17	0.63	15	0.48
Block questionnaire	15				
Unadjusted		5	0.48	5	0.40
Adjusted		4	0.54	9	0.46
Willett questionnaire	42				
Unadjusted		0	0.32	0	0.26
Adjusted		5	0.58	5	0.44
			Men		
Diet History Questionnaire	31				
Unadjusted		14	0.53	9	0.44
Adjusted		14	0.65	14	0.52
Block questionnaire	42				
Unadjusted		11	0.52	15	0.45
Adjusted		3	0.58	6	0.48
Willett questionnaire	65				
Unadjusted		0	0.38	1	0.32
Adjusted		9	0.63	9	0.50

* Data based on 26 nutrients (see table 1).

† Data based on 23 and 25 nutrients for adjusted and energy-adjusted data, respectively (% of energy from protein, carbohydrate, and fat included in energy-adjusted data); ties counted toward both instruments in calculating the number of times a food frequency questionnaire correlation or attenuation coefficient was higher.

‡ Data include all nutrients, not just those selected for presentation in table 6.

other two FFQs and why energy adjustment was so much more important for the Willett than for the DHQ or Block FFQ. A possible explanation for the differences in absolute intakes may be due to how portion size information is handled in the three FFQs. Unlike the other two instruments, the Willett FFQ does not ask portion size, and the standard data processing program provided by Harvard does not assign different portion sizes to men and women. Interestingly, this is not the case in Harvard cohort studies (Laura Sampson, Harvard School of Public Health, personal communication, 2001). There is evidence in our study that this lack of specificity in portion sizes creates error in absolute nutrient estimates: The Willett instrument tends to underestimate the nutrient intakes of men and to overestimate the nutrient intakes of women. In a highly heterogeneous population, such as in the present study, the assumption of a standard portion size across all groups is likely to create additional substantial error, diminishing all correlations. This may explain the difference in the correlation coefficients between the Willett FFQ and the other two FFQs.

Energy adjustment, in general, appears to reduce measurement error in all FFQs. Energy comes from nearly all items asked and therefore may serve as a good surrogate variable to adjust for all other nutrients. In this study, too, the correlation coefficients in all three instruments improved after energy adjustment. The particularly dramatic improvement in the Willett FFQ may be due to a substantial portion size error in absolute intakes that is reduced by energy adjustment.

The measurement model underlying all the calculations in this paper is based on the assumptions that error in the reference instrument (here the 24-hour recall) is unbiased and contains only within-person variation uncorrelated with errors in the FFQ. Recent evidence suggests that these assumptions may be unwarranted for self-reported reference instruments. Studies involving biomarkers, such as doubly labeled water for measuring energy intake and urinary nitrogen for protein intake (29–35), suggest that reports using food records or recalls are biased (on average toward underreporting) and that persons may systematically differ in their reporting accuracy. This could mean that all dietary report instruments involve bias at the individual level. Part of the bias may depend on true intake and manifest itself in what is often called a flattened slope syndrome. Part of the bias may also be person specific (36) and correlate with its counterpart in the FFQ.

For this reason, Kipnis et al. (36, 37) proposed a new measurement error model that allows for correlated person-specific biases in the dietary report reference instrument and in the FFQ. Using sensitivity analysis (36), they showed that, if the correlation between person-specific biases in the FFQ and reference instrument was 0.3 or greater, the usual measurement error model would be seriously wrong. It would substantially overestimate the correlation between nutrient and true intake and the slope of the regression of the FFQ on true intake, and it would underestimate true attenuation. In a subsequent paper (37), they estimated the correlation between person-specific biases in the 4-day weighed food record and a version of the Willett FFQ to be at least 0.35. However, currently we lack information on the magnitude of this correlation for recalls and different questionnaires and do not know how this correlation, if it exists, may alter our results.

Using this new model, one can speculate that the correlation between person-specific biases in 24-hour recalls and the Block FFQ or DHQ may be somewhat higher than the Willett FFQ because of similarities in how portion size is asked (these two FFQs and recalls require respondents to recall portion size). This may lead to somewhat inflated correlation coefficients for absolute intakes for the DHQ and Block instruments and be part of the explanation for why the Willett instrument produced substantially lower correlation coefficients with true intake. We think, however, that the correlation between the errors in FFQs and recalls induced by portion size questions constitutes only a small part of the correlation between person-specific biases; these biases may be induced by many other factors related to self-reporting, such as other personality and cultural characteristics. Therefore, the additional correlations between errors in different instruments that may be caused by portion size questions are likely to explain only a small part of the discrepancy in the estimated correlations between three different FFQs and truth.

To better understand the structure of measurement error in FFQs, further research needs to be conducted in varied populations with biomarker reference instruments that are not based on self-report, such as urinary nitrogen and doubly labeled water. Such research will provide the necessary information to assess properties of different FFQs both for absolute intakes and intakes after energy adjustment.

At the time the study was conducted, we used the versions of the Willett and Block instruments most widely used in diet and epidemiologic research. The Block instrument, originally created in the early 1990s at the National Cancer Institute, was updated in 1992, 1995, and 1998. Several of the cognitive improvements suggested in our previous research (8) and integral to the DHQ have been incorporated in the 1998 version. However, the method by which portion size is now asked is quite different from the original versions. The most recent Willett instrument, too, has been modified to incorporate more low-fat foods. Validation work on both of these newer instruments needs to be conducted to ensure and justify their use.

The DHQ represents a new FFQ available to researchers. Most investigators pause at its 36-page length. However, most of this length is due not so much to the increased amount of information asked, but rather to a non-grid formatting style that repeats response categories. Cognitively, this is recommended by survey designers because the questions and answers appear as a unit for every item and because the navigational flow is consistent across questions. Further, this format allows for more flexibility in asking nonstandard questions within the instrument. This study, as well as one earlier investigation in a large prospective screening trial (38), has shown that the increased page length or time required to complete the DHQ (about 1 hour) does not materially affect response rates. The screening trial data also showed that the non-grid format led to fewer missing data and scanning errors, particularly for portion size. The DHQ will, however, lead to increased printing, scanning, and mailing costs (though the software and nutrient analyses are in the public domain).

The current study includes a unique study design and new standard for validation by including a comparison of our new FFQ directly with two widely used FFQs. These data show clearly that the efforts put into creating and testing a cognitively based FFQ (8, 9, 12, 38) were worthwhile in a domain of data collection that is, at this point in time, still primarily self-reported. Overall the DHQ is as good as or better than the two FFQs with which it was compared. This makes it both a reasonable alternative for investigators to use in their diet research and a candidate for further development in diverse populations. Information about obtaining and using the DHQ can be found at http://appliedresearch.cancer.gov/DHQ/index.html.

REFERENCES

1. Greenwald P, Clifford C. Dietary prevention. In: Greenwald P, Kramer BS, Weed DL, eds. Cancer prevention and control. New York, NY: Marcel Dekker, Inc, 1995:303–27.
2. World Cancer Research Fund, American Institute for Cancer Research. Food, nutrition and the prevention of cancer: a global perspective. Washington, DC: American Institute for Cancer Research, 1997.
3. Freedman LS, Schatzkin A, Wax Y. The impact of dietary measurement error on planning sample size required in a cohort study. Am J Epidemiol 1990;132:1185–95.
4. Thompson FE, Byers T. Dietary assessment resource manual. J Nutr 1994;124(suppl):2245S–317S.
5. Friedenreich CM. Improving long-term recall in epidemiologic studies. Epidemiology 1994;5:1–4.
6. Buzzard IM, Sievert YA. Research priorities and recommendations for dietary assessment methodology. First International Conference on Dietary Assessment Methods. Am J Clin Nutr 1994;59(suppl):275S–80S.
7. Block G, Hartman AM, Dresser CM, et al. A data-based approach to diet questionnaire design and testing. Am J Epidemiol 1986;124:453–69.
8. Subar AF, Thompson FE, Smith AF, et al. Improving food frequency questionnaires: a qualitative approach using cognitive interviewing. J Am Diet Assoc 1995;95:781–8.
9. Thompson FE, Subar AF, Brown CC, et al. Cognitive research enhances accuracy of food frequency questionnaire reports: results of an experimental validation study. J Am Diet Assoc (in press).
10. Agricultural Research Service, US Department of Agriculture. Design and operation: the continuing survey of food intakes by individuals and the diet and health knowledge survey, 1994–96. Washington, DC: US Department of Agriculture, 1997.
11. Subar AF, Krebs-Smith SM, Cook A, et al. Dietary sources of nutrients among US adults, 1989 to 1991. J Am Diet Assoc 1998;98:537–47.
12. Subar AF, Midthune D, Kulldorff M, et al. An evaluation of

alternative approaches to assign nutrient values to food groups in food frequency questionnaires. Am J Epidemiol 2000;152:279–86.
13. Willett WC. Nutritional epidemiology. New York, NY: Oxford University Press, 1998.
14. Freedman LS, Carroll RJ, Wax Y. Estimating the relation between dietary intake obtained from a food frequency questionnaire and true average intake. Am J Epidemiol 1991;134:310–20.
15. Box G, Cox D. An analysis of transformations. J R Stat Soc (B) 1964;26:211–52.
16. Shapiro S, Wilk M. An analysis of variance test for normality (complete samples). Biometrika 1965;52:591–611.
17. Shapiro S, Francia R. An approximate analysis of variance test for normality. J Am Stat Assoc 1973;67:215–16.
18. Willett WC, Sampson L, Stampfer MJ, et al. Reproducibility and validity of a semiquantitative food frequency questionnaire. Am J Epidemiol 1985;122:51–65.
19. Rimm EB, Giovannucci EL, Stampfer MJ, et al. Reproducibility and validity of an expanded self-administered semiquantitative food frequency questionnaire among male health professionals. Am J Epidemiol 1992;135:1114–26.
20. Munger RG, Folsom AR, Kushi LH, et al. Dietary assessment of older Iowa women with a food frequency questionnaire: nutrient intake, reproducibility, and comparison with 24-hour dietary recall interviews. Am J Epidemiol 1992;136:192–200.
21. Longnecker MP, Lissner L, Holden JM, et al. The reproducibility and validity of a self-administered semiquantitative food frequency questionnaire in subjects from South Dakota and Wyoming. Epidemiology 1993;4:356–65.
22. Brown JE, Buzzard IM, Jacobs DR Jr, et al. A food frequency questionnaire can detect pregnancy-related changes in diet. J Am Diet Assoc 1996;96:262–6.
23. Rockett HR, Breitenbach M, Frazier AL, et al. Validation of a youth/adolescent food frequency questionnaire. Prev Med 1997;26:808–16.
24. Block G, Woods M, Potosky AL, et al. Validation of a self-administered diet history questionnaire using multiple diet records. J Clin Epidemiol 1990;43:1327–35.
25. Block G, Thompson FE, Hartman A, et al. Comparison of two dietary questionnaires validated against multiple dietary records collected during a 1-year period. J Am Diet Assoc 1992;92:686–93.
26. Mares-Perlman JA, Klein BE, Klein R, et al. A diet history questionnaire ranks nutrient intakes in middle-aged and older men and women similarly to multiple food records. J Nutr 1993;123:489–501.
27. Potischman N, Carroll RJ, Iturria SJ, et al. Comparison of the 60- and 100-item NCI-Block questionnaires with validation data. Nutr Cancer 1999;34:70–5.
28. Bingham SA, Day NE. Using biochemical markers to assess the validity of prospective dietary assessment methods and the effect of energy adjustment. Am J Clin Nutr 1997;65(suppl):1130S–7S.
29. Bandini LG, Schoeller DA, Cyr HN, et al. Validity of reported energy intake in obese and nonobese adolescents. Am J Clin Nutr 1990;52:421–5.
30. Livingstone MBE, Prentice AM, Strain JJ, et al. Accuracy of weighed dietary records in studies of diet and health. BMJ 1990;300:708–12.
31. Heitmann BL. The influence of fatness, weight change, slimming history and other lifestyle variables on diet reporting in Danish men and women aged 35–65 years. Int J Obes 1993;17:329–36.
32. Heitmann BL, Lissner L. Dietary underreporting by obese individuals—is it specific or non-specific? BMJ 1995;311:986–9.
33. Martin LJ, Su W, Jones PJ, et al. Comparison of energy intakes determined by food records and doubly labeled water in women participating in a dietary-intervention trial. Am J Clin Nutr 1996;63:483–90.
34. Sawaya AL, Tucker K, Tsay R, et al. Evaluation of four methods for determining energy intake in young and older women: comparison with doubly labeled water measurements of total energy expenditure. Am J Clin Nutr 1996;63:491–9.
35. Black AE, Bingham SA, Johansson G, et al. Validation of dietary intakes of protein and energy against 24 hour urinary N and DLW energy expenditure in middle-aged women, retired men and post-obese subjects: comparisons with validation against presumed energy requirements. Eur J Clin Nutr 1997;51:405–13.
36. Kipnis V, Carroll RJ, Freedman LS, et al. A new dietary measurement error model and its implications for the estimation of relative risk: application to four calibration studies. Am J Epidemiol 1999;150:642–51.
37. Kipnis V, Midthune D, Freedman LS, et al. Empirical evidence of correlated biases in dietary assessment instruments and its implications. Am J Epidemiol 2001;153:394–403.
38. Subar AF, Ziegler RG, Thompson FE, et al. Is shorter always better?: relative importance of dietary questionnaire length and cognitive ease on response rates and data quality. Am J Epidemiol 2001;153:404–9.