American Journal of Epidemiology

Vol. 174, No. 5
DOI: 10.1093/aje/kwr140
Advance Access publication: July 15, 2011

Practice of Epidemiology

Evaluation and Comparison of Food Records, Recalls, and Frequencies for Energy and Protein Assessment by Using Recovery Biomarkers

Ross L. Prentice*, Yasmin Mossavar-Rahmani, Ying Huang, Linda Van Horn, Shirley A. A. Beresford, Bette Caan, Lesley Tinker, Dale Schoeller, Sheila Bingham†, Charles B. Eaton, Cynthia Thomson, Karen C. Johnson, Judy Ockene, Gloria Sarto, Gerardo Heiss, and Marian L. Neuhouser

* Correspondence to Dr. Ross L. Prentice, Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, P.O. Box 19024, Seattle, WA 98109-1024 (e-mail: rprentic@whi.org).
† Deceased.

Initially submitted January 21, 2011; accepted for publication April 4, 2011.

The food frequency questionnaire approach to dietary assessment is ubiquitous in nutritional epidemiology research. Food records and recalls provide approaches that may also be adaptable for use in large epidemiologic cohorts, if warranted by better measurement properties. The authors collected (2007–2009) a 4-day food record, three 24-hour dietary recalls, and a food frequency questionnaire from 450 postmenopausal women in the Women's Health Initiative prospective cohort study (enrollment, 1994–1998), along with biomarkers of energy and protein consumption. Through comparison with biomarkers, the food record is shown to provide a stronger estimate of energy and protein than does the food frequency questionnaire, with 24-hour recalls mostly intermediate. Differences were smaller and nonsignificant for protein density. Food frequencies, records, and recalls were, respectively, able to "explain" 3.8%, 7.8%, and 2.8% of biomarker variation for energy; 8.4%, 22.6%, and 16.2% of biomarker variation for protein; and 6.5%, 11.0%, and 7.0% of biomarker variation for protein density. However, calibration equations that include body mass index, age, and ethnicity substantially improve these numbers to 41.7%, 44.7%, and 42.1% for energy; 20.3%, 32.7%, and 28.4% for protein; and 8.7%, 14.4%, and 10.4% for protein density. Calibration equations using any of the assessment procedures may yield suitable consumption estimates for epidemiologic study purposes.

bias (epidemiology); biological markers; diet; energy intake; epidemiologic methods; measurement error; nutrition assessment

Abbreviations: NPAAS, Nutrition and Physical Activity Assessment Study; SE, standard error; WHI, Women's Health Initiative.

Reliable information on the health effects of diet and nutrition on chronic disease is crucial to formulating appropriate dietary recommendations for individuals and to instituting food policy changes that may be needed to reverse the national obesity epidemic. However, in spite of clear obesity associations with major cardiovascular diseases and cancers, few diet and chronic disease associations are regarded as convincing or even probable (1, 2).

The food frequency questionnaire has been ubiquitous in nutritional epidemiology for the past 25 years, because its self-administered and machine-readable features make it practical and cost-effective for application to large epidemiologic cohorts. Other more detailed dietary assessment approaches, including food records (diaries) and dietary recalls, were applied retrospectively in early case-control studies. Prospective use of these approaches may offer cognitive advantages compared with the food frequency questionnaire, prompting a substantial effort to develop an automated, self-administered 24-hour recall (3).

A few cohort studies have collected food records prospectively, with subsequent nutrient analyses in a case-control mode. Positive associations between dietary fat and breast

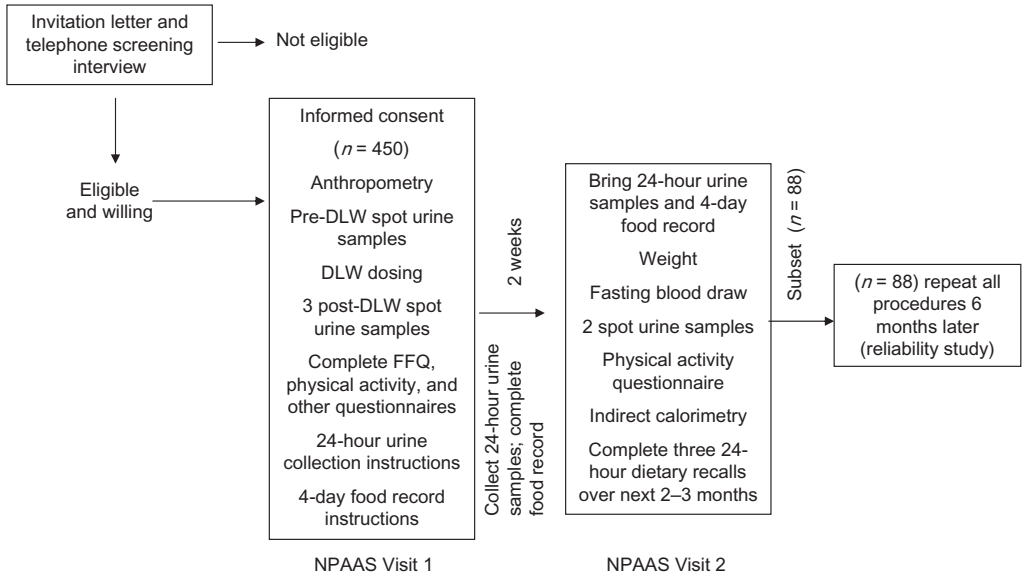

Figure 1. Women's Health Initiative Nutrition and Physical Activity Assessment Study (NPAAS; 2007–2009) procedures. DLW, doubly labeled water; FFQ, food frequency questionnaire.

cancer (4, 5) and an inverse association of fiber consumption and colorectal cancer (6) based on food records have been reported that were not evident from corresponding food frequency questionnaire data. These analyses highlight the importance of the dietary measurement error issue, but they do not indicate whether any available dietary approach leads to reliable diet and disease information.

The availability of urinary recovery biomarkers (7) for some dietary components allows the relative and absolute performance of dietary assessment methods to be evaluated in relation to short-term consumption. The Observing Protein and Energy Nutrition (OPEN) Study, among 484 men and women in Maryland, reported better measurement properties for 24-hour dietary recalls compared with food frequency questionnaires for energy and protein, both absolute and relative (8, 9), while a biomarker substudy among 179 men and women in the European Prospective Investigation of Cancer (EPIC)-Norfolk cohort reported better properties for 7-day food records compared with food frequency questionnaires for protein, potassium, and sodium consumptions (10), at least for absolute intakes (11). These studies reported measurement errors to be positively correlated among assessment procedures, arguing that a biomarker, rather than a second self-report, be used as ''reference'' instrument for measurement error correction.

Our Nutrient Biomarker Study among 544 postmenopausal women from the Women's Health Initiative (WHI) Dietary Modification trial of a low-fat eating pattern found only a weak correlation between food frequency questionnaire assessments of energy and protein consumptions and corresponding consumption biomarkers (12). Moreover, the food frequency questionnaire was found to incorporate important systematic biases related to body mass index, age, and ethnicity. Regression calibration equations were developed to provide estimates of energy, protein, and protein density (fraction of energy from protein) that incorporate adjustments for systematic and random aspects of measurement error. These equations were used to generate ''calibrated'' consumption estimates throughout WHI cohorts. Calibrated energy was found to be positively associated with total and site-specific cancer incidence (13) and with coronary disease (14) in WHI cohorts. These associations were not apparent from food frequency questionnaire consumption estimates without calibration. They appeared to be substantially mediated by body fat accumulation over time (13–15).

Important questions remain concerning the development and use of calibrated energy and protein consumption estimates: 1) Are the ''signal strengths'' from food frequency questionnaires, food records, and 24-hour dietary recalls materially different in corresponding calibration equations?; 2) To what extent can the calibration procedures from any of the 3 assessment procedures recover the nutrient consumption variation in the study population?; and 3) Are calibration equations transferable among study cohorts?

To address these questions, we conducted a further biomarker study, this time among 450 women enrolled in the WHI Observational Study. This Nutrition and Physical Activity Assessment Study (NPAAS) included the WHI food frequency questionnaire, a 4-day food record, and three 24-hour dietary recalls, along with doubly labeled water and urinary nitrogen assessments of energy and protein consumptions. Calibration equations involving the 3 dietary assessment procedures individually and combined were compared for their ability to explain variation among study subjects in the biomarker assessments, and food

Table 1. Baseline (1994–1998) Demographic and Lifestyle Characteristics of Participants in the NPAAS and Participants in the WHI Observational Study But Not the NPAAS

Characteristic	NPAAS (n = 450)		WHI Observational Study (n = 93,226)		P Value[a]
	No.	%	No.	%	
Age at WHI enrollment, years					
<60	304	67.6	29,406	31.5	<0.0001
60–69	119	26.4	41,081	44.1	
≥70	27	6.0	22,739	24.4	
Body mass index, kg/m^2					
<25	179	39.8	37,617	40.8	<0.0001
25–29	106	23.6	31,356	34	
≥30	165	36.7	23,148	25.1	
Race/ethnicity					
Black	84	18.7	7,551	8.1	<0.0001
Hispanic	64	14.2	3,545	3.8	
Other minority	14	3.1	4,402	4.7	
White	288	64.0	77,728	83.4	
Annual family income, $					
<20,000	43	9.9	13,975	16.2	0.0002
20,000–34,999	92	21.2	20,134	23.3	
35,000–49,999	84	19.4	17,346	20.1	
50,000–74,999	98	22.6	17,389	20.1	
≥75,000	117	27.0	17,491	20.3	
Education					
College degree or higher	226	50.6	38,777	41.9	0.0003
High school diploma/GED	48	10.7	15,074	16.3	
Less than high school	16	3.6	4,833	5.2	
School after high school	157	35.1	33,778	36.5	
Current smoking	21	4.7	5,769	6.3	0.21
Any use of dietary supplements	306	68.0	67,445	72.3	0.045
Recreational episodes/week					
<2	63	14.1	18,096	19.6	0.0073
2–4	81	18.1	17,324	18.8	
>4	303	67.8	56,758	61.6	

Abbreviations: GED, general equivalency diploma; NPAAS, Women's Health Initiative Nutrition and Physical Activity Assessment Study; WHI, Women's Health Initiative.

[a] The P values comparing the NPAAS and WHI Observational Study are based on chi-square tests.

frequency questionnaire calibration equations from the 2 WHI biomarker studies were compared to examine the transferability question.

MATERIALS AND METHODS

The WHI Observational Study and Dietary Modification trial

The WHI Observational Study is a prospective cohort study that enrolled 93,676 postmenopausal women in the age range 50–79 years during 1994–1998 (16, 17) at 40 US clinical centers. The Observational Study has considerable commonality with the Dietary Modification trial (16) among 48,835 postmenopausal women, in which the Nutrient Biomarker Study was conducted. The Observational Study and Dietary Modification cohorts were drawn from the same catchment populations, with substantial overlap in baseline data collection and in outcome ascertainment during cohort follow-up. The WHI food frequency questionnaire (18) was administered at baseline and at 3-years in the Observational Study and at baseline and 1-year in the Dietary Modification trial, where a baseline 4-day food record was also obtained.

The Nutrition and Physical Activity Assessment Study

NPAAS enrolled 450 postmenopausal women from the WHI Observational Study. Black and Hispanic women were oversampled to support comparisons of measurement properties among racial/ethnic groups. Three participating clinical centers recruited primarily these minority groups, with an odds ratio of 3 for Hispanic versus black, while the other 6 clinical centers recruited black and Hispanic women with an odds ratio of 5. Women in the extremes of body mass index were oversampled, with odds ratios of 10 and 2 for underweight women (body mass index, <18.5) and obese women (body mass index, ≥30), respectively. Because of the time lag between cohort enrollment and this biomarker substudy, younger postmenopausal women were oversampled, with odds ratios of 3 and 2 for women who were 50–54 and 55–59 years of age at enrollment. As in the Nutrient Biomarker Study, women were excluded for having any medical condition precluding participation, weight instability, or travel plans during the study period. Overall, 20.6% of women invited and screened for eligibility completed the protocol. An additional 4 women consented to, but did not complete, the study. A subsample of 88 women (19.6%) repeated the entire protocol about 6 months later to provide repeatability information. NPAAS women completed their participation in 2007–2009, with specimen analyses completed by June 2010. Study procedures were approved by the institutional review boards of participating institutions. Participants provided informed consent and received $100 upon study completion.

Study protocol and procedures

The study protocol involved 2 clinical center visits separated by a 2-week period, along with at-home activities (Figure 1). The first visit included eligibility confirmation; informed consent; anthropometric measurements; doubly labeled water dosing; training in 4-day food record

Table 2. Geometric Means and 95% Confidence Intervals for Biomarker and Self-Report Assessments of Energy and Protein Consumption in the NPAAS (2007–2009), Along With Geometric Means and 95% Confidence Intervals for Self-Report:Biomarker Assessment Ratios

Assessment	Sample Size, no.	Geometric Mean	95% CI	Ratios of Self-Report to Biomarker		
				Sample Size, no.	Geometric Mean	95% CI
			Energy, kcal/day			
Doubly labeled water assessment[a]	415	2,023	1,988, 2,058			
Food frequency questionnaire	450	1,455	1,399, 1,514	415	0.72	0.69, 0.76
4-Day food record	450	1,617	1,582, 1,652	415	0.80	0.78, 0.82
24-Hour dietary recall	447	1,556	1,519, 1,594	412	0.77	0.75, 0.79
			Protein, g/day			
Urinary nitrogen	443	69.3	67.3, 71.3			
Food frequency questionnaire	450	62.8	60.0, 65.6	443	0.91	0.87, 0.95
4-Day food record	450	66.7	65.0, 68.4	443	0.96	0.94, 0.99
24-Hour dietary recall	446	62.0	60.5, 63.6	439	0.90	0.87, 0.92
			Protein density			
Biomarker	408	13.8	13.4, 14.2			
Food frequency questionnaire	450	17.3	16.9, 17.6	408	1.25	1.22, 1.29
4-Day food record	450	16.6	16.3, 16.9	407	1.21	1.18, 1.25
24-Hour dietary recall	447	16.0	15.7, 16.3	405	1.16	1.13, 1.20

Abbreviations: CI, confidence interval for the geometric mean; NPAAS, Women's Health Initiative Nutrition and Physical Activity Assessment Study.

[a] Assessment of energy expenditure using the US average respiratory quotient.

completion; completion of food frequency questionnaire and physical activity, dietary supplement, and other questionnaires; and collection of a blood specimen and spot urine samples both before and after doubly labeled water dosing. Between the 2 clinic visits, participants completed a 4-day food record and collected 24-hour urine samples on the day prior to the second clinic visit.

At the second clinic visit, the 24-hour urine samples were received; 4-day food records were reviewed; and participants completed additional physical activity questionnaires, provided additional spot urine and fasting blood specimens, and had resting energy expenditure assessed via indirect calorimetry. The first of the 24-hour dietary recalls was obtained in the 1–3 weeks after visit 2 and then monthly thereafter for the other 2.

Recovery biomarkers

Total energy expenditure was estimated as in our previous biomarker study (19, 20). Briefly, after a 4-hour fast at visit 1, participants provided baseline urine samples, were weighed, and ingested a single dose of approximately 1.8 g of 10-atom percent oxygen-18-labeled water and 0.12 g of 99.9% deuterium-labeled water per kilogram of estimated total body water. The tracers equilibrate rapidly in body water, and the difference in elimination rates of oxygen-18 and deuterium is proportional to carbon dioxide production, from which total energy expenditure is calculated by using modified Weir equations (20). Elimination rates were estimated from 3 spot urine specimens over 4 hours following doubly labeled water dosing, with a blood specimen drawn at 3 hours post-doubly labeled water dosing among women of age $\geq$60 years used instead if corresponding spot urine specimens showed insufficient isotope enrichment. Elimination rates were also estimated from spot urine samples obtained at the second clinic visit. In weight-stable persons, total energy consumption over a 2-week period is objectively estimated by this procedure.

Similarly, protein consumption was objectively estimated by 6.25 $\times$ 24-hour urinary nitrogen $\div$ 0.81 (21). Participants collected urine for 24 hours on day 14, immediately preceding visit 2. PABACheck (para-aminobenzoic acid; Laboratories for Applied Biology, Ltd., London, United Kingdom) was used to assess the quality of urine collection (22), with recovery of 85%–110% of the dose considered as complete urine collection.

Specimen handling and quality assurance procedures were as previously described for the Nutrient Biomarker Study (12). Blind duplicates (5%) were included in the energy and protein biomarker assessments. A 6.5% quality control failure rate occurred for the doubly labeled water procedure. About half of the failures were due to low tracer enrichments or lack of equilibration, while the others were due to dilution space or other external reproducibility

Table 3. β Coefficients and Standard Errors From Regression of Log(Self-Report) − Log(Biomarker) on Body Mass Index, Age, and Ethnicity in the NPAAS (2007–2009) Among 450 Postmenopausal Women

Variable	Food Frequency Questionnaire		4-Day Food Record		24-Hour Dietary Recall	
	β	SE	β	SE	β	SE
			Energy			
Intercept	−0.645[a]	0.281	−0.267	0.163	−0.281	0.185
Body mass index	−0.0043	0.0034	−0.0114[a]	0.0020	−0.0139[a]	0.0022
Age, years	0.0075[a]	0.0035	0.0055[a]	0.0020	0.0062[a]	0.0023
Black	−0.265[a]	0.055	−0.056	0.032	−0.072[a]	0.036
Hispanic	−0.204[a]	0.061	0.0037	0.036	−0.033	0.041
Other minority	−0.220	0.117	−0.051	0.068	−0.109	0.077
			Protein			
Intercept	−0.578	0.319	−0.167	0.193	−0.458	0.204
Body mass index	0.0002	0.0039	−0.0097[a]	0.0023	−0.0106[a]	0.0025
Age, years	0.0078	0.0040	0.0055[a]	0.0024	0.0090[a]	0.0025
Black	−0.231[a]	0.064	0.095[a]	0.039	0.097[a]	0.041
Hispanic	−0.170[a]	0.068	0.010	0.041	0.010	0.044
Other minority	−0.111	0.133	−0.070	0.080	0.007	0.085
			Protein density			
Intercept	0.047	0.210	0.205	0.196	−0.135	0.206
Body mass index	0.0046	0.0025	0.0019	0.0023	0.0022	0.0025
Age, years	0.0004	0.0026	−0.0013	0.0024	0.0027	0.0026
Black	0.054	0.041	0.136[a]	0.039	0.144[a]	0.040
Hispanic	0.045	0.046	0.019	0.043	0.052	0.045
Other minority	0.064	0.086	−0.024	0.081	0.088	0.085

Abbreviations: NPAAS, Women's Health Initiative Nutrition and Physical Activity Assessment Study; SE, standard error.

[a] β Coefficient differs from zero at the 0.05 level of significance.

issues. These issues arose more frequently among elderly women.

Dietary assessment

Participants completed the self-administered WHI food frequency questionnaire (23) in English or Spanish. This food frequency questionnaire includes 122 foods or food groups, 19 adjustment questions, and 4 summary questions, and it was designed to assess typical dietary habits over the preceding 3 months in a multiethnic and geographically diverse population. Food frequency questionnaires were reviewed by clinic staff at the first clinic visit.

Participants viewed a 25-minute instructional video and received a food record instruction booklet at the first clinic visit. The English or Spanish booklet contains detailed instructions on recording food intake, including the description of food preparation methods, added fats, brand names, and ingredients of mixed dishes and recipes, and 12 questions on food-use patterns. Participants also received a 12-page serving size booklet with photographs and other measuring devices. They completed 4 days of recording on alternate days (Sunday through Saturday) between visits 1 and 2.

The 24-hour dietary recalls were conducted by trained and certified study staff by telephone, with data entered directly and computerized by using NDSR (Nutrition Data System for Research; Nutrition Coordinating Center, University of Minnesota, Minneapolis, Minnesota) software. Interviews targeted all food and beverages consumed during the previous 24 hours (midnight to midnight). The software prompts the interviewer to probe for detailed information on quantities, brands, and cooking methods, using the US Department of Agriculture multiple-pass method, assisted by the 12-page serving size booklet.

Dietary data from each of the 3 methods were analyzed for nutrient content by using the University of Minnesota nutrient database (24), which derives from the US Department of Agriculture Nutrient Database for Standard Reference and its periodic revisions.

Statistical methods

Analyses focused on log-transformed consumption estimates for each of energy, protein, and protein density, which were each approximately normally distributed. Daily food record and recall estimates were averaged over the reporting days prior to log transformation. Values that fell outside the

Table 4. Calibration Equation β Coefficients, Standard Errors, and Percentage of Biomarker Variation Explained as R^2 From Regression of Log(Biomarker) on Log(Self-Report), Body Mass Index, Age, and Ethnicity in the NPAAS (2007–2009) Among 450 Postmenopausal Women

Variable	Food Frequency Questionnaire				4-Day Food Record				24-Hour Dietary Recall				All Self-Reports			
	β	SE	R^2	Adjusted R^{2a}	β	SE	R^2	Adjusted R^2	β	SE	R^2	Adjusted R^2	β	SE	R^2	Adjusted R^2
Energy																
Intercept	7.614[b]	0.009			7.597[b]	0.009			7.607[b]	0.009			7.594[b]	0.009		
Food frequency questionnaire	0.054[b]	0.017	3.8	6.5									0.026	0.017	4.1	7.0
4-Day food record					0.161[b]	0.028	7.8	13.3					0.147[b]	0.036	5.3	9.0
24-Hour dietary recall									0.101[b]	0.026	2.8	4.8	0.004	0.033	0.2	0.3
Body mass index	0.013[b]	0.001	26.9	45.9	0.013[b]	0.001	27.0	46.0	0.013[b]	0.001	28.7	48.9	0.013[b]	0.001	25.9	44.2
Age	−0.010[b]	0.001	9.7	16.5	−0.009[b]	0.001	8.4	14.3	−0.009[b]	0.001	9.1	15.5	−0.009[b]	0.001	8.4	14.2
Black	−0.023	0.019			−0.024	0.018			−0.024	0.018			−0.013	0.018		
Hispanic	−0.062[b]	0.021	1.3	2.2	−0.065[b]	0.020	1.5	2.6	−0.063[b]	0.020	1.5	2.6	−0.056[b]	0.020	1.1	1.9
Other minority	−0.041	0.040			−0.039	0.038			−0.038	0.039			−0.031	0.038		
Total[c]			41.7	71.1			44.7	76.2			42.1	71.8			45.0	76.6
Protein																
Intercept	4.263	0.017			4.235	0.016			4.269	0.016			4.240	0.016		
Food frequency questionnaire	0.135[b]	0.021	8.4	16.4									0.006	0.029	8.6	16.8
4-Day food record					0.465[b]	0.045	22.6	44.2					0.350[b]	0.056	15.7	30.7
24-Hour dietary recall									0.404[b]	0.046	16.2	31.7	0.199[b]	0.055	1.6	3.2
Body mass index	0.012[b]	0.002	5.8	11.4	0.012[b]	0.002	5.1	10.0	0.012[b]	0.002	5.8	11.4	0.011[b]	0.002	4.3	8.5
Age	−0.012[b]	0.002	4.1	8.0	−0.009[b]	0.002	2.2	4.3	−0.011[b]	0.002	3.4	6.7	−0.009[b]	0.002	2.2	4.4
Black	−0.120[b]	0.038			−0.138[b]	0.034			−0.145[b]	0.035			−0.131[b]	0.035		
Hispanic	−0.078	0.040	2.0	3.9	−0.067	0.036	2.7	5.3	−0.069	0.037	3.0	5.9	−0.052	0.036	2.2	4.3
Other minority	−0.018	0.076			0.012	0.070			−0.026	0.072			0.006	0.06		
Total[c]			20.3	39.7			32.7	63.8			28.4	55.6			34.6	67.9

	Protein density															
Intercept	2.652[b]	0.017			2.671[b]	0.017			2.687[b]	0.018			2.679[b]	0.018		
Food frequency questionnaire	0.344[b]	0.068	6.5	38.5									0.092	0.080	6.1	36.7
4-Day food record					0.488[b]	0.067	11.0	65.9					0.360[b]	0.084	5.7	34.3
24-Hour dietary recall									0.393[b]	0.068	7.0	41.7	0.151[b]	0.083	0.5	3.2
Body mass index	−0.002	0.002	0.6	3.3	−0.001	0.002	0.5	3.2	−0.001	0.002	0.4	2.3	−0.002	0.002	0.6	3.8
Age	−0.002	0.002	0.04	0.2	−0.001	0.002	0.01	0.04	−0.003	0.002	0.1	0.7	−0.001	0.002	0.0	0.0
Black	−0.100[b]	0.037			−0.130[b]	0.036			−0.133[b]	0.037			−0.127[b]	0.036		
Hispanic	−0.043	0.041	1.7	10.0	−0.035	0.040	2.8	17.0	−0.043	0.041	2.9	17.6	−0.034	0.040	2.6	15.4
Other minority	−0.030	0.078			−0.006	0.075			−0.042	0.078			−0.016	0.076		
Total[c]			8.7	52.1			14.4	86.1			10.4	62.3			15.5	93.4

Abbreviations: NPAAS, Women's Health Initiative Nutrition and Physical Activity Assessment Study; SE, standard error.
[a] Adjusted R^2 values ($R^2 \div$ by log(biomarker) correlation in the reliability subsample) correct for biomarker measurement error.
[b] β Coefficient differs from zero at the $P = 0.05$ significance level.
[c] Total percentage of variation explained by all the variables. R^2 values for specific variables arise from analyses with only these regression variables, with subsequent rescaling so that these R^2 values add to the total regression R^2. R^2 values for race/ethnicity pertain to comparisons among the 4 groups (white, black, Hispanic, other minority).

interquartile range by more than 3 times its width were excluded as outliers. Our measurement model (25, 26) assumes a log(biomarker) assessment W to adhere to a classical measurement model,

$$W = Z + e,$$

where Z is the targeted nutritional variable, and e is an independent error term that is assumed to be independent of Z and other study subject characteristics. Z can be regarded as the logarithm of average daily consumption for the nutritional factor under study over a fairly short period of time, such as 6–12 months, in proximity to the biomarker data collection period.

A more flexible measurement model,

$$Q = a_0 + a_1 Z + a_2{}^{\mathrm{T}}\mathrm{V} + \epsilon,$$

is considered for a corresponding log-transformed self-report assessment Q. Here, V is a vector of study subject characteristics that may relate to the self-report assessment; a_0, a_1, and a_2 are regression parameters; and ϵ is an error term that is independent of Z, V, and the biomarker error e.

Initial analyses apply a more restrictive model with $a_1 = 1$ for the self-report assessments. This model permits a specific focus on systematic bias in the self-report in relation to V, through linear regression of $Q - W$ on V. Our analyses focus on body mass index, age, and ethnicity, characteristics that surfaced as the major sources of systematic bias in the Nutrient Biomarker Study (12). Age and body mass index were coded as quantitative variables, while indicator variables were used to contrast minority group women to white women.

Our principal analyses aimed to develop ''calibrated'' consumption estimates that allow for systematic and random measurement error in the self-report assessments. These involve linear regression of W on Z and V, as arises under our measurement model with a joint normality assumption (13). These regression equations allow consumption estimates to be calculated from (Q, V), for use in disease association analyses.

The percentage of biomarker variation explained (R^2) by the (log-transformed) self-report assessment in these calibration equations is used to evaluate the ''signal'' strength from the self-report, and traditional correlation coefficients between Q and W are also given. R^2 values for the calibrated consumption estimates are also examined.

The biomarker data include measurement error that may primarily reflect temporal consumption variation. The (log-transformed) biomarker values W_1 and W_2 for the initial and repeat assessments in our reliability sample are modeled as $W_1 = Z + e_1$ and $W_2 = Z + e_2$, with error terms e_1 and e_2 that are independent with a common variance, in which case the correlation between W_1 and W_2 estimates the variance of $Z \div$ the variance of W. Hence, we also provided ''adjusted'' R^2 values by dividing the R^2 values from linear regression by the squared sample biomarker correlation in the reliability subsample. The adjusted R^2 values can be considered as estimating the percentage of variation explained in the underlying Z value.

To allow for possible departures from normally distributed response variables, we used bootstrap procedures to estimate standard errors and significance levels (10,000 bootstrap samples). These procedures are particularly convenient for testing the equality of coefficients between regression analyses of W on Q and V, for differing choices of the self-report Q. Calibration equations arising from food frequency questionnaire assessments from the nonoverlapping Nutrient Biomarker Study and NPAAS data sets were compared by using likelihood ratio tests based on the combined data set.

Calibration equations were developed separately for NPAAS subsets defined by race/ethnicity and body mass index.

The urinary nitrogen biomarker was analyzed with and without exclusions based on the PABACheck assessment of urine collection completeness. Even though 14.7% of samples did not meet our completeness criteria, calibration equations differed little, and results are presented without PABACheck exclusion, as in our Nutrient Biomarker Study report (12).

RESULTS

Table 1 shows the distribution of demographic and lifestyle characteristics in NPAAS, along with those for the remainder of the Observational Study cohort. The oversampling according to race/ethnicity, body mass index, and age at enrollment is evident. NPAAS women were somewhat more highly educated, more affluent, and more frequently engaged in recreational activities compared with other cohort members.

Table 2 shows geometric means for biomarker and dietary assessments of energy, protein, and protein density, for assessments meeting quality control criteria. The geometric means of the self-report:biomarker assessment ratios are also shown. Each of the 3 self-report procedures underestimates energy substantially (20%–27%) and protein to a lesser extent (4%–10%), and each overestimates protein density compared with the biomarker (16%–25%).

Table 3 shows some results from linear regression of log(self-report) − log(biomarker) on body mass index, age at NPAAS participation, and race/ethnicity. Each of the 3 self-report procedures shows evidence of systematic biases related to 1 or more of these factors, for both energy and protein. For 4-day food record and 24-hour dietary recall assessments, energy and protein underreporting was more severe among women with a high body mass index or a younger age, while black women tended to further modestly underestimate energy and to overestimate protein and protein density. Food frequency questionnaire systematic bias patterns included greater energy underestimation by younger women and substantially greater underestimation of both energy and protein by minority group women. Food frequency questionnaire bias for energy in relation to body mass index was greater ($P < 0.05$) in corresponding analyses that excluded the ethnicity variables from the regression model. Systematic biases were not evident for food frequency questionnaire protein density.

Correlation coefficients between log-transformed biomarker and log-transformed food frequency questionnaire, 4-day food record, and 24-hour dietary recall assessments were, respectively, 0.196 (standard error (SE), 0.044), 0.297 (SE, 0.046), and 0.167 (SE, 0.051) for energy; 0.289 (SE, 0.042), 0.476 (SE, 0.043), and 0.403 (SE, 0.041) for protein; and 0.254 (SE, 0.041), 0.332 (SE, 0.049), and 0.264 (SE, 0.046) for protein density.

Table 4 shows regression coefficients from linear regression of log(biomarker) on log(self-report), as well as body mass index, age, and ethnicity, thereby adjusting for the systematic biases noted in Table 3, while also allowing these study subject characteristics to help explain biomarker variation more generally. For energy, the resulting ''calibration equations'' that use food frequency questionnaire, 4-day food record, or 24-hour dietary recall assessments, respectively, explain 41.7%, 44.7%, and 42.1% of the biomarker variation. These percentages are much larger than those from analyses using the self-report data alone (3.8%, 7.8%, and 2.8%, respectively), with much of the added value deriving from body mass index and age. For protein, the food frequency questionnaire, 4-day food record, and 24-hour dietary recall-based calibration equations provide an explanation for 20.3%, 32.7%, and 28.4% of the biomarker variation. For protein density, the corresponding percentages are 8.7%, 14.4%, and 10.4%. Calibration equations are also shown using all 3 self-reports simultaneously with the other variables. The percentages of biomarker variation explained were 45.0%, 34.6%, and 15.5% for energy, protein, and protein density. The strongest self-report ''signal'' for each of the 3 nutritional variables arises from the 4-day food record, and the variation explained is not significantly greater than that from the calibration equation with only the 4-day food record self-report for energy ($P = 0.67$), protein ($P = 0.10$), or protein density ($P = 0.23$).

The adjusted R^2 values in Table 4 suggest that the calibration equations recover a large fraction of the log-transformed consumption variation in the underlying dietary factor (e.g., 71%–77% for energy), using any of the self-report assessments, though less so for protein and protein density if the calibration procedure uses the food frequency questionnaire.

We also estimated measurement error correlations among pairs of assessment methods, under our measurement model and joint normality assumptions. For energy, the estimated measurement error correlation was 0.30 (SE, 0.05) for the food frequency questionnaire and 4-day food record, 0.30 (SE, 0.05) for the food frequency questionnaire and 24-hour dietary recall, and 0.50 (SE, 0.05) for the 4-day food record and 24-hour dietary recall. The corresponding numbers for protein were 0.35 (SE, 0.07), 0.33 (SE, 0.07), and 0.27 (SE, 0.18) and for protein density were 0.38 (SE, 0.14), 0.38 (SE, 0.12), and 0.40 (SE, 0.17).

Table 5 compares food frequency questionnaire-based calibration equations between the 2 WHI biomarker studies. Dietary Modification trial women tended to be slightly younger and of higher body mass index compared with Observational Study women. A likelihood ratio test of equality of all coefficients is not significant for protein ($P = 0.23$) or for protein density ($P = 0.95$). This test is

Table 5. Comparison of Calibration Equation β Coefficients and Standard Errors From Regression on Log(Biomarker) on Corresponding Log(Food Frequency Questionnaire), Body Mass Index, Age, and Ethnicity Between the NBS (2004–2006) and the NPAAS (2007–2009)

Variable	NBS		NPAAS		*P* Values		
	β	SE	β	SE	1[a]	2[a]	3[a]
			Energy				
Intercept	7.628[b]	0.006	7.614[b]	0.009	0.206	0.003	0.006
Food frequency questionnaire	0.058[b]	0.016	0.054[b]	0.017	0.858		
Body mass index	0.012[b]	0.001	0.013[b]	0.001	0.495		
Age	−0.005[b]	0.001	−0.010[b]	0.001	0.002		
Black	−0.030	0.019	−0.023	0.019	0.790		
Hispanic	0.015	0.026	−0.062[b]	0.021	0.025		
Other minority	−0.081[b]	0.039	−0.041	0.040	0.468		
			Protein				
Intercept	4.293[b]	0.013	4.263[b]	0.017	0.140	0.227	0.421
Food frequency questionnaire	0.215[b]	0.03	0.135[b]	0.029[b]	0.056		
Body mass index	0.011[b]	0.002	0.012[b]	0.002[b]	0.792		
Age	−0.010[b]	0.002	−0.012[b]	0.002[b]	0.635		
Black	−0.136[b]	0.039	−0.120[b]	0.038[b]	0.779		
Hispanic	−0.008	0.054	−0.078	0.04	0.300		
Other minority	−0.096	0.084	−0.018	0.076	0.494		
			Protein density				
Intercept	2.658[b]	0.013	2.652[b]	0.017	0.773	0.950	0.934
Food frequency questionnaire	0.409[b]	0.062	0.344[b]	0.068	0.482		
Body mass index	−0.002	0.002	−0.002	0.002	0.909		
Age	−0.005[b]	0.002	−0.002	0.002	0.359		
Black	−0.080[b]	0.04	−0.1[b]	0.037	0.717		
Hispanic	−0.042	0.052	−0.043	0.041	0.984		
Other minority	0.001	0.082	−0.03	0.078	0.785		

Abbreviations: NBS, Nutritional Biomarker Study; NPAAS, Women's Health Initiative Nutrition and Physical Activity Assessment Study; SE, standard error.

[a] *P* value 1 compares NBS and NPAAS coefficients for specific variable; *P* value 2 does so (likelihood-ratio test) for the entire set of coefficients; and *P* value 3 does so for all coefficients except the intercept. The NBS calibration equation did not depend significantly on the Dietary Modification trial randomization assignment.

[b] Coefficient significant at the $P = 0.05$ level.

significant ($P = 0.003$) for energy, but the differences derive from coefficients for age and for Hispanic ethnicity, rather than from the food frequency questionnaire coefficient. The correlations between consumption estimates using the Nutrient Biomarker Study and NPAAS calibration equations are 0.95 for energy, 0.96 for protein, and 0.96 for protein density.

Figure 2 provides scatterplots and correlation coefficients between NPAAS visit 1 and NPAAS visit 3, for women in the reliability subsample for log(biomarker) and each self-report. Food frequency questionnaire correlations are somewhat larger than those from the other self-reports, while the correlation for the protein density biomarker is low ($r = 0.24$).

The WHI food frequency questionnaire aims to assess consumption over the preceding 3 months, whereas the 4-day food record and 24-hour dietary recalls target consumption over a few days or weeks, respectively, in proximity to biomarker assessment. Calibration equations of the type shown in Table 4 were also carried out from reliability subsample data by averaging the visit 1 and visit 3 log(biomarker) assessments and using either the visit 3 log(food frequency questionnaire) or the average of visit 1 and visit 3 log(4-day food record) or log(24-hour dietary recall) assessments as predictor variables. These analyses led to somewhat higher percentages of biomarker variation explained, compared with Table 4. Specifically, for the food frequency questionnaire, 4-day food record, and 24-hour dietary recall, these percentages were, respectively, 52.3%, 58.1%, and 53.6% for energy; 24.8%, 42.6%, and 37.4% for protein; and 15.0%, 22.4%, and 20.0% for protein density. The percentages of variation explained by the food frequency

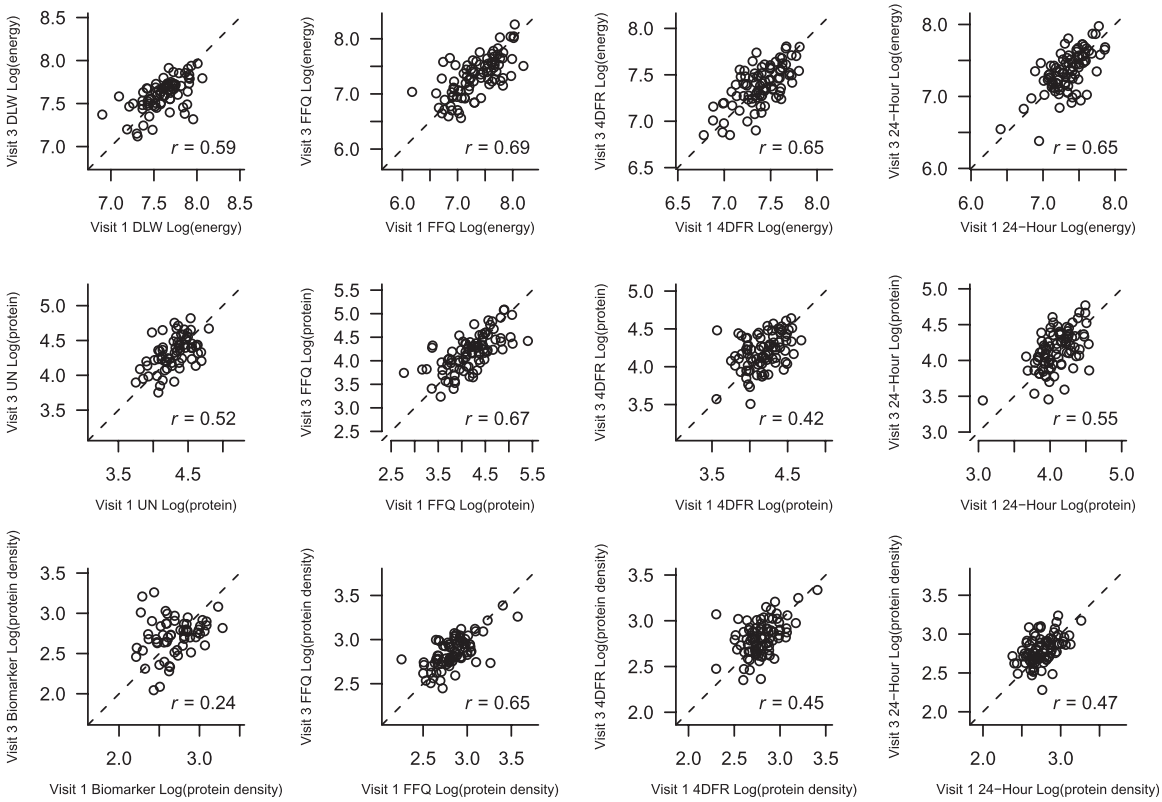

Figure 2. Scatterplot of the Women's Health Initiative Nutrition and Physical Activity Assessment Study (NPAAS; 2007–2009) primary versus reliability sample. Each plot provides the Pearson correlation for the log-measure. DLW, doubly labeled water; FFQ, food frequency questionnaire; UN, urinary nitrogen; 4DFR, 4-day food record.

questionnaire, 4-day food record, and 24-hour dietary recall data alone in these calibration equations were, respectively, 6.7%, 11.9%, and 4.3% for energy; 7.4%, 28.2%, and 18.1% for protein, and 4.9%, 12.3%, and 8.6% for protein density.

Calibration equations were also developed separately by race/ethnicity (white, black, Hispanic) and body mass index (<25.0, 25.0–29.9, ≥30.0). The ''signals'' from the self-report assessment were comparatively weaker for black women for each assessment procedure. Similarly, the signals for overweight and obese women were weaker than those for normal weight women for each assessment procedure. As shown in the Web Appendix, which is posted on the *Journal*'s Web site (http://aje.oupjournals.org/), the fraction of biomarker variation explained by these calibration equations was somewhat higher for Hispanic compared with black women, with white women intermediate; and somewhat higher for obese compared with normal weight women for energy, but higher for normal weight versus obese women for protein density, with overweight women intermediate.

DISCUSSION

Four-day food records and, to a lesser extent, 24-hour dietary recalls ''recover'' more of the variation in short-term energy and protein consumption biomarkers than does the food frequency questionnaire in our study population, providing a possible explanation for differential association study findings between food records and food frequency questionnaires (4–6). However, when combined with readily available data on body mass index, age, and ethnicity, much larger fractions of biomarker variation can be explained: about 40%–45% for energy; 20%–35% for protein; and 8%–16% for protein density. Furthermore, when these R^2 values are adjusted (Table 4) to eliminate the ''noise'' component of biomarker variation, the calibration equations provide an explanation for 70%–80% of the consumption variation for energy, 40%–68% for protein, and 52%–93% for protein density.

These adjusted R^2 values suggest that calibrated estimates using any of the 3 assessment procedures may be sufficient for epidemiologic association study purposes. The adjusted R^2 values are noticeably higher for consumption estimates using the 4-day food record versus those using the other assessment procedures. However, these adjusted R^2 values may be somewhat optimistic for the 4-day food records, in that the 4-day food record recording times corresponded closely to the biomarker assessment time period, whereas the food frequency questionnaire targeted a preceding 3-month period, and the three 24-hour dietary recalls were obtained over a 2–3-month period following biomarker assessments. R^2 values were somewhat larger and more

similar among assessment procedures, when based on the repeat biomarker and dietary data in the reliability subsample. The adjusted R^2 values using any of the assessment procedures could also be somewhat inflated by seasonal consumption variations that would tend to reduce initial and repeat log
(biomarker) correlations in the reliability subsample.

Our study examined the calibrated consumption transferability issue under near-optimal conditions of cohorts drawn from the same catchment population in the same period of time, but with different eligibility criteria and study demands. Although some minor difference could be detected in the energy calibration equation from the 2 studies, resulting consumption estimates were very highly correlated for each of the dietary variables, and the equations developed for 1 cohort can be readily applied for consumption estimation in the other.

The calibrated consumption estimates can be used rather directly in disease association studies in WHI and, potentially, other cohorts of postmenopausal women, assuming that the variables used in calibration (body mass index, age, ethnicity) are also included in the disease risk model, although nonstandard variance estimates are needed to acknowledge uncertainty in calibration equation coefficients (13, 14). Some important analyses, however, will need to allow for the possibility that body mass index change is a key variable in mediating any diet and disease association. For this purpose, analyses that exclude body mass index from the disease risk model can be induced from these that include body mass index but require reliability subsample data sufficient to estimate biomarker measurement error correlations relative to dietary consumption over the perhaps lengthy time period that may be relevant to disease risk (15).

Positive measurement error correlations among the 3 assessment procedures were estimated for each of the dietary factors, strongly arguing that biomarkers are needed for measurement error correction. The fact that biomarkers adhering to a classical measurement model have been developed for only a few dietary components precludes a comprehensive application of the biomarker approach to nutritional epidemiology. The future research agenda needs to place priority on biomarker development for additional dietary factors.

In summary, a simple calibration procedure involving dietary self-report, body mass index, age, and ethnicity appears able to estimate short- to intermediate-term dietary consumption of energy, protein, and protein density among postmenopausal US women with adequate reliability for most epidemiologic study purposes, regardless of which of the 3 dietary assessment procedures is utilized.

ACKNOWLEDGMENTS

Author affiliations: Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington (Ross L. Prentice, Ying Huang, Shirley A. A. Beresford, Lesley Tinker, Marian L. Neuhouser); Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, New York (Yasmin Mossavar-Rahmani); Department of Preventive Medicine, Northwestern University, Chicago, Illinois (Linda Van Horn); Division of Research, Kaiser Permanente, Oakland, California (Bette Caan); Department of Nutritional Sciences, University of Wisconsin, Madison, Wisconsin (Dale Schoeller); Medical Research Council, Dunn Human Nutrition Unit, University of Cambridge, Cambridge, United Kingdom (Sheila Bingham); Center for Primary Care Prevention, Memorial Hospital of Rhode Island, Alpert Medical School of Brown University, Pawtucket, Rhode Island (Charles B. Eaton); Department of Nutritional Sciences, University of Arizona, Tucson, Arizona (Cynthia Thomson); Department of Preventive Medicine, University of Tennessee Health Sciences Center, Memphis, Tennessee (Karen C. Johnson); Department of Medicine, University of Massachusetts Medical School, Worcester, Massachusetts (Judy Ockene); Department of Obstetrics and Gynecology, University of Wisconsin, Madison, Wisconsin (Gloria Sarto); and Departments of Epidemiology and Medicine, University of North Carolina, Chapel Hill, North Carolina (Gerardo Heiss).

This work was supported by the National Heart, Lung, and Blood Institute, National Institutes of Health, Department of Health and Human Services (contracts N01WH22110, 24152, 32100-2, 32105-6, 32108-9, 32111-13, 32115, 32118-19, 32122, 42107-26, 42129-32, and 44221); National Cancer Institute (grants CA119171 and CA53996); and National Center for Research Resources (grant 5UL1RR025750). Clinical Trials Registration: ClinicalTrials.gov identifier: NCT00000611.

The authors thank the WHI investigators and staff for their outstanding dedication and commitment.

A full listing of WHI investigators can be found at the following website: http://www.whi.org. A list of key investigators involved in this research follows. *Program Office:* Jacques Rossouw, Shari Ludlam, Joan McGowan, Leslie Ford, and Nancy Geller (National Heart, Lung, and Blood Institute, Bethesda, Maryland). *Clinical Coordinating Center:* Ross Prentice, Garnet Anderson, Andrea LaCroix, and Charles Kooperberg (Fred Hutchinson Cancer Research Center, Seattle, Washington); Evan Stein (Medical Research Labs, Highland Heights, Kentucky); Steven Cummings (University of California at San Francisco, San Francisco, California). *Investigators and Academic Centers:* Sylvia Wassertheil-Smoller (Albert Einstein College of Medicine, Bronx, New York); Haleh Sangi-Haghpeykar (Baylor College of Medicine, Houston, Texas); JoAnn E. Manson (Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts); Charles B. Eaton (Brown University, Providence, Rhode Island); Lawrence S. Phillips (Emory University, Atlanta, Georgia); Shirley Beresford (Fred Hutchinson Cancer Research Center, Seattle, Washington); Lisa Martin (George Washington University Medical Center, Washington, DC); Rowan Chlebowski (Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, California); Erin LeBlanc (Kaiser Permanente Center for Health Research, Portland, Oregon); Bette Caan (Kaiser Permanente Division of Research, Oakland, California); Jane Morley Kotchen (Medical College of Wisconsin, Milwaukee, Wisconsin); Barbara V. Howard (MedStar Health Research Institute/Howard University,

Washington, DC); Linda Van Horn (Northwestern University, Chicago/Evanston, Illinois); Henry Black (Rush Medical Center, Chicago, Illinois); Marcia L. Stefanick (Stanford Prevention Research Center, Stanford, California); Dorothy Lane (State University of New York at Stony Brook, Stony Brook, New York); Rebecca Jackson (The Ohio State University, Columbus, Ohio); Cora E. Lewis (University of Alabama at Birmingham, Birmingham, Alabama); Cynthia A. Thomson (University of Arizona, Tucson/Phoenix, Arizona); Jean Wactawski-Wende (University at Buffalo, Buffalo, New York); John Robbins (University of California at Davis, Sacramento, California); Hoda Anton-Culver (University of California at Irvine, Irvine, California); Lauren Nathan (University of California at Los Angeles, Los Angeles, California); Robert D. Langer (University of California at San Diego, LaJolla/Chula Vista, California); Margery Gass (University of Cincinnati, Cincinnati, Ohio); Marian Limacher (University of Florida, Gainesville/Jacksonville, Florida); J. David Curb (University of Hawaii, Honolulu, Hawaii); Robert Wallace (University of Iowa, Iowa City/Davenport, Iowa); Judith Ockene (University of Massachusetts/Fallon Clinic, Worcester, Massachusetts); Norman Lasser (University of Medicine and Dentistry of New Jersey, Newark, New Jersey); Mary Jo O'Sullivan (University of Miami, Miami, Florida); Karen Margolis (University of Minnesota, Minneapolis, Minnesota); Robert Brunner (University of Nevada, Reno, Nevada); Gerardo Heiss (University of North Carolina, Chapel Hill, North Carolina); Lewis Kuller (University of Pittsburgh, Pittsburgh, Pennsylvania); Karen C. Johnson (University of Tennessee Health Science Center, Memphis, Tennessee); Robert Brzyski (University of Texas Health Science Center, San Antonio, Texas); Gloria E. Sarto (University of Wisconsin, Madison, Wisconsin); Mara Vitolins (Wake Forest University School of Medicine, Winston-Salem, North Carolina); and Michael S. Simon (Wayne State University School of Medicine/Hutzel Hospital, Detroit, Michigan). *Women's Health Initiative Memory Study:* Sally Shumaker (Wake Forest University School of Medicine, Winston-Salem, North Carolina).

Although decisions concerning study design, data collection and analysis, interpretation of the results, preparation of the manuscript, and whether to submit the manuscript for publication resided with committees composed of WHI investigators and National Heart, Lung, and Blood Institute representatives, the contents of the paper are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.

Conflict of interest: none declared.

REFERENCES

1. *Diet, Nutrition, and the Prevention of Chronic Diseases: Report of a Joint WHO/FAO Expert Consultation*. Geneva, Switzerland: World Health Organization; 2003: 88. (WHO technical report 916).
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; 1997:371.
3. Schatzkin A, Subar AF, Moore S, et al. Observational epidemiologic studies of nutrition and cancer: the next generation (with better observation). *Cancer Epidemiol Biomarkers Prev.* 2009;18(4):1026–1032.
4. Bingham SA, Luben R, Welch A, et al. Are imprecise methods obscuring a relation between fat and breast cancer? *Lancet*. 2003;362(9379):212–214.
5. Freedman LS, Potischman N, Kipnis V, et al. A comparison of two dietary instruments for evaluating the fat-breast cancer relationship. *Int J Epidemiol*. 2006;35(4):1011–1021.
6. Dahm CC, Keogh RH, Spencer EA, et al. Dietary fiber and colorectal cancer risk: a nested case-control study using food diaries. *J Natl Cancer Inst*. 2010;102(9):614–626.
7. Kaaks RJ. Biochemical markers as additional measurements in studies of the accuracy of dietary questionnaire measurements: conceptual issues. *Am J Clin Nutr.* 1997;65(suppl 4): S1232–S1239.
8. Subar AF, Kipnis V, Troiano RP, et al. Using intake biomarkers to evaluate the extent of dietary misreporting in a large sample of adults: the OPEN Study. *Am J Epidemiol*. 2003;158(1):1–13.
9. Kipnis V, Subar AF, Midthune D, et al. Structure of dietary measurement error: results of the OPEN biomarker study. *Am J Epidemiol*. 2003;158(1):14–21.
10. Day N, McKeown N, Wong M, et al. Epidemiological assessment of diet: a comparison of a 7-day diary with a food frequency questionnaire using urinary markers of nitrogen, potassium and sodium. *Int J Epidemiol*. 2001;30(2): 309–317.
11. Willett W. Commentary: dietary diaries versus food frequency questionnaires—a case of undigestible data. *Int J Epidemiol*. 2001;30(2):317–319.
12. Neuhouser ML, Tinker L, Shaw PA, et al. Use of recovery biomarkers to calibrate nutrient consumption self-reports in the Women's Health Initiative. *Am J Epidemiol*. 2008;167(10): 1247–1259.
13. Prentice RL, Shaw PA, Bingham SA, et al. Biomarker-calibrated energy and protein consumption and increased cancer risk among postmenopausal women. *Am J Epidemiol*. 2009;169(8):977–989.
14. Prentice RL, Huang Y, Kuller LH, et al. Biomarker-calibrated energy and protein consumption and cardiovascular disease risk among postmenopausal women. *Epidemiology.* 2011; 22(2):170–179.
15. Prentice RL, Huang Y. Measurement error modeling and nutritional epidemiology association analyses. *Can J Stat*. In press.
16. Women's Health Initiative Study Group. Design of the Women's Health Initiative Clinical Trial and Observational Study. *Control Clin Trials*. 1998;19(1):61–109.
17. Langer RD, White E, Lewis CE, et al. The Women's Health Initiative Observational Study: baseline characteristics of participants and reliability of baseline measures. *Ann Epidemiol*. 2003;13(suppl 9):S107–S121.
18. Kristal AR, Shattuck AL, Williams AE. Food frequency questionnaires for diet intervention research. *17th National Nutrient Databank Conference: Celebrating the First 100 Years of Food Composition Data 1982—1992*. Baltimore, MD: International Life Sciences Institute; 1992:110–125.
19. Schoeller DA, Hnilicka JM. Reliability of the doubly labeled water method for the measurement of total daily energy expenditure in free-living subjects. *J Nutr.* 1996;126(suppl): S348–S354.
20. Schoeller DA. Recent advances from application of doubly labeled water to measurement of human energy expenditure. *J Nutr.* 1999;129(10):1765–1768.

21. Bingham SA. The use of 24-h urine samples and energy expenditure to validate dietary assessments. *Am J Clin Nutr.* 1994;59(suppl 1):S227–S231.
22. Bingham SA, Murphy J, Waller E, et al. Para-amino benzoic acid in the assessment of completeness of 24-hour urine collections from hospital outpatients and the effect of impaired renal function. *Eur J Clin Nutr.* 1992;46(2):131–135.
23. Patterson RE, Kristal AR, Tinker LF, et al. Measurement characteristics of the Women's Health Initiative food frequency questionnaire. *Ann Epidemiol*. 1999;9(3):178–187.
24. Schakel SF, Buzzard IM, Gebhardt SE. Procedures for estimating nutrient values for food composition databases. *J Food Compost Anal*. 1997;10(2):102–114.
25. Prentice RL, Sugar E, Wang CY, et al. Research strategies and the use of nutrient biomarkers in studies of diet and chronic disease. *Public Health Nutr.* 2002;5(6A): 977–984.
26. Sugar EA, Wang CY, Prentice RL. Logistic regression with exposure biomarkers and flexible measurement error. *Biometrics*. 2007;63(1):143–151.