American Journal of Epidemiology

Vol. 167, No. 10
DOI: 10.1093/aje/kwn026
Advance Access publication March 15, 2008

Practice of Epidemiology

Use of Recovery Biomarkers to Calibrate Nutrient Consumption Self-Reports in the Women's Health Initiative

Marian L. Neuhouser[1], Lesley Tinker[1], Pamela A. Shaw[2], Dale Schoeller[3], Sheila A. Bingham[4], Linda Van Horn[5], Shirley A. A. Beresford[1,6], Bette Caan[7], Cynthia Thomson[8,9], Suzanne Satterfield[10], Lew Kuller[11], Gerardo Heiss[12], Ellen Smit[13], Gloria Sarto[14], Judith Ockene[15], Marcia L. Stefanick[16], Annlouise Assaf[17], Shirley Runswick[4], and Ross L. Prentice[1]

[1] Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA.
[2] Biostatistics Research Branch, National Institute of Allergy and Infectious Diseases, Bethesda, MD.
[3] Department of Nutritional Sciences, University of Wisconsin, Madison, WI.
[4] Medical Research Council Dunn Human Nutrition Unit, University of Cambridge, Cambridge, United Kingdom.
[5] Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL.
[6] Department of Epidemiology, School of Public Health and Community Medicine, University of Washington, Seattle, WA.
[7] Division of Research, Kaiser Permanente, Oakland, CA.
[8] Department of Nutritional Sciences, College of Agriculture and Life Sciences, University of Arizona, Tucson, AZ.
[9] Division of Prevention and Control, Arizona Cancer Center, Tucson, AZ.
[10] Department of Preventive Medicine, College of Medicine, University of Tennessee Health Science Center, Memphis, TN.
[11] Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA.
[12] Department of Epidemiology, School of Public Health, University of North Carolina, Chapel Hill, NC.
[13] School of Community Health, College of Urban and Public Affairs, Portland State University, Portland, OR.
[14] Department of Obstetrics and Gynecology, Medical School, University of Wisconsin–Madison, Madison, WI.
[15] Department of Medicine, University of Massachusetts Medical Center, Worcester, MA.
[16] Stanford Prevention Research Center, Stanford University, Palo Alto, CA.
[17] Pfizer, Inc., New York, NY.

Received for publication July 10, 2007; accepted for publication January 24, 2008.

Underreporting of energy consumption by self-report is well-recognized, but previous studies using recovery biomarkers have not been sufficiently large to establish whether participant characteristics predict misreporting. In 2004–2005, 544 participants in the Women's Health Initiative Dietary Modification Trial completed a doubly labeled water protocol (energy biomarker), 24-hour urine collection (protein biomarker), and self-reports of diet (assessed by food frequency questionnaire (FFQ)), exercise, and lifestyle habits; 111 women repeated all procedures after 6 months. Using linear regression, the authors estimated associations of participant characteristics with misreporting, defined as the extent to which the log ratio (self-reported FFQ/nutritional biomarker) was less than zero. Intervention women in the trial underreported energy intake by 32% (vs. 27% in the comparison arm) and protein intake by 15% (vs. 10%). Younger women had more underreporting of energy ($p = 0.02$) and protein ($p = 0.001$), while increasing body mass index predicted increased underreporting of energy and overreporting of percentage of energy derived from protein ($p = 0.001$ and $p = 0.004$, respectively). Blacks and Hispanics underreported more than did Caucasians. Correlations of initial measures with repeat measures ($n = 111$) were 0.72, 0.70, 0.46, and 0.64 for biomarker energy, FFQ energy, biomarker protein, and FFQ protein, respectively. Recovery biomarker data were used in regression equations to calibrate self-reports; the potential application of these equations to disease risk modeling is presented. The authors confirm the existence of systematic bias in dietary self-reports and provide methods of correcting for measurement error.

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

Correspondence to Dr. Marian L. Neuhouser, Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, M4B402, Seattle, WA 98109-1024 (e-mail: mneuhous@fhcrc.org).

Abbreviations: DM, Dietary Modification [Trial]; FFQ, food frequency questionnaire; NBS, Nutritional Biomarkers Study; PABA, para-aminobenzoic acid; TEE, total energy expenditure; WHI, Women's Health Initiative.

Dietary factors play an important role in determining the risk of chronic diseases such as cardiovascular disease and cancer (1–3). However, the magnitude and certainty of diet-disease associations from nutritional epidemiology studies are often lower than one might expect, given the strong biologic evidence motivating such hypotheses. Further, differences among studies in the direction of association between dietary exposures and disease endpoints are common. One reason for such discrepancies is measurement error associated with self-reported dietary assessment, which leads to serious attenuation or other distortion of relative risk estimates (4–7). Distorted risk estimates and the subsequent lack of consistent findings across studies generate controversy within the field of nutritional epidemiology (8, 9) and impede progress towards the formulation of diet-related strategies for risk reduction (2, 4, 10). New analytic approaches for dietary assessment are needed to bridge the gap between important biologic concepts regarding diet-disease associations and the many modest or null relative risk estimates obtained from observational studies (4, 6, 10–12).

Among the important issues with respect to investigating dietary measurement error are the following. Firstly, caution should be applied to previous assumptions about the measurement properties of dietary assessment procedures. Evidence suggests that errors in food frequency questionnaires (FFQs) are correlated with errors in typical ''reference'' instruments, such as food records or 24-hour recalls (7). This leads some epidemiologists to question the interpretation of ''validity'' studies and renders these types of reference instruments unsuited for the calibration of data arising from the standard instrument (i.e., the FFQ). Instead, correlated biases between assessment instruments relying upon self-reports necessitate the use of objective biomarker measures of diet for calibration, since biomarker measurement errors are likely to be independent of the errors associated with self-reported estimates (13, 14). Few studies have utilized objective measures of diet to describe the measurement properties (including the error variance) of self-reported nutrient consumption data; however, biomarkers can be effectively used in statistical approaches, such as regression calibration, to approximate actual diet-disease associations (10, 11). These approaches will be valuable for large studies that cannot collect biomarker data on the entire study sample, particularly when appropriately calibrated estimates depend on participant characteristics. Secondly, the self-reported diet errors probably include both random and systematic errors. Scant attention has been focused on sources of systematic bias, but evidence suggests that some subgroups, such as obese persons, may underreport their energy consumption (15–19), perhaps in a manner that depends on the self-report assessment procedure and the study population.

We aimed to assess these methodological issues in the dietary modification trial of the Women's Health Initiative (WHI) (20, 21). Our objectives were to: 1) use biomarkers to characterize the measurement error distributions for FFQ-assessed energy and protein; 2) examine whether the measurement error structure varied by participant characteristics such as age, race/ethnicity, or obesity; and 3) develop equations to calibrate FFQ nutrient consumption estimates for use in subsequent WHI disease association studies.

MATERIALS AND METHODS

The WHI Dietary Modification Trial

The WHI Dietary Modification Trial (WHI-DM) was a randomized, controlled dietary intervention trial testing whether a low-fat dietary pattern would reduce the incidence of invasive breast cancer and colorectal cancer (primary endpoints) and coronary heart disease (secondary endpoint) among 48,835 postmenopausal US women. Details about the design of the WHI-DM have been published elsewhere (20, 21). The WHI-DM began in 1993. After a mean follow-up period of 8.1 years, the trial ended in March 2005. Results have been published for breast cancer (22), colorectal cancer (23), and coronary heart disease (24).

The WHI Nutritional Biomarkers Study

The WHI Nutritional Biomarkers Study (WHI-NBS) was undertaken to determine the measurement properties of the WHI FFQ, which was designed specifically for WHI as the primary assessment instrument and principal monitoring tool for the WHI-DM (22–25). The WHI-NBS also aimed to facilitate disease risk modeling using biomarker-calibrated estimates of diet. The collection and use of objective nutritional biomarkers, particularly recovery biomarkers, can provide important information about the measurement error in FFQ self-reports (5, 13, 26, 27). Recovery biomarkers have a known quantitative time-associated relation between dietary intake and recovery (excretion) in human waste (6, 13).

Participants/recruitment

Participants in the WHI-NBS were 544 postmenopausal women enrolled in the WHI-DM. We sought to obtain a sample that was representative of the WHI-DM in terms of age, race/ethnicity, body mass index (weight (kg)/height $(m)^2$), and randomization assignment (50 percent intervention group, 50 percent comparison group). We sent letters of invitation to potential participants, followed by telephone screening. Women were excluded for having any medical conditions precluding participation, weight instability, and planning to travel during the study period. Of the 1,456 WHI-DM participants invited, 677 (46.5 percent) declined, 223 (15.3 percent) were ineligible, and 556 (38.2 percent) agreed to participate. Twelve women (2 percent) dropped out, leaving 544 women who completed their study participation

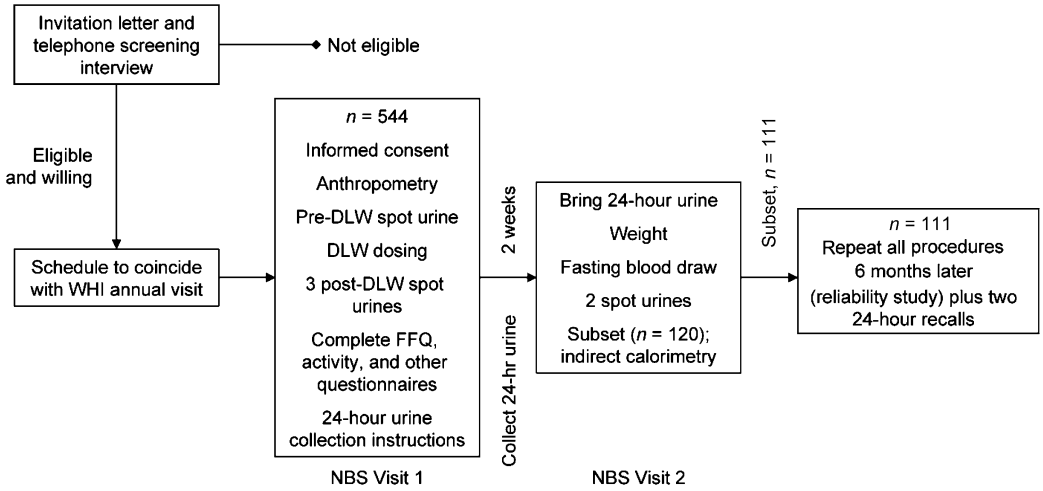

FIGURE 1. Procedures used in the Women's Health Initiative (WHI) Nutritional Biomarkers Study (NBS), 2004–2005. Of the 1,456 participants in the WHI Dietary Modification Trial who were invited to participate in the WHI-NBS, 677 (46.5%) declined, 223 (15.3%) were ineligible, and 556 (38.2%) agreed to participate. Twelve women (2%) dropped out, leaving a total sample size of 544. DLW, doubly labeled water; FFQ, food frequency questionnaire.

between May 2004 and March 2005. Study procedures were approved by the institutional review boards of the WHI Clinical Coordinating Center (Fred Hutchinson Cancer Research Center, Seattle, Washington) and 12 geographically dispersed WHI clinical centers. All participants gave written informed consent. Participants received $100 upon study completion.

Figure 1 illustrates the study protocol. Briefly, activities included two visits to local WHI clinics where participants completed a doubly labeled water protocol (recovery biomarker of total energy expenditure (TEE)), had a 24-hour urine sample collected (recovery biomarker of protein intake), gave a fasting blood specimen, and completed self-report instruments on diet, physical activity, and lifestyle habits. Height and weight were measured. A total of 111 women (20 percent of enrollees) repeated the protocol 6 months later as part of a reliability study. We recruited the reliability study subset from the early enrollees in order to ensure this 6-month window.

Doubly labeled water biomarker of TEE

TEE is optimally assessed using doubly labeled water, which measures energy expenditure over a 2-week period (27, 28). In weight-stable persons, TEE is approximately equivalent to energy intake, such that if TEE is measured with precision, it provides an objective estimate of energy intake. Briefly, after a loading dose of water labeled with deuterium plus the stable isotope oxygen-18, the tracers rapidly equilibrate in body water. The deuterium is eliminated from the body as water, and the elimination rate is proportional to water turnover. The oxygen-18 is eliminated as water plus carbon dioxide, and the oxygen-18 elimination is proportional to the sum of water and carbon dioxide production. The difference between these two elimination rates is proportional to the production of carbon dioxide, which is the end product of energy metabolism from which TEE is estimated (27).

Participants arrived for study visit 1 after a 4-hour fast, provided a baseline urine specimen, and were weighed. Participants ingested the doubly labeled water in a single dose of approximately 1.8 g of 10 atom percent oxygen-18-labeled water and 0.12 g of 99.9 percent deuterium-labeled water per kilogram of estimated total body water (29). Participants remained in the clinic for 4 hours and provided three additional spot urine specimens (5, 27). Participants received a meal replacement beverage and additional fluids as necessary for urine production. Women aged $\geq$60 years provided a blood sample 3 hours post-isotope to compensate for age-related postvoid urine retention (30). For participants in whom the 3-hour urine sample showed insufficient isotope enrichment, the isotope enrichment values from the 3-hour blood drawing were used in lieu of the urine values. A protocol and validated methods for use of the blood values instead of urine values for the 3-hour time point have been published (30). Women returned on day 15 and gave a fasting blood sample for future analyses of concentration biomarkers, were weighed, and provided two additional spot urine samples.

Isotopes in the biospecimens were measured by mass spectrometry at the University of Wisconsin, Madison, Stable Isotope Laboratory. Blinded duplicates (5 percent) were included for quality control. TEE was calculated from carbon dioxide production using the modified Weir equation, which calculates TEE as a function of oxygen consumption and carbon dioxide release. This calculation assumes a respiratory quotient of 0.86 based on a high-fat Western diet (31, 32). Since the goals of the WHI-DM intervention included reducing fat intake to 20 percent of energy intake, we calculated TEE both from a respiratory quotient based on 34 percent

energy derived from fat (the standard American diet) (31, 32) and from values estimated towards the end of the WHI-DM intervention (29.8 percent energy from fat for the intervention arm and 38.1 percent for the comparison arm) (33).

Urinary nitrogen biomarker of protein

Urinary nitrogen is a recovery biomarker for protein: protein intake (g/day) = 6.25 × (24-hour urinary nitrogen/0.81) (34, 35). Participants had urine collected for 24 hours on day 14, immediately preceding visit 2, with details about missed or spilled voids being noted. Urine samples were not analyzed if two or more voids were missed or if ≥240 ml of urine was spilled ($n = 6$). PABACheck (para-aminobenzoic acid; Laboratories for Applied Biology Ltd., London, United Kingdom) (34) was used to assess the quality of urine collection (14, 36, 37). PABACheck was unavailable from the supplier during the early months of WHI-NBS, so only 83 of 544 participants in the primary study but all participants in the repeat application of the protocol took one 80-mg PABACheck pill with each meal on the day of urine collection (total $n = 194$) (34). The Dunn Human Nutrition Unit at the University of Cambridge (Cambridge, United Kingdom) analyzed specimens by means of the Kjeldahl technique (Tecator 1015 digestor and Kjeltec 1035 analyzer; Foss UK Ltd., Warrington, United Kingdom) (34). Blinded duplicates (5 percent) were included for quality control. The recovery of PABA from urine was determined colorimetrically (37); recovery of 85–110 percent of the dose equaled complete urine collection. Specimens with recovery of greater than 110 percent were reanalyzed by high-performance liquid chromatography (5, 36, 37).

Dietary assessment

Participants completed the WHI FFQ (25). This self-administered questionnaire includes 19 adjustment questions, 122 line items for individual foods/food groups, and summary questions. The nutrient database used is the Nutrition Data System for Research (version 2005; University of Minnesota, Minneapolis, Minnesota). Nutritional epidemiologists matched the FFQ foods to appropriate database selections. Our approach to analyzing FFQs and the algorithms used for analysis have been described elsewhere (38, 39).

Other measures

Participants completed standard, self-reported assessment instruments on physical activity, alcohol consumption, and smoking and an interviewer-administered dietary supplement inventory (20, 21, 40). Data on demographic characteristics had been collected previously in WHI and were not reassessed in WHI-NBS.

Statistical analyses

Our analytic goals included developing measurement-error models for the FFQ values for energy and protein using the biomarkers, estimating the reliability of self-report and biomarker measures in a subsample, and constructing regression equations to calibrate (i.e., correct) FFQ nutrient estimates for selected nutrients. For the measurement error analysis, we applied logarithmic transformation to data on energy, protein, and percentage of energy derived from protein. The calibration equations rely on the assumption that the (log-transformed) biomarker adheres to a classical measurement model, where the errors in the biomarker are independent of nutrient intake and participant characteristics. They also assume that measurement errors in self-reports are independent of errors in objective biomarkers; thus, the biomarker functions as an anchor with which to calibrate the self-report (11, 41). Details about these statistical methods are given in the Appendix, and simulation studies supporting this approach have been published elsewhere (11, 41).

To study the relation between dietary measurement error and participant characteristics, we analyzed the differences between log-transformed FFQ values for energy, protein, and percentage of energy derived from protein and corresponding recovery biomarkers (11). In linear regression models, we estimated the association of the difference between the log-transformed values for energy and protein with these participant characteristics (hypothesized a priori): WHI-DM randomization assignment, age, body mass index, race/ethnicity, education, income, smoking, use of dietary supplements, and physical activity. Outliers that fell outside the interquartile range by more than three times the width of the interquartile range were excluded (FFQ: $n = 4$; biomarkers: $n = 3$). Models were fitted using linear regression with "robust" variance estimates, using the generalized estimating equations package in R software (version 2.2.1; R Foundation for Statistical Computing, Vienna, Austria (http://www.r-project.org)). The robust variance (Huber/White or sandwich estimates) provides a consistent estimate for the variance of the regression coefficients, allowing for dependence between repeated measures and potential heteroscedasticity, and it does not require specific distributional assumptions for the regression residuals (42). Statistical significance was determined by means of Wald tests with robust standard errors ($\alpha = 0.05$).

The second analytic goal was to estimate the reliability of the self-reports and biomarkers. Unadjusted correlations were used to assess agreement between the repeat measures of the (log-transformed) biomarkers and the FFQ.

Finally, to create regression calibration equations for use in subsequent disease risk association studies, we used multivariate regression models to predict true intakes of energy and protein, given the observed self-report and biomarker data (see Appendix). We began with a full list of covariates and applied a backward selection procedure ($p = 0.10$) to eliminate covariates. An interaction term for the FFQ variable and body mass index was also considered for each calibration model. The retained covariates for this reduced measurement error model were used to create regression calibration equations for a single FFQ measure for energy, protein, and percentage of energy derived from protein (42). In addition, we conducted a sensitivity analysis for the protein measures by comparing data with and without PABA adjustment.

TABLE 1. Baseline demographic and lifestyle characteristics of participants in the Women's Health Initiative Nutritional Biomarkers Study (WHI-NBS) and participants in the Women's Health Initiative Dietary Modification Trial (WHI-DM), 2004–2005

Characteristic	WHI-NBS sample ($n = 544$)		WHI-DM sample* ($n = 48,291$)		p value†
	No.	%‡	No.	%‡	
Age (years)					0.02
50–59	212	39.0	17,791	36.8	
60–69	265	48.7	22,447	46.5	
70–79	67	12.3	8,053	16.7	
Body mass index§					0.005
Normal (<25.0)	165	30.3	12,492	25.9	
Overweight (25.0–29.9)	203	37.3	17,183	35.6	
Obese (≥30)	172	31.6	18,396	38.1	
Race					0.16
White	448	82.4	39,311	81.4	
Black	59	10.8	5,207	10.8	
Hispanic	26	4.8	1,828	3.8	
Other¶	11	2.0	1,945	4.0	
Annual income					0.05
<$20,000	60	11.0	7,017	14.5	
$20,000–$34,999	133	24.4	11,182	23.2	
$35,000–$49,999	123	22.6	9,699	20.1	
$50,000–$74,999	107	19.7	9,442	19.6	
≥$75,000	94	17.3	8,147	16.9	
Education					0.16
Less than high school	20	3.7	2,195	4.5	
High school/General Educational Development diploma	82	15.1	8,436	17.5	
Schooling after high school	207	38.1	19,101	39.6	
College degree or higher	233	42.8	18,254	37.8	
Current smoking	35	6.4	3,215	6.7	0.89
Any use of dietary supplements	367	67.5	31,202	64.6	0.18
Recreational physical activity (no. of times per week)					0.47
<2	192	35.3	17,297	35.8	
2–3	119	21.9	9,558	19.8	
≥4	233	42.8	21,436	44.4	

* The WHI-DM sample characteristics shown include all participants in the dietary modification trial who did not participate in the biomarker substudy.

† From Pearson's χ^2 test of independence for WHI-NBS membership and WHI-DM cohort ($n = 48,835$) baseline characteristics.

‡ Percentages in individual categories may not sum to 100 because of missing values.

§ Weight (kg)/height $(m)^2$.

¶ For the WHI-NBS sample, "other" race included American Indian ($n = 2$), Asian/Pacific Islander ($n = 4$), and unspecified ($n = 5$).

RESULTS

Table 1 presents demographic and lifestyle characteristics for both WHI-NBS participants ($n = 544$) and WHI-DM participants ($n = 48,291$). The mean age at WHI-DM enrollment was 61.7 years; women were, on average, 9 years older at WHI-NBS enrollment. Eighty-two percent of WHI-NBS participants were White, 10.8 percent were Black, and 4.8 percent were Hispanic. The mean body mass index was 28.3; 37.3 percent of participants were overweight (body mass index 25.0–29.9) and 31.6 percent were obese (body mass index ≥30.0). WHI-NBS participants were well-educated; 42.8

TABLE 2. Distribution of nutritional biomarkers and self-reported measures of energy and protein intake among postmenopausal women in the Women's Health Initiative Nutritional Biomarkers Study, 2004–2005*

Variable	WHI† Dietary Modification Trial treatment assignment					
	Intervention arm ($n = 268$)			Comparison arm ($n = 276$)		
	Geometric mean	95% CI†	IQR†	Geometric mean	95% CI	IQR
FFQ† total energy (kcal)	1,379	1,320, 1,440	1,143–1,728	1,505	1,441, 1,571	1,204–1,917
Doubly labeled water total energy						
US respiratory quotient‡ (kcal)	2,059	2,020, 2,099	1,842–2,278	2,053	2,017, 2,091	1,888–2,257
WHI respiratory quotient§ (kcal)	2,029	1,990, 2,069	1,815–2,244	2,067	2,030, 2,104	1,900–2,271
FFQ/doubly labeled water energy						
US respiratory quotient‡	0.67	0.64, 0.70	0.56–0.85	0.73	0.70, 0.77	0.58–0.94
WHI respiratory quotient§	0.68	0.65, 0.71	0.56–0.87	0.73	0.69, 0.76	0.58–0.93
FFQ protein (g)	61.3	58.5, 64.2	48.3–82.4	63.8	60.9, 67.0	51.0–84.0
24-hour urinary nitrogen (g)	9.4	9.1, 9.8	7.7–11.9	9.2	8.9, 9.5	7.7–11.5
Urinary nitrogen-based protein (g)	72.8	70.3, 75.5	59.3–91.6	71.3	68.9, 73.9	59.6–89.0
FFQ/urinary nitrogen protein	0.85	0.80, 0.89	0.65–1.12	0.90	0.86, 0.94	0.71–1.17
% energy from protein						
FFQ	17.7	17.3, 18.1	15.8–20.1	17.0	16.6, 17.4	15.1–19.3
US respiratory quotient‡	14.1	13.6, 14.6	11.9–17.1	13.9	13.4, 14.4	11.8–16.8
WHI respiratory quotient§	14.3	13.8, 14.8	12.0–17.4	13.8	13.4, 14.3	11.7–16.7
FFQ/biomarker % energy from protein						
US respiratory quotient‡	1.25	1.20, 1.30	1.02–1.51	1.22	1.18, 1.26	1.03–1.42
WHI respiratory quotient§	1.23	1.19, 1.28	1.00–1.49	1.23	1.19, 1.27	1.04–1.43

* Results are from the primary sample and include available data from all study participants.
† WHI, Women's Health Initiative; CI, confidence interval; IQR, interquartile range (25th–75th percentiles); FFQ, food frequency questionnaire.
‡ Energy expenditure calculated using the US average percentage of energy derived from fat (see Materials and Methods for details).
§ Energy expenditure calculated using the average percentage of energy derived from fat in the WHI Dietary Modification Trial (see Materials and Methods for details).

percent had at least a college degree. Few women smoked (6.4 percent), and 67.5 percent used dietary supplements. There were small but significant ($p < 0.05$) differences for age and body mass index between WHI-NBS participants and the remaining WHI-DM participants.

The distributions of data on the nutritional biomarkers and FFQ measures of energy, protein, and percentage of energy derived from protein are shown in table 2. The unadjusted geometric mean value for FFQ energy was lower for the intervention arm ($p = 0.005$) than for the comparison arm. There was no difference in FFQ protein, whereas FFQ percentage of energy from protein was higher for women in the intervention arm ($p = 0.02$). Compared with the TEE biomarker, women in the intervention arm underreported energy intake by 32 percent (geometric mean for FFQ/TEE = 0.68), while women in the comparison arm underreported energy intake by 27 percent (FFQ/TEE = 0.73) ($p = 0.03$). The geometric mean for the ratio of FFQ protein to biomarker protein was 0.85 in the intervention group and 0.90 in the comparison group ($p = 0.10$), suggesting slightly more underreporting in the intervention group, but this difference was not statistically significant. The FFQ percentage of energy derived from protein was higher than the biomarker percentage (geometric mean = 1.23 in both study arms).

Table 3 shows results from multivariate regression models giving the associations of participant characteristics with measurement error in self-reported diet. For energy, underreporting (i.e., the extent to which the log ratio (FFQ/biomarker) was less than zero) decreased with age ($p = 0.02$) and increased with body mass index ($p = 0.001$). Race/ethnicity was associated with misreporting (global $p = 0.0009$). Blacks and Hispanics had more underreporting than did Caucasians. The difference in underreporting by WHI-DM intervention arm was not significant when we used TEE based on the WHI dietary composition-calculated respiratory quotient ($p = 0.08$) but was significant when we used the US diet-based respiratory quotient ($p = 0.02$). For protein, underreporting decreased with age ($p = 0.001$). The global test for race/ethnicity was significant for protein ($p = 0.002$); Blacks ($p = 0.02$) and Hispanics ($p = 0.09$) had more protein underreporting than did Caucasians. Misreporting of percentage of energy derived from protein was significantly associated with body mass index ($p = 0.004$) and smoking ($p = 0.006$).

TABLE 3. Beta coefficients for regression of measurement error in self-reported diet on participant characteristics in the Women's Health Initiative Nutritional Biomarkers Study, 2004–2005†

	Total energy: log(FFQ‡/TEE‡,§)	Protein: log(FFQ/UN‡,¶)	% energy from protein: log(FFQ/UN)
Diet change intervention arm [comparison arm]#	−0.055 (0.032)**	−0.019 (0.034)	0.029 (0.024)
Body mass index††	−0.011 (0.003)*	−0.004 (0.003)	0.007 (0.002)*
Age (years)	0.006 (0.003)*	0.009 (0.003)*	0.004 (0.002)
Black race/ethnicity [Caucasian]	−0.140 (0.068)*	−0.165 (0.071)*	−0.027 (0.048)
Hispanic race/ethnicity [Caucasian]	−0.138 (0.068)*	−0.139 (0.083)*	−0.003 (0.050)
Other race/ethnicity [Caucasian]	0.134 (0.054)*	0.118 (0.055)*	−0.006 (0.060)
Education [some college]			
High school/General Educational Development diploma or less	−0.001 (0.045)	−0.075 (0.050)	−0.057 (0.036)
College degree or more	0.074 (0.037)	0.047 (0.038)	−0.010 (0.027)
Annual income [$35,000–$49,999]			
<$20,000	0.014 (0.064)	−0.021 (0.075)	−0.016 (0.046)
$20,000–$34,999	0.009 (0.044)	0.021 (0.048)	0.021 (0.036)
$50,000–$74,999	−0.026 (0.045)	−0.064 (0.051)	−0.010 (0.038)
≥$75,000	−0.112 (0.054)	−0.127 (0.054)	−0.026 (0.041)
Current smoking [nonsmoking]	0.043 (0.077)	0.0158 (0.092)	0.159 (0.058)*
Any use of dietary supplements [no use]	0.043 (0.052)	0.053 (0.058)	−0.004 (0.033)
Physical activity (metabolic equivalents/week)	0.0004 (0.001)	0.0002 (0.001)	0.0002 (0.0008)

* $p < 0.05$ (global χ^2 test).

† Each column represents a single model. Positive beta coefficients indicate decreased underreporting, while negative coefficients indicate increased underreporting, since underreporting is defined as the extent to which the log ratio (FFQ/biomarker) is less than zero.

‡ FFQ, food frequency questionnaire; TEE, total energy expenditure; UN, urinary nitrogen.

§ Doubly labeled water biomarker for TEE calculated with the Women's Health Initiative cohort respiratory quotient.

¶ UN protein marker.

Information in brackets, reference category. Age, body mass index, and physical activity were modeled as continuous variables.

** Numbers in parentheses, standard error.

†† Weight (kg)/height $(m)^2$.

Figure 2 shows the results of the reliability study. Pearson correlations of initial measures with repeat measures were 0.72, 0.70, 0.46, and 0.64 for biomarker energy, FFQ energy, biomarker protein, and FFQ protein, respectively. The coefficient of variation between the two visits was 1.1 percent for biomarker TEE and 2.7 percent for FFQ energy. For protein, these coefficients of variation were 4.5 percent and 5.4 percent for the biomarker and the self-reports, respectively. The correlation between the average of the two FFQ energy measures and the average of the two TEE measures was 0.14. This correlation for the two FFQ protein measures and corresponding duplicate biomarker measures was 0.26. Coefficients of variation for the blinded duplicates were 0.6 percent for TEE and 2.8 percent for urinary nitrogen.

Table 4 presents the regression calibration model coefficients for the logarithm of energy and protein. These coefficients, along with the observed covariate values for an individual, estimate true dietary intake under the measurement model (11, 41). The intercept estimates the expected log intake for a Caucasian woman in the WHI comparison arm with an average body mass index of 28.2, an average age of 70.9 years, and WHI-NBS-measured values for the other included covariates. For a 25 percent increase in FFQ energy, the estimated median increase in total energy intake was 1.4 percent. By comparison, for a 1-standard-deviation increase in body mass index (5.5), the estimated median increase in energy intake was 7.5 percent, and a 5-year increase in age was associated with a median decrease in energy intake of 2.4 percent. For dietary protein, a 5-year increase in age was associated with a median decrease in true intake of 4.0 percent, and a 1-standard-deviation increase in body mass index was associated with an estimated median increase of 6.8 percent. Further, for a 25 percent increase in FFQ protein, there was an estimated median increase in dietary intake of 4.8 percent. Similarly, for a 25 percent increase in FFQ percentage of energy from protein, the median increase in intake was 10.3 percent.

The model R^2 value was 0.31 for the energy calibration model (table 4); thus, the model explained 31 percent of the variation in log energy. Body mass index had the highest

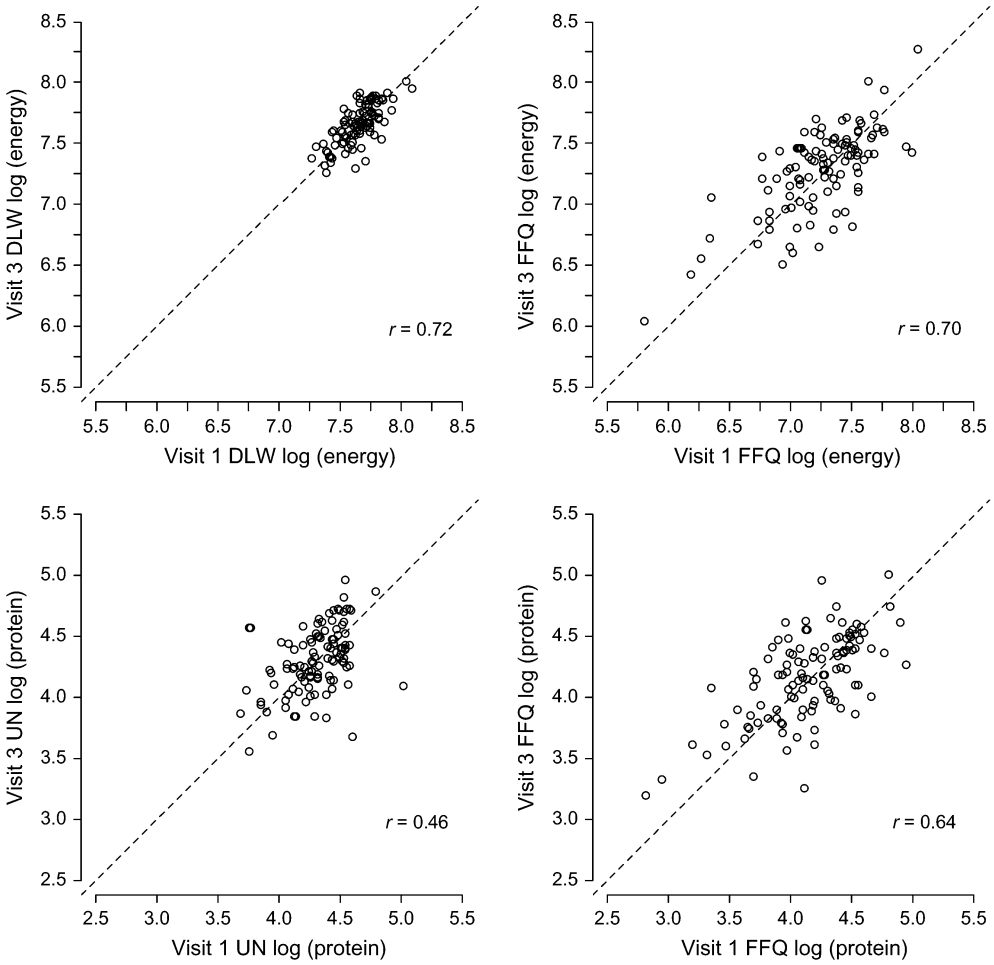

FIGURE 2. Comparisons of repeat measures of doubly labeled water (DLW) and urinary nitrogen (UN) with corresponding intakes from the self-reported food frequency questionnaire (FFQ) used in the Women's Health Initiative Nutritional Biomarkers Study, 2004–2005. The dotted line denotes the 45° line ($y = x$). Each plot gives Pearson correlations for the logarithm of visit 1 measures versus visit 3 measures of DLW total energy expenditure (kcal) ($n = 101$), FFQ energy intake (kcal) ($n = 111$), 24-hour UN-based protein intake (g) ($n = 111$), and FFQ protein intake (g) ($n = 111$).

partial R^2 at 0.22; thus, body mass index explained 22 percent of the variance in log energy that was not explained by other variables. By comparison, this figure was 3 percent for log(FFQ energy). For the protein model, R^2 was 0.23, and self-reported intake had the highest partial R^2 at 0.08. For percentage of energy derived from protein, R^2 was 0.12, and self-reported intake had the highest partial R^2 at 0.10 (data not shown). Note that a high R^2 coefficient is not necessary for regression calibration to be valid, since a large random component of measurement error could also be present.

In a sensitivity analysis, we compared the PABA-adjusted and unadjusted urinary nitrogen biomarkers. Of the 194 samples with PABA, only 10 (5 percent) had PABA recovery less than 70 percent (i.e., incomplete). For collections considered complete, 83 percent required no upward adjustment. The average difference between PABA-adjusted nitrogen and unadjusted nitrogen was 0.15 g (95 percent confidence interval: 0.08, 0.23). The geometric mean for protein estimated from PABA-adjusted samples was 15.4 (95 percent confidence interval: 14.8, 15.9), while the same samples' PABA-unadjusted geometric mean was 15.0 (95 percent confidence interval: 14.4, 15.5). We repeated the table 4 protein regression analyses using only PABA-adjusted biomarkers. Results were slightly different but did not materially alter our interpretation (data not shown).

DISCUSSION

In this study of 544 postmenopausal women, recovery biomarkers were used to characterize measurement properties of the WHI FFQ. Women randomized to the WHI-DM intervention arm underreported energy intake by 32 percent, while those in the WHI-DM comparison arm underreported energy intake by 27 percent. The similar levels of biomarker-measured energy and greater underreporting of FFQ energy (by approximately 125 kcal/day) in the intervention arm versus the comparison arm provide

TABLE 4. Regression calibration coefficients (*a*) for log-transformed total energy intake, total protein intake, and percentage of energy derived from protein in the Women's Health Initiative Nutritional Biomarkers Study, 2004–2005*

Characteristic	Total energy intake	Protein intake	% energy from protein
Intercept	7.61 (0.013)†	4.28 (0.024)	2.66 (0.011)
FFQ‡,§ (kcal)	0.062 (0.018)	0.211 (0.032)	0.439 (0.058)
Body mass index§,¶	0.013 (0.001)	0.012 (0.002)	−0.004 (0.002)
Age§ (years)	−0.005 (0.001)	−0.008 (0.002)	−0.005 (0.002)
Black race/ethnicity [Caucasian]#	−0.016 (0.017)	−0.130 (0.047)	
Hispanic race/ethnicity [Caucasian]	−0.004 (0.030)	−0.021 (0.056)	
Other race/ethnicity [Caucasian]	−0.093 (0.027)	−0.100 (0.058)	
Education [some college]			
High school/General Educational Development diploma or less		0.065 (0.032)	
College degree or more		0.033 (0.025)	
Annual income [$35,000–$49,999]			
<$20,000	−0.019 (0.021)	−0.053 (0.043)	
$20,000–$34,999	0.037 (0.017)	−0.009 (0.031)	
$50,000–$74,999	0.013 (0.018)	0.042 (0.035)	
≥$75,000	0.019 (0.018)	0.067 (0.035)	
Current smoking [nonsmoking]			−0.129 (0.064)
Physical activity (metabolic equivalents/week)	0.001 (0.0004)		
FFQ × body mass index§		−0.009 (0.005)	

* $E(Z|Q,V) = a_0 + a_1Q + a_2V$, where Z is the nutrient intake of interest, Q is the self-reported value for that nutrient, and V is a multivariate vector of personal characteristics.

† Numbers in parentheses, robust standard error.

‡ FFQ, food frequency questionnaire.

§ The FFQ variable was centered at mean self-reported intake (7.27 log kcal/day, 4.14 log g protein/day, and 2.85 log % energy from protein/day); body mass index was centered at mean body mass index (28.2); and age was centered at mean age (70.9 years). All dietary intake variables are presented on the log scale.

¶ Weight (kg)/height $(m)^2$.

Information in brackets, reference category. Body mass index and physical activity were modeled as continuous variables.

a possible explanation for the lack of sustained weight difference between the randomized groups (33). Specifically, it was previously reported that the intervention-group women reported an energy consumption approximately 100 kcal/day lower than did comparison-group women, on average, after the first year from enrollment (33). Hence, this difference may be attributable to a systematic bias in energy reporting. In a future analysis, we plan to apply the table 4 energy regression model to FFQ energy data collected over the course of the WHI-DM to assess whether any long-term weight differential should be expected between randomization groups. Note that this intervention program did not aim to alter long-term energy consumption.

We observed modest underreporting of protein intake in both the intervention (15 percent) and comparison (10 percent) groups and nearly identical (23 percent) overreporting of percentage of energy derived from protein. Unfortunately, there are no established recovery biomarkers for either fat or carbohydrate, so our inferences here are limited to those based on the energy and protein results. However, the overreporting of percentage of energy derived from protein, together with the underreporting of energy intake, suggests that both groups of participants may disproportionately underreport fat plus carbohydrate.

Few studies employing TEE as an objective biomarker of energy intake have been large enough to thoroughly examine systematic bias in relation to dietary misreporting. Results from the OPEN [Observing Protein and Energy Nutrition] Study suggested increased underreporting as body mass index increased but no clear associations with other participant characteristics (5). Among 35 low-income women who completed a TEE protocol and a dietary self-report, energy underreporting was strongly associated with increased body fatness (17). One previous study suggested that psychosocial characteristics predict underreporting (18). A few studies have utilized resting energy expenditure as a proxy for TEE (16, 43). Resting energy expenditure is a crude measure of energy expenditure, but these studies have also supported associations of more underreporting with increasing body mass index (15, 16, 43). In the WHI-NBS, our sample size was sufficiently large to observe systematic bias in energy reporting by obesity, age, and race/ethnicity.

Of the published studies on protein misreporting using recovery biomarkers (5, 43–45), few examined the role of

participant characteristics (5, 43). Heerstrass et al. (43) reported a positive association of obesity with protein underreporting, while in the OPEN Study, Subar et al. (5) reported no clear associations of participant characteristics with protein misreporting. Our results showed slight underreporting of protein by younger women and significant underreporting by income and race/ethnicity.

The findings that participant characteristics predicted energy and protein misreporting allowed us to compute separate calibration equations for energy, protein, and percentage of energy derived from protein. We plan to apply these biomarker-calibrated estimates of dietary self-reports in diet-disease analyses in WHI, where these nutrients may be important exposures. As such, the calibrated estimates may provide insights into diet-disease relations in WHI that would otherwise have been obscured by the measurement error in self-reports. While other investigators have provided extremely useful insights into the structure of measurement error using nutritional biomarkers to inform the regression models (6, 46), no other large cohort studies are in a position to use internally calibrated estimates of energy and protein for analyses of diet-disease associations. The fact that the calibrated estimate of energy consumption depends only weakly on FFQ energy and more strongly on other participant characteristics suggests that disease associations with calibrated energy may differ greatly from corresponding associations using uncalibrated energy (see Appendix table).

Most biomarker studies rely on measures from a single biospecimen. These measures are assumed to be representative of a person's usual nutritional status, presuming low within-person variation (13, 47). Here we included a reliability study, where 20 percent of participants completed all study procedures twice in a 6-month period. The coefficient of variation for the reliability sample TEE results was 1.1 percent. These coefficient of variation results, together with a Pearson correlation of 0.72, provide further evidence that TEE measures are reliable and reproducible (48). They also demonstrate that a reliability subset is necessary so that the variance in the biomarker measure can be adjusted for using an appropriate measurement error model.

There are several strengths of this study. WHI-NBS is one of the largest studies to date to use recovery biomarkers. The large sample size allowed us to test hypotheses related to systematic bias in underreporting, which is not possible with smaller sample sizes. Another strength is the reliability study; these replicate measures offer important data about the reproducibility of the recovery biomarkers and are essential to the calibration model developments. The reliability of the TEE estimates was quite good ($r = 0.72$). The reliability of the protein measures was more modest ($r = 0.46$), perhaps because of age-related fluid-balance issues or specimen collection errors. Some studies have used multiple days of urine collection to minimize day-to-day variability (44), and our correlations were similar to those of studies with multiple urine collections (45).

There are also limitations of this study. The number of minority participants was relatively small, preventing precise estimates of a calibration coefficient for race/ethnicity. A biomarker study with additional non-Caucasians will be necessary to more precisely correct measurement error for various racial/ethnic groups. Further, our objective biomarkers were limited to energy and protein, since there are no recovery biomarkers for fat, carbohydrate, or micronutrients. Additional work is needed to model blood micronutrient concentrations in relation to actual intake, perhaps in the context of human feeding studies, to develop methods for using these concentrations to calibrate corresponding self-report estimates of micronutrient consumption. Finally, there may be other unmeasured covariates with systematic measurement error that were not assessed as part of WHI.

In conclusion, among 544 postmenopausal women, recovery biomarkers confirmed that underreporting of nutrient consumption from a self-reported FFQ was common. The extent of misreporting was predicted by age, body mass index, and race/ethnicity. We used these data to create regression calibration equations for energy, protein, and percentage of energy derived from protein that will be applied to subsequent diet-disease analyses in WHI. Such analyses can be expected to further our understanding of the relation between dietary factors and chronic disease risk, while substantially addressing measurement error problems that have long plagued the field of nutritional epidemiology.

ACKNOWLEDGMENTS

This research was funded by grant N01WH22110 from the National Heart, Lung, and Blood Institute.

The Women's Health Initiative (WHI) Investigators—(National Heart, Lung, and Blood Institute, Bethesda, Maryland): Elizabeth Nabel, Jacques Rossouw, Shari Ludlam, Linda Pottern, Joan McGowan, Leslie Ford, and Nancy Geller; *WHI Clinical Coordinating Center*—(Fred Hutchinson Cancer Research Center, Seattle, Washington): Ross Prentice, Garnet Anderson, Andrea LaCroix, Charles L. Kooperberg, Ruth E. Patterson, and Anne McTiernan; (Wake Forest University School of Medicine, Winston-Salem, North Carolina) Sally Shumaker; (Medical Research Labs, Highland Heights, Kentucky) Evan Stein; (University of California at San Francisco, San Francisco, California) Steven Cummings; *WHI Clinical Centers*—(Albert Einstein College of Medicine, Bronx, New York) Sylvia Wassertheil-Smoller; (Baylor College of Medicine, Houston, Texas) Aleksandar Rajkovic; (Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts) JoAnn Manson; (Brown University, Providence, Rhode Island) Annlouise R. Assaf; (Emory University, Atlanta, Georgia) Lawrence Phillips; (Fred Hutchinson Cancer Research Center, Seattle, Washington) Shirley Beresford; (George Washington University Medical Center, Washington, DC) Judith Hsia; (Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, California) Rowan Chlebowski; (Kaiser Permanente Center for Health Research, Portland, Oregon) Evelyn Whitlock; (Kaiser Permanente Division of Research, Oakland, California) Bette Caan; (Medical College of Wisconsin, Milwaukee, Wisconsin) Jane Morley Kotchen; (MedStar Research Institute/Howard University, Washington, DC) Barbara V. Howard; (Northwestern University, Chicago/Evanston,

Illinois) Linda Van Horn; (Rush Medical Center, Chicago, Illinois) Henry Black; (Stanford Prevention Research Center, Stanford, California) Marcia L. Stefanick; (State University of New York at Stony Brook, Stony Brook, New York) Dorothy Lane; (The Ohio State University, Columbus, Ohio) Rebecca Jackson; (University of Alabama at Birmingham, Birmingham, Alabama) Cora E. Lewis; (University of Arizona, Tucson/Phoenix, Arizona) Tamsen Bassford; (State University of New York at Buffalo, Buffalo, New York) Jean Wactawski-Wende; (University of California at Davis, Sacramento, California) John Robbins; (University of California at Irvine, Irvine, California) F. Allan Hubbell; (University of California at Los Angeles, Los Angeles, California) Howard Judd; (University of California at San Diego, LaJolla/Chula Vista, California) Robert D. Langer; (University of Cincinnati, Cincinnati, Ohio) Margery Gass; (University of Florida, Gainesville/Jacksonville, Florida) Marian Limacher; (University of Hawaii, Honolulu, Hawaii) David Curb; (University of Iowa, Iowa City/Davenport, Iowa) Robert Wallace; (University of Massachusetts/Fallon Clinic, Worcester, Massachusetts) Judith Ockene; (University of Medicine and Dentistry of New Jersey, Newark, New Jersey) Norman Lasser; (University of Miami, Miami, Florida) Mary Jo O'Sullivan; (University of Minnesota, Minneapolis, Minnesota) Karen Margolis; (University of Nevada, Reno, Nevada) Robert Brunner; (University of North Carolina, Chapel Hill, North Carolina) Gerardo Heiss; (University of Pittsburgh, Pittsburgh, Pennsylvania) Lewis Kuller; (University of Tennessee, Memphis, Tennessee) Karen C. Johnson; (University of Texas Health Science Center, San Antonio, Texas) Robert Brzyski; (University of Wisconsin, Madison, Wisconsin) Gloria E. Sarto; (Wake Forest University School of Medicine, Winston-Salem, North Carolina) Denise Bonds; (Wayne State University School of Medicine/Hutzel Hospital, Detroit, Michigan) Susan Hendrix.

Conflict of interest: none declared.

REFERENCES

1. World Cancer Research Fund. Food, nutrition and the prevention of cancer: a global perspective. Washington, DC: American Institute for Cancer Research, 1997.
2. Prentice RL, Willett WC, Greenwald P, et al. Nutrition and physical activity and chronic disease prevention: research strategies and recommendations. J Natl Cancer Inst 2004;96:1276–87.
3. Public Health Service, US Department of Health and Human Services. The Surgeon General's report on nutrition and health. Washington, DC: US GPO, 1988. (DHHS publication no. (PHS) 88-50210).
4. Prentice RL. Measurement error and results from analytic epidemiology: dietary fat and breast cancer. J Natl Cancer Inst 1996;88:1738–47.
5. 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;158:1–13.
6. Kipnis V, Subar AF, Midthune D, et al. Structure of dietary measurement error: results of the OPEN biomarker study. Am J Epidemiol 2003;158:14–21.
7. 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.
8. Byers T. Food frequency dietary assessment: how bad is good enough? Am J Epidemiol 2001;154:1087–8.
9. Fraser GE. A search for truth in dietary epidemiology. Am J Clin Nutr 2003;78(suppl):521S–5S.
10. Kipnis V, Carroll RJ, Freedman LS, et al. Implications of a new dietary measurement error model for estimation of relative risk: application to four calibration studies. Am J Epidemiol 1999;150:642–51.
11. 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;5:977–84.
12. Kipnis V, Freedman LS, Brown CC, et al. Interpretation of energy adjustment models for nutritional epidemiology. Am J Epidemiol 1993;137:1376–80.
13. Kaaks R. Biochemical markers as additional measurements in studies of the accuracy of dietary questionnaire measurements: conceptual issues. Am J Clin Nutr 1997;65(suppl):1232S–9S.
14. Bingham SA, Day DE. 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.
15. Heitmann BL, Lissner L. Dietary underreporting by obese individuals: is it specific or non-specific? Br Med J 1995;311:986–9.
16. Horner NK, Patterson RE, Neuhouser ML, et al. Participant characteristics associated with errors in self-reported energy intake from the Women's Health Initiative food-frequency questionnaire. Am J Clin Nutr 2002;76:766–73.
17. Johnson RK, Soultanakis RP, Matthers DE. Literacy and body fatness are associated with underreporting of energy intake in US low-income women using the multiple-pass 24-hour recall: a doubly labeled water experiment. J Am Diet Assoc 1998;98:1136–40.
18. Tooze JA, Subar AF, Thompson FE, et al. Psychosocial predictors of energy underreporting in a large doubly labeled water study. Am J Clin Nutr 2004;79:795–804.
19. Hebert JR, Patterson RE, Gorfine M, et al. Differences between estimated caloric requirements and self-reported caloric intake in the Women's Health Initiative. Ann Epidemiol 2003;13:629–37.
20. Women's Health Initiative Study Group. Design of the Women's Health Initiative Clinical Trial and Observational Study. Control Clin Trials 1998;19:61–109.
21. Ritenbaugh C, Patterson RE, Chlebowski RT, et al. The Women's Health Initiative Dietary Modification Trial: overview and baseline characteristics of participants. Ann Epidemiol 2003;13(suppl):S87–97.
22. Prentice RL, Caan B, Chlebowski RT, et al. Low-fat dietary pattern and risk of invasive breast cancer: the Women's Health Initiative Randomized Controlled Dietary Modification Trial. JAMA 2006;295:629–42.
23. Beresford SA, Johnson KC, Ritenbaugh C, et al. Low-fat dietary pattern and risk of colorectal cancer: the Women's Health Initiative Randomized Controlled Dietary Modification Trial. JAMA 2006;295:643–54.
24. Howard BV, Van Horn L, Hsia J, et al. Low-fat dietary pattern and risk of cardiovascular disease: the Women's Health Initiative Randomized Controlled Dietary Modification Trial. JAMA 2006;295:655–66.
25. Patterson RE, Kristal AR, Carter RA, et al. Measurement characteristics of the Women's Health Initiative food frequency questionnaire. Ann Epidemiol 1999;9:178–87.

26. Bingham SA. Urine nitrogen as a biomarker for the validation of dietary protein intake. J Nutr 2003;133(suppl):921S–4S.
27. 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: 348S–54S.
28. Schoeller DA. Recent advances from application of doubly labeled water to measurement of human energy expenditure. J Nutr 1999;129:1765–8.
29. Cole TJ, Coward WA. Precision and accuracy of doubly labeled water energy expenditure by multipoint and two-point methods. Am J Physiol 1992;263:E965–73.
30. Blanc S, Colligan AS, Trabulsi J, et al. Influence of delayed isotopic equilibration in urine on the accuracy of the $^{2}H_{2}{}^{18}O$ method in the elderly. J Appl Physiol 2002;92:1036–44.
31. Krebs-Smith SM, Cleveland LE, Ballard-Barbash R, et al. Characterizing food intake patterns of American adults. Am J Clin Nutr 1997;65(suppl):1264S–8S.
32. Slesinski MJ, Subar AF, Kahle LL. Dietary intake of fat, fiber and other nutrients is related to the use of vitamin and mineral supplements in the United States: the 1992 National Health Interview Survey. J Nutr 1996;126:3001–8.
33. Howard BV, Manson JE, Stefanick ML, et al. Low-fat dietary pattern and weight change over 7 years: the Women's Health Initiative Dietary Modification Trial. JAMA 2006;295:39–49.
34. Bingham S. The use of 24-h urine samples and energy expenditure to validate dietary assessments. Am J Clin Nutr 1994;59(suppl):227S–31S.
35. Bingham SA, Williams R, Cole TJ, et al. Reference values for analytes of 24-h urine collections known to be complete. Ann Clin Biochem 1988;25:610–19.
36. 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;46:131–5.
37. Bingham SA, Cummings JH. Creatinine and PABA as markers for completeness of collection of 24-hour urine samples. Hum Nutr Clin Nutr 1986;40:473–6.
38. Kristal AR, Shattuck AL, Williams AE. Food frequency questionnaires for diet intervention research. In: 17th National Nutrient Databank conference proceedings. Baltimore, MD: International Life Sciences Institute, 1992:110–25.
39. Schakel SF, Buzzard IM, Gebhardt SE. Procedures for estimating nutrient values for food composition databases. J Food Compost Anal 1997;10:102–14.
40. Patterson RE, Levy L, Tinker LF, et al. Evaluation of a simplified vitamin supplement inventory developed for the Women's Health Initiative. Public Health Nutr 1999;2:273–6.
41. Sugar EA, Wang CY, Prentice RL. Logistic regression with exposure biomarkers and flexible measurement error. Biometrics 2007;63:143–51.
42. Diggle PJ, Heagerty PJ, Liang KY, et al. Analysis of longitudinal data. New York, NY: Oxford University Press, 2002.
43. Heerstrass D, Ocke M, Bueno-de-Mesquita H, et al. Underreporting of energy, protein and potassium intake in relation to body mass index. Int J Epidemiol 1998;27:186–91.
44. McKeown NM, Day NE, Welch AA, et al. Use of biological markers to validate self-reported dietary intake in a random sample of the European Prospective Investigation into Cancer United Kingdom Norfolk cohort. Am J Clin Nutr 2001;74: 188–96.
45. 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 2000;30:309–17.
46. Day N, Wong M, Bingham S, et al. Correlated measurement error—implications for nutritional epidemiology. Int J Epidemiol 2004;33:1373–81.
47. Neuhouser ML, Patterson RE, King IB, et al. Selected nutritional biomarkers predict diet quality. Public Health Nutr 2003;6:703–9.
48. Willett W. Invited commentary: OPEN questions. Am J Epidemiol 2003;158:22–4.
49. Liang K, Zeger A. Longitudinal data analysis using generalized linear models. Biometrika 1986;73:13–22.

APPENDIX

Regression Calibration Estimator

Let $Z(i,j)$ represent a nutrient intake, such as long-term average daily energy consumption, protein intake, or protein density (or their log-transformed values) for the ith study subject. Instead of directly observing this value, we obtain an estimate $Q(i,j)$ from a food frequency questionnaire (FFQ) or an estimate $W(i,j)$ from a biomarker, where the subscript j refers to the jth replicate measurement for the ith individual. To determine the regression calibration estimate of Z, the following statistical model is assumed:

$$\begin{aligned} Q_{ij} &= \delta_0 + \delta_1 Z_i + \delta_2 V_i + \delta_3 V_i Z_i + r_i + u_{ij}. \\ W_{ij} &= Z_i + \varepsilon_{ij}. \end{aligned} \quad (1)$$

Here V_i is a covariate (or multidimensional vector) of personal characteristics for the ith subject, such as body mass index or obesity status, that may affect the amount of systematic reporting bias in the FFQ. Related measurement error models have previously been considered (10, 11). In equation 1, δ_0 and δ_1 allow the FFQ assessment to be relocated and rescaled from the actual intake, while δ_2 and δ_3 allow the magnitude of the relocation and rescaling to depend on participant characteristics. For example, suppose that Z is energy consumption in kilocalories and V is a binary indicator for obesity ($V = 1$ if body mass index ≥ 30 and $V = 0$ otherwise) and that $\delta_0 = 50$ and $\delta_1 = 0.40$; then the expected value for the intake reported on the FFQ by a nonobese person would be 40 percent of the true value, plus an offset of 50 kcal. If $\delta_2 = 10$ and $\delta_3 = -0.10$, then one would expect an obese person to be reporting only 30 percent of his or her true intake, with an offset of 60. The r_i term in equation 1 is a subject-specific bias term assumed to have a mean of zero across all subjects, which allows one to correlate the measurement error for repeated observations of $Q(i,j)$, $j = 1, 2, \ldots$ for an individual. The u_{ij} term is mean-zero random error, which is uncorrelated with the other terms on the right-hand side of equation 1. This statistical model assumes that the biomarker W contains no systematic error but rather is composed of the true intake plus a mean-zero random error term. This error, ε_{ij}, is assumed to be independent of all terms on the right-hand side of the equation for Q; that is, a classical measurement model is assumed for W.

The regression calibration estimator for Z for a person for whom no biomarker is observed is an estimate of $E(Z|Q,V)$,

the expected value of the actual intake given the reported FFQ intake and relevant personal characteristics. From equation 1, the form of this conditional expectation under normality assumptions for (Z,r,u), given V, is

$$E(Z|Q,V) = a_0 + a_1 Q + a_2 V + a_3 VQ. \quad (2)$$

Because the biomarker follows the classical measurement error model, with $W = Z + \varepsilon$, one has $E(W|Q,V) = E(Z|Q,V)$. It follows that a linear regression of W on the covariates Q and V and their interaction VQ provides unbiased estimates for the a coefficients. These coefficients are estimated using the Women's Health Initiative Nutritional Biomarkers Study cohort and are presented in table 4. The estimates shown in table 4 were obtained using multivariate linear regression with robust variance estimation (42, 49).

Prentice et al. (11) allowed the variance of r_i, the subject-specific random effect in equation 1, to depend on aspects of V. The model shown in equation 1 could easily be generalized in this manner. In an exploratory analysis of the residuals from the models in table 4, we examined the data for possible relations between the residual variance and body mass index and age. This analysis did not suggest the need for either an age-specific variance term or a body mass index-specific variance term in equation 1.

APPENDIX TABLE. Estimates of energy intake (kcal/day) obtained by self-reported food frequency questionnaire, a biomarker (total energy expenditure), and a calibrated food frequency questionnaire, according to body mass index category, Women's Health Initiative Nutritional Biomarkers Study, 2004–2005*

Body mass index† category	Self-reported FFQ‡		Total energy expenditure		Calibrated FFQ	
	Geometric mean	IQR‡	Geometric mean	IQR	Geometric mean	IQR
Normal (<25.0)	1,407	1,157–1,759	1,894	1,714–2,083	1,912	1,853–1,980
Overweight (25.0–29.9)	1,462	1,196–1,837	2,043	1,904–2,232	2,028	1,962–2,103
Obese (≥30)	1,454	1,161–1,897	2,213	2,034–2,415	2,247	2,156–2,338

* Note that the difference between FFQ energy intake (self-report) and total energy expenditure (biomarker) increases as body mass index increases. The biomarker-calibrated estimates, for the same women, correct for the measurement error using the model shown in table 4.

† Weight (kg)/height $(m)^2$.

‡ FFQ, food frequency questionnaire; IQR, interquartile range (25th–75th percentiles).