



Systematic Reviews and Meta- and Pooled Analyses

Pooled Results From 5 Validation Studies of Dietary Self-Report Instruments Using Recovery Biomarkers for Energy and Protein Intake

Laurence S. Freedman*, John M. Commins, James E. Moler, Lenore Arab, David J. Baer, Victor Kipnis, Douglas Midthune, Alanna J. Moshfegh, Marian L. Neuhouser, Ross L. Prentice, Arthur Schatzkin†, Donna Spiegelman, Amy F. Subar, Lesley F. Tinker, and Walter Willett

* Correspondence to Laurence S. Freedman, Biostatistics Unit, Gertner Institute for Epidemiology and Health Policy Research, Sheba Medical Center, Tel Hashomer 52621, Israel (e-mail: lsf@actcom.co.il).

† Deceased.

Initially submitted November 20, 2013; accepted for publication April 14, 2014.

We pooled data from 5 large validation studies of dietary self-report instruments that used recovery biomarkers as references to clarify the measurement properties of food frequency questionnaires (FFQs) and 24-hour recalls. The studies were conducted in widely differing US adult populations from 1999 to 2009. We report on total energy, protein, and protein density intakes. Results were similar across sexes, but there was heterogeneity across studies. Using a FFQ, the average correlation coefficients for reported versus true intakes for energy, protein, and protein density were 0.21, 0.29, and 0.41, respectively. Using a single 24-hour recall, the coefficients were 0.26, 0.40, and 0.36, respectively, for the same nutrients and rose to 0.31, 0.49, and 0.46 when three 24-hour recalls were averaged. The average rate of under-reporting of energy intake was 28% with a FFQ and 15% with a single 24-hour recall, but the percentages were lower for protein. Personal characteristics related to under-reporting were body mass index, educational level, and age. Calibration equations for true intake that included personal characteristics provided improved prediction. This project establishes that FFQs have stronger correlations with truth for protein density than for absolute protein intake, that the use of multiple 24-hour recalls substantially increases the correlations when compared with a single 24-hour recall, and that body mass index strongly predicts under-reporting of energy and protein intakes.

24-hour recall; attenuation factors; calibration equations; dietary measurement error; food frequency questionnaire; under-reporting

Abbreviations: AMPM, Automated Multiple Pass Method; BMI, body mass index; FFQ, food frequency questionnaire; NBS, Nutrition Biomarker Study; NPAAS, Nutrition and Physical Activity Assessment Study; OPEN, Observing Protein and Energy Nutrition; USDA, United States Department of Agriculture; WHI, Women's Health Initiative.

Most studies of dietary intakes or their relations to health outcomes use a dietary self-report instrument that is completed by participants (1). However, data from such instruments contain reporting errors (2). Investigators need to know the magnitude and direction of such errors in order to assess their impact on research results. Therefore, validity assessment of the self-report instrument is commonly performed. Relatively brief self-report instruments, such as a food frequency questionnaire (FFQ), are often compared in the validation exercise to more detailed self-reports, such as 24-hour recalls (3, 4). Dietary intake recovery biomarkers (5)

that provide accurate assessments of short-term intakes of a limited set of dietary components (e.g., energy, protein, potassium, and sodium) have also been used for validation. However, these biomarkers are expensive or inconvenient to measure and typical sample sizes are small, yielding limited information.

Recently, a series of larger validation studies that used recovery biomarkers, starting with the Observing Protein and Energy Nutrition (OPEN) Study in 2000 (6), was conducted in various US populations. In 2009, investigators from 5 such studies agreed to pool their data for common analysis with the

aim of describing with greater precision the nature and magnitude of reporting errors in FFQs and 24-hour recalls and investigating the personal characteristics associated with such errors. We present here results for intakes of energy and protein from this Validation Studies Pooling Project.

METHODS

Validation studies and their populations

The 5 validation studies were conducted with different aims and in diverse populations within the United States (Table 1). The OPEN Study was conducted to elucidate the measurement properties of self-report instruments in adult volunteers who were 40–69 years of age and resided in Maryland (6). The Energetics Study investigated similar questions, emphasizing multiple 24-hour recalls in younger white and black adults (7). The Automated Multiple Pass Method (AMPM) Study evaluated reporting by adults using the United States Department of Agriculture’s (USDA) primary 24-hour recall assessment tool for dietary intakes in the National Health and Nutrition Examination Survey (NHANES) (8). The Nutrition Biomarker Study (NBS) studied dietary reporting by participants in the Women’s Health Initiative (WHI) Dietary Modification Trial, using various self-report instruments and biomarkers (9). The Nutrition and Physical Activity Assessment Study (NPAAS) studied dietary and physical activity self-reports and biomarker levels among participants in the WHI Observational Cohort (10). The latter 2 studies included only women, nearly all of whom were older than 60 years of age.

In each study, at least 70% of the participants had college or postgraduate education (Table 1), and more than 90% were nonsmokers. Each study proposal received institutional review board approval, including approval of the manner in which informed consent was obtained from participants.

Self-report instruments

In each study, FFQs were administered to participants. Although repeated administrations were performed in the OPEN, NBS, NPAAS, and AMPM studies, the present analysis includes only the first administration. Three different FFQs were used; all included the most frequently consumed foods and the foods that contributed the most to nutrient (especially fat) intakes in the United States. The FFQs queried intakes over the past year in the OPEN and AMPM studies and over the past 3 months in NBS and NPAAS. The OPEN and Energetics studies used the Diet History Questionnaire, which includes questions about 124 food and beverage items, with follow-up questions regarding food preparation and type; portion size is categorized as falling into 1 of 3 portion size ranges (11). The Harvard FFQ (used in the AMPM Study) includes questions about 146 items, with a single reference portion size (3). The WHI FFQ (used in NBS and NPASS) includes questions about 122 items and includes summary and adjustment questions; portion size is categorized as small, medium (with a reference size), or large (12, 13).

Each study included 2 or more 24-hour recall assessments. These were administered to all participants in 4 studies and

Table 1. Characteristics of Participants in the OPEN (1999–2000), Energetics (2006–2009), AMPM (2002–2004), NBS (2004–2005), and NPAAS (2007–2009) Studies

Study Name	First Author, Year (Reference No.)	Organization	No. of Participants	Mean Age (SD), years	% of Subjects Who Were Male	Mean BMI ^a (SD)	% Who Were Non-Hispanic White	% Who Were Non-Hispanic Black	% With a College Education	% With a Postgraduate Education
OPEN	Subar, 2003 (6)	National Cancer Institute	484	53.4 (8.3)	54	27.9 (5.3)	83	6	54	32
Energetics	Arab, 2010 (7)	University of California, Los Angeles	263	37.8 (12.6)	36	26.8 (6.2)	49	51	81	15
AMPM	Moshfegh, 2008 (8)	United States Department of Agriculture	524	49.5 (10.9)	50	26.6 (4.6)	77	13	54	39
NBS	Neuhouser, 2008 (9)	Women’s Health Initiative	544	70.9 (6.3)	0	28.2 (5.5)	83	11	40	31
NPAAS	Prentice, 2011 (10)	Women’s Health Initiative	450	70.5 (6.0)	0	28.5 (6.4)	64	18	38	38

Abbreviations: AMPM, Automated Multiple Pass Method; BMI, body mass index; NBS, Nutrition Biomarker Study; NPAAS, Nutrition and Physical Activity Assessment Study; OPEN, Observing Protein and Energy; SD, standard deviation.
^a Weight (kg)/height (m)².

to a subset of 20% in NBS (Table 2). Different versions of the 24-hour recall were used. The OPEN Study used a pencil-and-paper version of USDA's in-person interviewer-administered automated multiple-pass method that included 2- and 3-dimensional models to assist with estimation of portion sizes (6). The AMPM Study used a computer-automated version of this same method. The first recall was conducted in person with the 2- and 3-dimensional portion size aids; the second and third recalls were administered via telephone with the participants using a food model booklet and measuring cups and spoons as portion size aids (8). Both the OPEN and AMPM studies analyzed recalls using the Food and Nutrient Database for Dietary Studies, version 1.0 (14). NBS and NPAAS used the Nutrition Data System for Research (2005 nutrient database) interviewer-administered multiple-pass method administered via telephone, with a food model booklet to aid in estimation of portion sizes (12, 15). The Energetics Study used DietDay, a web-based self-administered 24-hour recall; 4 computer images per food aided in portion size estimation. Analysis used a nutrient database comprising values from the USDA's Standard Reference 23, Food and Nutrient Database for Dietary Studies version 4.1, product labeling information, and recipes for some mixed dishes (7).

Biomarkers

Each study used recovery biomarkers, including doubly labeled water for energy intake (16) and 24-hour urinary nitrogen for protein intake (17) (Table 2). Data on 24-hour urinary potassium and sodium intakes will be presented elsewhere.

Doubly labeled water measures energy expenditure over a 10–14 day period and, assuming individuals are in energy balance, is used to measure average daily energy intake over this period (16). In 4 studies, it was measured at the University of Wisconsin, and in the AMPM Study, it was measured at the USDA laboratory.

Twenty-four-hour urinary nitrogen level provides a measure of protein intake over a 24-hour period (17). In 3 studies, it was measured at the Medical Research Council (Cambridge, United Kingdom) using the Kjeldahl method, and in the other 2 studies it was measured at the USDA laboratory using the Dumas combustion method (Table 2). Three studies included repeat determinations in the main study protocol, separated by approximately 5 days; NBS and NPAAS included repeat determinations in a reliability substudy (see below). Urinary nitrogen in grams was divided by 0.81 to convert the measurement to dietary nitrogen (17). That number was then multiplied by 6.25 to convert dietary nitrogen to dietary protein.

Reliability substudies

Each study included a substudy, of varying sample size, to examine the reliability of self-reports and biomarker assessments. The time between initial and repeat administrations varied considerably, ranging from 2 weeks in the OPEN Study to approximately 6 months in the Energetics Study, NBS, and NPAAS and up to 10–23 months in the AMPM Study (Table 2). The extent of the repeat data collection also varied. In the OPEN study, only doubly labeled water

Table 2. Self-Report Instruments and Biomarkers Used in the OPEN (1999–2000), Energetics (2006–2009), AMPM (2002–2004), NBS (2004–2005), and NPAAS (2007–2009) Studies

Study	First Author, Year (Reference No.)	FFQ			24-Hour Recall		% of Patients Doing 24-Hour Recall	Laboratory That Measured the DLW	% of Patients Doing Repeated DLW Measurements	Time Between Repeat DLW Measurements	Urinary Nitrogen		Time Between Repeat Urine Assessments	% of Patients Doing Repeat Set of Urine Assessments
		Type	No. ^a	Type	No. ^a	Laboratory					Method	No. ^b		
OPEN	Subar, 2003 (6)	DHQ	2	AMPM, version 1	2	UW	UW	5	2 weeks	MRC	Kjeldahl	2	12 days	0
Energetics	Arab, 2010 (7)	DHQ	1	DayDiet	8	UW	UW	23	6 months	USDA	Dumas	2	10 days	23
AMPM	Moshfegh, 2008 (8)	Harvard	1	AMPM, version 2	3	USDA	USDA	11	10–23 months	USDA	Dumas	2	5 days	11
NBS	Neuhouser, 2008 (9)	WHI	1	AMPM with NDS-R	2	UW	UW	20	6 months	MRC	Kjeldahl	1	5 months ^c	20
NPAAS	Prentice, 2011 (10)	WHI	1	AMPM with NDS-R	3	UW	UW	20	6 months	MRC	Kjeldahl	1	6 months ^c	20

Abbreviations: AMPM, Automated Multiple Pass Method; DHQ, Diet-History Questionnaire; DLW, doubly labeled water; FFQ, food frequency questionnaire; MRC, Medical Research Council; NBS, Nutrition Biomarker Study; NDS-R, Nutrition Data System for Research; NPAAS, Nutrition and Physical Activity Assessment Study; OPEN, Observing Protein and Energy; USDA, United States Department of Agriculture; UW, University of Wisconsin; WHI, Women's Health Initiative.

^a Number of assessments.

^b Number of assessments in a set.

^c One urine assessment per set.

administration was repeated, whereas other studies repeated both biomarker assessments and self-reports. For example, NBS and NPAAS repeated the entire study protocol in a subsample comprising 20% of the study population (Table 2). Data on repeat biomarker and 24-hour recall determinations are included in our analyses.

Statistical methods

We report on 3 dietary components: energy, protein, and protein density. Protein density is defined as the ratio (%) of energy from protein to total energy. We excluded urinary protein values from the analysis if participants indicated missing 2 or more voids during the 24-hour collection. No exclusions were made on the basis of para-amino-benzoic acid results (18). The exclusion of outliers is described in Web Appendix 1 (available at <http://aje.oxfordjournals.org/>). Using the repeated 24-hour recall assessments, we investigated reporting characteristics for a single 24-hour recall, as well as for 2 and three 24-hour recalls (where available), using the mean log reported intake as the value derived from multiple assessments.

All analyses were based on the premise that recovery biomarker levels provide, on a logarithmic scale, unbiased estimates of short-term intake. With the additional conventional assumption that short-term intake does not vary systematically with time, short-term biomarkers are then unbiased for longer-term “usual” (i.e., average) intake.

We investigated several characteristics of dietary reporting error, including reporting bias, the attenuation factor, and the correlation coefficient between reported and true usual intakes. Reporting bias, the group mean difference between the reported and true usual intakes, is important when estimating or comparing mean intakes in populations. It was estimated as the mean difference between the log first reported intake and the log biomarker value and was re-expressed as relative bias by exponentiation.

The attenuation factor and correlation coefficient between reported and true intakes are important when estimating diet-health relationships. The attenuation factor (usually between 0 and 1) is the multiplicative bias or shrinkage factor in the estimated regression coefficient when a health outcome is regressed on continuous self-reported intake rather than true dietary intake. It was estimated as the slope in the linear regression of log biomarker value on log first reported intake. To accommodate multiple determinations of biomarker levels, linear mixed models (19) with a random intercept for participants were used (see Web Appendix 2 for further details). Across-study average attenuation factors were weighted by the inverse of their variances.

The correlation coefficient between reported and true intakes is used to measure loss of statistical power to detect diet-health associations when using reported intake instead of true intake (20). In simple models, it can also serve to de-attenuate relative risks between 2 categories of intake (21). It was estimated as the correlation between first reported intake and biomarker value adjusted for within-person biomarker variation, using a method similar to Rosner and Willett's (22) (Web Appendix 3). Low values of attenuation and correlation, for example, less than 0.4, are undesirable,

although there is no sharp cut off. A value of 0.4 would mean that a true relative risk of 2.0 would on average be attenuated to a value of $2^{0.4} = 1.32$ (see Discussion).

We also investigated how personal characteristics were associated with reporting bias and attenuation. We examined sex, age (<40, 40–49, 50–59, 60–69, 70–79, and ≥ 80 years), body mass index (BMI; weight (kg)/height (m)², log-transformed), race (black, white/other), and educational level (high school, college, postgraduate education). Their relation to bias was investigated through linear regressions of reported intake minus biomarker value on these characteristics, examining their regression coefficients. Their relation with attenuation was investigated through linear regressions of log biomarker value on log reported intake, the characteristics, and the interaction between a characteristic and reported intake. The coefficient of the interaction was interpreted as a measure of the change in attenuation associated with that characteristic. Calibration equations for predicting true usual intake were obtained from regressions of log biomarker value on log reported intake and personal characteristics. Accuracy of prediction was measured by the multiple correlation coefficient of the regression, adjusted for within-person biomarker variation (10) (Web Appendix 3).

We performed all of these analyses as meta-analyses with the study entered as a variable into the regression model (Web Appendix 2). Between-study heterogeneity was assessed through interactions between the study variable and other terms in the model and quantified by I^2 (23). Between-sex heterogeneity in attenuation factors was assessed through interaction between sex and self-report instrument. The statistical significance of coefficients was tested using 2-sided *t* tests or *F* tests. Although we used the 5% level as a guide for statistical significance, the tables presented cite many *P* values that were not adjusted for multiple testing. We interpret these *P* values cautiously and draw conclusions based on the consistency of results across studies, as well as the *P* values themselves. Statistical analyses were implemented in SAS, version 9 (SAS Institute, Inc., Cary, North Carolina) (24).

RESULTS

Reporting bias

Geometric mean intakes are shown in Table 3. Self-reported means for energy and protein intakes were uniformly lower than those based on biomarker levels. However, self-reported protein density means tended to exceed the biomarker means.

FFQ energy intake reporting bias was approximately 30% (range, 24%–32%) under-reporting across all studies for both sexes (Table 4). Twenty-four-hour recall energy intake under-reporting was approximately 10% (range, 6%–16%) in the OPEN, Energetics, and AMPM studies but approximately 25% (range, 24%–28%) in NBS and NPAAS. FFQ and 24-hour recall reporting biases for protein intake were generally lower than those for energy intake. With a FFQ, the rate of under-reporting of protein intake was approximately 10% (range, 5% over-reporting to 16% under-reporting) for all studies except the OPEN Study, for which the range was 26%–29% (Table 4). Protein under-reporting in

Table 3. Average Intakes of Energy, Protein, and Protein Density by Study, Sex, and Instrument in the OPEN (1999–2000), Energetics (2006–2009), AMPM (2002–2004), NBS (2004–2005), and NPAAS (2007–2009) Studies

Instrument by Sex	Study									
	OPEN		Energetics		AMPM		NBS ^a		NPAAS ^a	
	Geometric Mean	95% CI	Geometric Mean	95% CI	Geometric Mean	95% CI	Geometric Mean	95% CI	Geometric Mean	95% CI
<i>Energy, kcal^b</i>										
Men										
Biomarker	2,826	2,768, 2,885	2,997	2,892, 3,105	2,861	2,797, 2,925				
24-Hour recall ^c	2,522	2,428, 2,619	2,737	2,453, 3,053	2,480	2,388, 2,576				
FFQ	1,961	1,869, 2,057	2,275	2,086, 2,482	1,929	1,851, 2,010				
Women										
Biomarker	2,273	2,222, 2,325	2,241	2,179, 2,305	2,196	2,147, 2,246	2,059	2,033, 2,086	2,025	1,992, 2,059
24-Hour recall ^c	1,919	1,833, 2,009	2,096	1,951, 2,252	1,942	1,870, 2,017	1,520	1,436, 1,608	1,544	1,495, 1,594
FFQ	1,524	1,447, 1,605	1,658	1,543, 1,781	1,647	1,578, 1,720	1,461	1,420, 1,504	1,465	1,409, 1,522
<i>Protein, g</i>										
Men										
Biomarker	105.5	102.1, 109.1	104.5	97.4, 112.1	97.4	94.1, 100.9				
24-Hour recall ^c	92.2	88.1, 96.5	109.0	96.8, 122.9	94.6	90.4, 99.0				
FFQ	74.7	70.9, 78.6	89.1	81.5, 97.3	80.3	76.9, 83.7				
Women										
Biomarker	77.5	74.6, 80.5	70.2	66.4, 74.2	69.8	67.3, 72.5	72.4	70.7, 74.2	69.3	67.3, 71.3
24-Hour recall ^c	70.9	67.3, 74.7	83.4	77.1, 90.2	70.5	67.3, 73.8	64.1	60.1, 68.3	61.0	58.9, 63.2
FFQ	57.2	54.1, 60.5	61.5	56.7, 66.8	72.9	69.7, 76.4	63.2	61.2, 65.3	63.0	60.2, 65.8
<i>Protein Density, %</i>										
Men										
Biomarker	14.8	14.4, 15.3	13.9	12.9, 15.0	13.6	13.1, 14.0				
24-Hour recall ^c	14.6	14.1, 15.1	15.1	14.2, 16.1	15.3	14.9, 15.8				
FFQ	15.2	14.9, 15.6	15.9	15.2, 16.6	16.8	16.4, 17.2				
Women										
Biomarker	13.5	13.0, 14.1	12.8	12.2, 13.5	12.9	12.4, 13.4	14.1	13.8, 14.4	13.8	13.5, 14.2
24-hour recall ^c	14.5	14.0, 15.1	15.9	15.1, 16.8	14.5	14.0, 15.0	16.6	15.7, 17.5	15.8	15.4, 16.2
FFQ	15.1	14.6, 15.5	14.9	14.3, 15.5	17.7	17.3, 18.1	17.4	17.1, 17.7	17.3	17.0, 17.6

Abbreviations: AMPM, Automated Multiple Pass Method; CI, confidence interval; FFQ, food frequency questionnaire; NBS, Nutrition Biomarker Study; NPAAS, Nutrition and Physical Activity Assessment Study; OPEN, Observing Protein and Energy.

^a NBS and NPAAS included only women.

^b 1 kcal = 4.184 kJ.

^c Single administration of a 24-hour recall; data from the first recall were used except in the Energetics Study, in which data from the second recall were used (Web Appendix 1).

Table 4. Bias in Reported Intakes of Energy, Protein, and Protein Density by Study, Sex, and Instrument in the OPEN (1999–2000), Energetics (2006–2009), AMPM (2002–2004), NBS (2004–2005), and NPAAS (2007–2009) Studies

Instrument by Sex	Study										Average Relative Bias ^c , %	P Value ^d
	OPEN		Energetics		AMPM		NBS ^b		NPAAS ^b			
	Relative Bias ^a , %	95% CI	Relative Bias ^a , %	95% CI	Relative Bias ^a , %	95% CI	Relative Bias ^a , %	95% CI	Relative Bias ^a , %	95% CI		
<i>Energy</i>												
Men												
24-Hour recall ^e	-12	-15, -9	-13	-23, -2	-13	-16, -10					-13	0.92
FFQ	-32	-35, -28	-24	-31, -16	-32	-35, -28					-31	0.10
Women												
24-Hour recall ^e	-16	-20, -11	-6	-12, 1	-11	-15, -8	-28	-32, -23	-24	-27, -21	-18	<0.001
FFQ	-32	-36, -28	-27	-32, -23	-25	-28, -21	-30	-32, -27	-27	-30, -24	-28	0.067
<i>Protein</i>												
Men												
24-Hour recall ^e	-12	-16, -8	7	-6, 21	-2	-6, 3					-5	<0.001
FFQ	-29	-32, -25	-12	-21, -3	-16	-21, -12					-22	<0.001
Women												
24-Hour recall ^e	-9	-14, -3	20	11, 30	0	-4, 5	-21	-26, -15	-12	-16, -9	-5	<0.001
FFQ	-26	-30, -21	-12	-21, -3	5	-1, 11	-12	-15, -10	-9	-13, -5	-11	<0.001
<i>Protein Density</i>												
Men												
24-Hour recall ^e	-1	-4, 3	14	6, 23	14	10, 18					7	<0.001
FFQ	3	0, 6	18	10, 26	25	20, 29					14	<0.001
Women												
24-Hour recall ^e	7	1, 12	24	17, 32	12	8, 17	8	2, 16	13	9, 17	14	0.001
FFQ	11	7, 15	17	11, 23	38	32, 44	23	20, 26	25	22, 29	23	<0.001

Abbreviations: AMPM, Automated Multiple Pass Method; CI, confidence interval; FFQ, food frequency questionnaire; NBS, Nutrition Biomarker Study; NPAAS, Nutrition and Physical Activity Assessment Study; OPEN, Observing Protein and Energy.

^a % Relative Bias = 100 × exponential (mean log self-report – mean log biomarker value) – 100. Negative values indicate under-reporting.

^b NBS and NPAAS included only women.

^c Average weighted by the inverse of the variance.

^d P value for heterogeneity across studies.

^e Single administration of a 24-hour recall; data from the first recall were used except in the Energetics Study, in which data from the second recall were used (Web Appendix 1).

Table 5. Added Bias in Reported Intakes of Energy, Protein, and Protein Density According to Personal Characteristics Averaged Across the OPEN (1999–2000), Energetics (2006–2009), AMPM (2002–2004), NBS (2004–2005), and NPAAS (2007–2009) Studies

Covariate	Energy				Protein				Protein Density			
	% Added Bias ^a	95% CI	P Value ^b	P _{heterogeneity} ^c	% Added Bias ^a	95% CI	P Value ^b	P _{heterogeneity} ^c	% Added Bias ^a	95% CI	P Value ^b	P _{heterogeneity} ^c
	<i>24-Hour Recall^d</i>											
Age, years			0.65	0.11			0.11	0.40			0.56	0.89
<40 vs. 50–59	2.6	–3.9, 9.5			2.8	–4.5, 10.6			–0.2	–6.4, 6.2		
40–49 vs. 50–59	3.1	–2.2, 8.6			4.5	–1.5, 10.8			1.3	–3.7, 6.5		
60–69 vs. 50–59	2.1	–3.7, 8.2			–1.1	–7.4, 5.6			–2.7	–8.0, 2.9		
70–79 vs. 50–59	6.8	–2.1, 16.5			1.7	–7.7, 12.1			–4.1	–11.8, 4.3		
>80 vs. 50–59	6.0	–6.7, 20.2			15.1	–0.2, 32.8			3.6	–8.4, 17.0		
Men vs. women	–0.5	–4.6, 3.7	0.80	0.14	–3.3	–7.6, 1.3	0.16	0.30	–3.0	–6.8, 0.9	0.13	0.08
BMI ^e of 30 vs. 25	–6.7	–8.3, –5.1	<0.001	0.29	–4.8	–6.6, –3.0	<0.001	0.71	1.7	0.0, 3.4	0.05	0.12
Black race vs. other ^f	–2.1	–6.9, 2.9	0.41	0.07	5.9	0.1, 12.0	0.05	0.03	6.5	1.5, 11.8	0.01	<0.001
Educational level			<0.001	0.53			0.54	0.24			0.16	0.40
High school vs. college	–9.9	–14.7, –4.9			–3.4	–9.0, 2.6			5.2	–0.1, 10.8		
Postgraduate vs. college	–0.9	–4.7, 3.0			–0.7	–4.9, 3.7			0.5	–3.1, 4.3		
	<i>Food Frequency Questionnaire</i>											
Age, years			0.007	0.92			<0.001	0.87			0.003	0.57
<40 vs. 50–59	2.7	–4.5, 10.5			2.4	–5.7, 11.2			–0.5	–5.8, 5.0		
40–49 vs. 50–59	3.7	–2.2, 9.9			7.8	1.0, 15.1			3.0	–1.3, 7.6		
60–69 vs. 50–59	6.6	0.0, 13.5			1.1	–5.8, 8.6			–5.0	–9.4, –0.4		
70–79 vs. 50–59	11.4	2.7, 20.9			7.1	–2.2, 17.4			–4.3	–10.0, 1.7		
>80 vs. 50–59	24.7	11.7, 39.2			25.0	10.6, 41.2			3.8	–4.4, 12.6		
Men vs. women	–2.7	–7.1, 1.9	0.24	0.05	–10.3	5.9, 17.4	<0.001	<0.001	–7.0	–10.1, –3.8	<0.001	0.06
BMI ^e of 30 vs. 25	–5.0	–6.6, –3.4	<0.001	0.19	–2.9	–4.7, –1.1	0.002	0.22	2.0	0.7, 3.3	0.002	0.33
Black race vs. other ^f	–3.6	–8.4, 1.4	0.15	0.002	–3.6	–9.0, 2.1	0.21	0.33	–0.4	–4.2, 3.4	0.83	0.007
Educational level			0.05	0.58			0.05	0.99			0.52	0.24
High school vs. college	–5.9	–10.5, –1.1			–6.6	–11.6, –1.2			–1.5	–5.1, 2.2		
Postgraduate vs. college	–2.1	–5.8, 1.7			–0.9	–5.1, 3.5			0.7	–2.1, 3.7		

Abbreviations: AMPM, Automated Multiple Pass Method; BMI, body mass index; CI, confidence interval; NBS, Nutrition Biomarker Study; NPAAS, Nutrition and Physical Activity Assessment Study; OPEN, Observing Protein and Energy.

^a % added bias over and above the average bias. Calculated as $100 \times \text{exponential}(\text{regression coefficient}) - 100$. Negative values indicate an association with under-reporting. Use of added bias: The model-based average percentages of relative bias for the reference group (non-Black women who were 50–59 years of age with a BMI of 25 and a college education) were as follows: energy, –15% for 24-hour recall and –29% for FFQ; protein, –4% for 24-hour recall and –12% for FFQ; and protein density, 15% for 24-hour recall and 24% for FFQ. Table entries show the added bias associated with personal characteristics. For example, for a black man who was 60–69 years of age, had a BMI of 30 and a high school education, and reported energy on a FFQ, one should expect an extra relative bias of approximately -3.6 (Black) -2.7 (man) $+6.6$ (60–69 years of age) -5.0 (BMI 30) -5.9 (high school education) = -10.6% over and above the -29% for the reference group, that is, underestimation of approximately 40%.

^b P value for the covariate (based on log-likelihood ratio test).

^c P value for heterogeneity across studies (based on log-likelihood ratio test).

^d Single administration of a 24-hour recall.

^e Weight (kg)/height (m)².

^f Other includes non-Hispanic whites.

24-hour recalls averaged 5% but exhibited much heterogeneity across studies (range, 20% over-reporting to 21% under-reporting). The level of under-reporting of protein intake was lower than that for energy intake, leading to a tendency for over-reporting of protein density that was greater with a FFQ than with a 24-hour recall (Table 4).

A higher BMI was consistently associated with increased under-reporting of both energy and protein intakes using both FFQs and 24-hour recalls (Table 5). Having a high school education was also associated with more under-reporting of energy and protein intakes using either instrument than was having some college education. Compared with an age of 50–59 years, an age older than 59 years was associated with less under-reporting of energy intake on a FFQ. Other personal characteristics were not consistently associated with reporting bias across the studies.

Attenuation and correlation of reported intake with true usual intake

Attenuation factors for FFQ-reported energy intake were extremely low for both men and women (Table 6), with an average below 0.1. For a single 24-hour recall they were not much higher, with an average of approximately 0.1. Using the mean of 2 or three 24-hour recall administrations increased the attenuation factor to only approximately 0.15.

Attenuation factors for reported protein intake were higher than those for energy intake (Table 6); the average value for FFQs was 0.17, and the range for a single 24-hour recall was 0.22–0.24. When values from 2 and three 24-hour recall administrations were averaged, the attenuation factors were substantially higher at approximately 0.3 and 0.4, respectively.

FFQ attenuation factors for protein density were markedly higher than those for protein or energy, with an average value of approximately 0.4 (Table 6). These were higher than for a single 24-hour recall (average of 0.2–0.3). With 2 and three 24-hour recall administrations, the average value increased to around 0.4 and 0.5 respectively. Thus, for protein density, FFQ attenuation levels were on average similar to those for two 24-hour recall administrations. Twenty-four-hour recall attenuation factors for protein density also appeared higher than those for protein, but not as markedly so as for FFQs (Table 6).

Considerable across-study heterogeneity in attenuation factors values was seen, particularly for 24-hour recalls, with protein and protein density values in the AMPM Study generally higher than those in other studies (Table 6). Attenuation factors did not differ substantially between men and women (Table 6).

Correlation coefficients between reported and true usual intakes displayed patterns similar to those seen with attenuation factors (Table 7). For FFQs, the correlations for protein density were higher than those for protein (approximately 0.4 vs. 0.3). However, for 24-hour recalls, unlike attenuation factors, correlations for protein density were generally not higher than those for protein. For example, for two 24-hour recalls, the average correlation was approximately 0.45 for protein and approximately 0.40 for protein density (Table 7).

There was no clear evidence that personal characteristics were substantially related to 24-hour recall attenuation factors in men or women (Appendix Table 1) or to FFQ

attenuation factors in men. However, for FFQ-reported energy and protein intakes among women, there was evidence that having a higher BMI and being black were associated with lower attenuation factors, and for protein, that a higher educational level was associated with higher attenuation factors.

Calibration (prediction) equations for true usual intake

Appendix Tables 2 and 3 present calibration equations for predicting the logarithm of true usual intake based on a self-report instrument and personal characteristics for men and women, respectively. The coefficient for the logarithm of self-reported intake is provided for each study. BMI, age, and race were all strong predictors of energy intake and together raised the multiple correlation for prediction from less than 0.1 with the self-report instrument alone to 0.3 and higher, dependent on the study.

For protein, these same characteristics were important predictors of true intake among women. Among men, age was less important and educational level was more important (Appendix Table 2). With inclusion of personal characteristics in the prediction, multiple correlations rose substantially above those achieved with a self-report instrument only, but not to the same extent as for energy.

For protein density, personal characteristics did not add much to the prediction of usual intake. Interestingly, after introduction of personal characteristics, energy was predicted best of the 3 dietary components, followed by protein and then protein density, an order very different from the level of prediction achieved by self-report instruments alone.

DISCUSSION

Dietary self-reporting is currently indispensable for population surveillance of dietary intake, many studies of interventions to modify dietary intake, and most studies of diet-health outcome relationships. However, reporting errors and daily variations in dietary intakes may be barriers to achieving reliable results from these studies. Knowledge of the measurement properties of self-report instruments is required to interpret the results of studies that rely on such instruments. In the present study, we examined reported intakes of energy, protein, and protein density.

In some studies, estimating the group average intake is important. These include studies for estimating the population distribution of intakes (25) and behavioral intervention studies in which the outcome is intake of a nutrient or food group (26). In such studies, average intake is estimated directly from the self-report instrument, and our analysis has shown that energy intakes are under-reported with both FFQs and 24-hour recalls, whereas absolute protein is under-reported with FFQs. We have also shown that in most of the studies, under-reporting of energy and protein intakes was greater with a FFQ than with a 24-hour recall, the exceptions being the WHI studies that showed somewhat more under-reporting of protein with a 24-hour recall than with a FFQ (Table 4). Furthermore, with either instrument, under-reporting of energy intake is greater than that of protein intake, and consequently protein density tends to be over-reported.

Table 6. Attenuation Factors in Reported Intakes of Energy, Protein, and Protein Density in the OPEN (1999–2000), Energetics (2006–2009), AMPM (2002–2004), NBS (2004–2005), and NPAAS (2007–2009) Studies

Instrument by Sex	OPEN		Energetics		AMPM		NBS ^a		NPAAS ^a		Average ^b		P Value ^c	I ^{2d}
	AF	95% CI	AF	95% CI	AF	95% CI	AF	95% CI	AF	95% CI	AF	95% CI		
<i>Energy</i>														
Men														
One 24-hour recall	0.18	0.12, 0.25	0.07	0.01, 0.13	0.11	0.04, 0.18					0.12	0.08, 0.16	0.04	0.69
Two 24-hour recalls	0.25	0.17, 0.32	0.08	0.01, 0.14	0.16	0.08, 0.24					0.15	0.11, 0.20	0.004	0.82
Three 24-hour recalls			0.08	0.01, 0.15	0.19	0.10, 0.28					0.12	0.07, 0.18	0.08	0.67
FFQ	0.07	0.01, 0.12	0.07	0.00, 0.15	-0.03	-0.11, 0.04					0.04	0.00, 0.08	0.05	0.66
Women														
One 24-hour recall	0.10	0.04, 0.17	0.08	0.03, 0.14	0.09	0.01, 0.16	0.05	-0.04, 0.14	0.07	0.03, 0.12	0.08	0.05, 0.11	0.90	0.00
Two 24-hour recalls	0.15	0.07, 0.23	0.12	0.05, 0.19	0.09	0.00, 0.18	0.11	0.01, 0.21	0.14	0.08, 0.20	0.13	0.09, 0.16	0.87	0.00
Three 24-hour recalls			0.13	0.05, 0.21	0.15	0.05, 0.24			0.18	0.11, 0.25	0.15	0.11, 0.20	0.66	0.00
FFQ	0.04	-0.01, 0.10	0.11	0.06, 0.17	0.05	-0.01, 0.12	0.05	0.01, 0.08	0.10	0.06, 0.14	0.07	0.05, 0.09	0.14	0.43
<i>Protein</i>														
Men														
One 24-hour recall	0.21	0.13, 0.29	0.11	0.01, 0.20	0.30	0.23, 0.38					0.22	0.17, 0.27	0.007	0.80
Two 24-hour recalls	0.28	0.19, 0.37	0.11	-0.01, 0.23	0.46	0.37, 0.54					0.32	0.26, 0.37	<0.001	0.91
Three 24-hour recalls			0.10	-0.03, 0.24	0.54	0.44, 0.64					0.39	0.31, 0.47	<0.001	0.96
FFQ	0.16	0.10, 0.23	0.14	0.00, 0.28	0.19	0.09, 0.28					0.17	0.12, 0.22	0.85	0.00
Women														
One 24-hour recall	0.14	0.05, 0.22	0.16	0.06, 0.25	0.33	0.25, 0.41	0.24	0.13, 0.35	0.28	0.21, 0.36	0.24	0.20, 0.28	0.004	0.74
Two 24-hour recalls	0.19	0.08, 0.30	0.21	0.11, 0.31	0.42	0.33, 0.51	0.35	0.22, 0.47	0.34	0.24, 0.44	0.31	0.26, 0.36	0.005	0.73
Three 24-hour recalls			0.31	0.19, 0.42	0.50	0.40, 0.60			0.43	0.32, 0.54	0.42	0.36, 0.48	0.05	0.66
FFQ	0.14	0.06, 0.22	0.04	-0.06, 0.13	0.17	0.08, 0.27	0.22	0.16, 0.28	0.18	0.13, 0.24	0.17	0.14, 0.20	0.02	0.66
<i>Protein Density</i>														
Men														
One 24-hour recall	0.27	0.18, 0.36	0.33	0.16, 0.50	0.36	0.23, 0.48					0.30	0.24, 0.37	0.52	0.00
Two 24-hour recalls	0.39	0.28, 0.50	0.30	0.08, 0.51	0.50	0.36, 0.65					0.41	0.33, 0.49	0.25	0.28
Three 24-hour recalls			0.43	0.19, 0.67	0.60	0.40, 0.76					0.55	0.42, 0.68	0.26	0.21
FFQ	0.43	0.29, 0.57	0.45	0.16, 0.75	0.41	0.23, 0.59					0.42	0.32, 0.53	0.97	0.00

Table continues

Table 6. Continued

Instrument by Sex	OPEN		Energetics		AMPM		NBS ^a		NPAAS ^a		Average ^b		P Value ^c	I ^{2d}
	AF	95% CI	AF	95% CI	AF	95% CI	AF	95% CI	AF	95% CI	AF	95% CI		
Women														
One 24-hour recall	0.07	-0.04, 0.19	0.18	0.06, 0.31	0.42	0.30, 0.53	0.18	0.05, 0.31	0.20	0.10, 0.29	0.21	0.16, 0.26	0.001	0.77
Two 24-hour recalls	0.23	0.07, 0.38	0.24	0.07, 0.40	0.61	0.47, 0.76	0.32	0.17, 0.47	0.29	0.16, 0.42	0.34	0.27, 0.41	0.001	0.77
Three 24-hour recalls			0.43	0.26, 0.61	0.69	0.51, 0.86			0.44	0.29, 0.58	0.51	0.41, 0.60	0.06	0.65
FFQ	0.32	0.16, 0.48	0.39	0.23, 0.56	0.42	0.23, 0.62	0.41	0.30, 0.52	0.35	0.22, 0.49	0.38	0.32, 0.45	0.87	0.00

Abbreviations: AF, attenuation factor; AMPM, Automated Multiple Pass Method; CI, confidence interval; FFQ, food frequency questionnaire; NBS, Nutrition Biomarker Study; NPAAS, Nutrition and Physical Activity Assessment Study; OPEN, Observing Protein and Energy.

- ^a NBS and NPAAS included only women.
- ^b Average weighted by the inverse of the variance.
- ^c P value for heterogeneity across studies.
- ^d I² measure of heterogeneity.

These biases need to be considered when interpreting results of such studies.

Our study clearly confirms previous reports that a higher BMI is strongly related to under-reporting of energy and protein intakes (Table 5). Therefore, careful control for baseline BMI is needed when analyzing studies in which energy, protein, or protein density intake is the main outcome, such as in comparisons of intake levels among subpopulations. Additionally, such studies could be analyzed using a prediction equation for intake (Appendix Tables 2 and 3) as the outcome, thus removing, or at least reducing, reporting bias from the outcome measure. Similarly, it seems important to control carefully for age in studies using FFQ-reported protein or energy intake as the main outcome.

For studies relating a dietary intake to a health outcome, the attenuation factor and correlation coefficient between reported and true intakes are important. Attenuation factors are useful at the analysis stage for de-attenuating observed relative risks measured on the continuous scale of intake. At the design stage, correlation coefficients are useful for judging how much a plausible relative risk between categories of intake will be attenuated by using the self-report instrument. For example, if a relative risk between upper and lower quintiles of 2.0 is plausible and the correlation coefficient is ρ , then in simple situations the expected observed relative risk will be 2.0^ρ (21). (This result is parallel to a result of Fraser and Yan (27) for standardized intakes on the continuous scale.) In this case, if the correlation coefficient is less than 0.38, the expected observed relative risk will be less than 1.3. The required sample size is also related to this correlation, being proportional to its inverse-square. To avoid needing hugely inflated sample sizes, one usually needs a correlation of approximately 0.4 or more. Table 7 shows that, in our studies, the FFQs reach this level for protein density intake but not for absolute energy or protein intake. Also, averaging results from 2 or three 24-hour recall administrations attains this level of correlation for protein density and absolute protein but not for energy, although the protein results for men differ between the AMPM and Energetics studies. It is worth considering combining reports from multiple 24-hour recalls and a FFQ to further increase the correlations of reported intakes with true intakes, as in the study by Carroll et al. (28).

The FFQs in our studies assessed intake over the past 3 (NPASS, NBS) or 12 (OPEN, AMPM, Energetics) months, whereas biomarkers and 24-hour recalls assessed short-term intake. The biomarker assessments were therefore more proximal to the period assessed by the 24-hour recalls than the period assessed by the FFQs. This could cause overestimation of 24-hour recall correlations with long-term true intake and underestimation of FFQ correlations. Preliminary investigations using statistical modeling indicate that this does occur but not to a degree that would change our overall conclusions. Further examination of this issue under a variety of statistical models is needed.

Because attenuation factors are used to adjust attenuated estimates of relative risks, it is important to know whether they are modified by personal characteristics. One important observation from our study is that attenuation factors for men and women seem comparable. However, the information in Appendix Table 1 suggests that when women report using

Table 7. Correlation Between Reported Intakes and True Usual Intakes of Energy, Protein, and Protein Density in the OPEN (1999–2000), Energetics (2006–2009), AMPM (2002–2004), NBS (2004–2005), and NPAAS (2007–2009) Studies

Instrument by Sex	Correlation ^a Between Reported and True Intakes					
	OPEN	Energetics	AMPM	NBS ^b	NPAAS ^b	Average ^c
<i>Energy</i>						
Men						
One 24-hour recall	0.36	0.28	0.23			0.29
Two 24-hour recalls	0.41	0.27	0.27			0.32
Three 24-hour recalls		0.27	0.29			0.28
FFQ	0.16	0.21	0.08			0.16
Women						
One 24-hour recall	0.23	0.24	0.34	0.12	0.20	0.24
Two 24-hour recalls	0.26	0.29	0.30	0.23	0.28	0.27
Three 24-hour recalls		0.27	0.42		0.32	0.34
FFQ	0.11	0.33	0.22	0.13	0.34	0.25
<i>Protein</i>						
Men						
One 24-hour recall	0.35	0.26	0.49			0.38
Two 24-hour recalls	0.42	0.24	0.64			0.46
Three 24-hour recalls		0.20	0.65			0.48
FFQ	0.32	0.25	0.27			0.28
Women						
One 24-hour recall	0.26	0.30	0.52	0.50	0.42	0.41
Two 24-hour recalls	0.29	0.37	0.56	0.63	0.38	0.46
Three 24-hour recalls		0.46	0.59		0.43	0.49
FFQ	0.29	0.07	0.25	0.39	0.35	0.29
<i>Protein Density</i>						
Men						
One 24-hour recall	0.40	0.44	0.39			0.41
Two 24-hour recalls	0.48	0.33	0.45			0.42
Three 24-hour recalls		0.43	0.48			0.45
FFQ	0.43	0.38	0.33			0.38
Women						
One 24-hour recall	0.10	0.29	0.47	0.35	0.30	0.32
Two 24-hour recalls	0.25	0.29	0.54	0.51	0.30	0.40
Three 24-hour recalls		0.48	0.52		0.40	0.47
FFQ	0.33	0.47	0.30	0.48	0.51	0.43

Abbreviations: AMPM, Automated Multiple Pass Method; FFQ, food frequency questionnaire; NBS, Nutrition Biomarker Study; NPAAS, Nutrition and Physical Activity Assessment Study; OPEN, Observing Protein and Energy.

^a Correlation coefficient adjusted for within-person variation in the biomarker (see Web Appendix 2 for method).

^b NBS and NPAAS included only women.

^c Root mean square of the individual study values.

a FFQ, attenuation factors may differ according to BMI and race, although this phenomenon does not appear to apply to men or to reporting using a 24-hour recall. The result for women using a FFQ suggests the need for further research into the effects of such attenuation modification on the results of cohort studies of women and into methods of estimating relative risks in such circumstances. The article by Prentice

(29) provided an early effort related to fat intake and BMI in breast cancer cases. Related to this point, in measurement error-adjusted survival analysis with age as the time scale, attenuation factors may vary with age and need to be age-specific, as in risk set regression calibration (30, 31).

Appendix Tables 2 and 3 show that the correlation between true and predicted usual intakes can be greatly increased for

energy and protein by including personal characteristics alongside self-reported intake in a calibration (or prediction) equation. However, gains are more modest for protein density. The important predictors aside from self-report are BMI and race, as well as age for energy. Neuhouser et al. (9) and Tinker et al. (32) have proposed that such prediction (calibration) equations be used to estimate usual intakes and be entered in place of reported intake into regression models that relate dietary intake to health outcomes. In its simplest form, where the prediction is based only on the self-report and there are no confounders in the health outcome model, this approach coincides with the method of correcting estimation bias, known as linear regression calibration (33–35). When there are confounders in the health outcome model, regression calibration requires that the confounders are used also in the prediction equation (34). In addition, according to theory, additional predictors may be added to the calibration equation if they are independent of the health outcome variable conditional on the explanatory variables in the health outcome model (36). The choice of which variables to use in the prediction equation is complex, closely tied to the time period targeted for usual intake, and beyond the scope of this article. It is clear, however, that the principle of increasing the accuracy of prediction of usual intake is centrally important in nutritional epidemiology research.

Overall, our pooling study has clarified the strengths and weaknesses of 2 commonly used types of self-report instrument. The different FFQs used in these studies were all self-administered using paper and pencil and were developed with the intent to capture total usual nutrient intakes for most Americans, with special attention to measuring fat intake (11–12, 37). However, further research has led to a general acceptance that FFQs do not measure absolute energy intake well (6, 9). The modes of administration of the 24-hour recalls varied across studies from web-based to interviewer-administered (including in-person and telephone administration) and in the type of portion size aids used. Additionally, the populations differed quite widely in age and racial/ethnic composition. These factors no doubt contributed to between-study heterogeneity in some measures. Despite this heterogeneity, the present study has established firmly that the attenuations and correlations with truth for the FFQs studied are much improved for protein density compared with absolute protein, that multiple 24-hour recalls substantially decrease attenuation and increase correlations over those for a single 24-hour recall, and that BMI strongly predicts under-reporting of energy and protein intakes. Our analysis is based only on total energy, a single nutrient, protein, and protein density and does not necessarily generalize to other nutrients. Our analyses of potassium and sodium intakes, which are to be reported separately, support the view of others (38) that levels of dietary reporting error differ across nutrients. Also, the fact that the participants in these studies were nearly all nonsmokers may imply that they were more health-conscious than average and may therefore have reported their intakes more accurately than average. Thus, it is possible that reporting in the total population may be somewhat poorer than indicated by these studies.

Clearly, improvements in our current dietary assessment methods are desirable, and ongoing work on new automated

instruments, new dietary biomarkers, and incorporating these with the current self-report instruments should be supported and encouraged. Research to develop further recovery biomarkers likewise should be strongly supported.

ACKNOWLEDGMENTS

Author affiliations: Information Management Systems, Inc., Rockville, Maryland (Laurence S. Freedman, John M. Commins, James E. Moler); Biostatistics Unit, Gertner Institute for Epidemiology and Health Policy Research, Tel Hashomer, Israel (Laurence S. Freedman); Division of General Internal Medicine, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, California (Lenore Arab); Agricultural Research Service, Beltsville Human Nutrition Research Center, United States Department of Agriculture, Beltsville, Maryland (David J. Baer, Alanna J. Moshfegh); Biometry Research Group, Division of Cancer Prevention, National Cancer Institute, Bethesda, Maryland (Victor Kipnis, Douglas Midthune); Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington (Marian L. Neuhouser, Ross L. Prentice, Lesley F. Tinker); Nutritional Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland (Arthur Schatzkin); Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts (Donna Spiegelman, Walter Willett); Department of Nutrition, Harvard School of Public Health, Boston, Massachusetts (Donna Spiegelman, Walter Willett); Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts (Donna Spiegelman); Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, Maryland (Amy F. Subar); Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts (Walter Willett); and Harvard Medical School, Harvard University, Boston, Massachusetts (Walter Willett).

The Women's Health Initiative program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, US Department of Health and Human Services through contracts HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C, and HHSN271201100004C.

We thank the following people for their valuable contributions to this work: Dr. Alfonso Ang (University of California, Los Angeles), Dr. Sheila Bingham (deceased, Dunn Nutrition Unit, British Medical Research Council, Cambridge), Dr. Kevin Dodd (US National Cancer Institute), Dr. Nancy Potischman (National Cancer Institute), Dr. Dale Schoeller (University of Wisconsin), and Dr. Richard Troiano (National Cancer Institute). We also thank the WHI investigators and staff for their dedication.

A full listing of Women's Health Initiative investigators can be found at <http://www.whi.org/researchers/Documents%20%20Write%20a%20Paper/WHI%20Investigator%20Short%20List.pdf>.

Conflict of interest: Lenore Arab has intellectual property interests in DietDay, the web-based 24-hour recall instrument

used in the Energetics Study. All of the other authors report no conflicts.

REFERENCES

1. Willett WC. *Nutritional Epidemiology*. 3rd ed. New York: Oxford University Press; 2013.
2. Thompson FE, Subar AF. Dietary assessment methodology. In: Coulston AM, Boushey CJ, Ferruzzi MG, eds, *Nutrition in the Prevention and Treatment of Disease*. 3rd ed. San Diego, CA: Academic Press; 2013:3–40.
3. Willett WC, Sampson L, Stampfer MJ, et al. Reproducibility and validity of a semiquantitative food frequency questionnaire. *Am J Epidemiol*. 1985;122(1):51–65.
4. Slimani N, Kaaks R, Ferrari P, et al. European Prospective Investigation Into Cancer and Nutrition (EPIC) Calibration Study: rationale, design and population characteristics. *Public Health Nutr*. 2002;5(6B):1125–1145.
5. Kaaks R, Ferrari P, Ciampi A, et al. Uses and limitations of statistical accounting for random error correlations, in the validation of dietary questionnaire assessments. *Public Health Nutr*. 2002;5(6A):969–976.
6. 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.
7. Arab L, Wesseling-Perry K, Jardack P, et al. Eight self-administered 24-hour dietary recalls using the internet are feasible in African Americans and Whites: The Energetics Study. *J Am Diet Assoc*. 2010;110(6):857–864.
8. Moshfegh AJ, Rhodes DG, Baer DJ, et al. The US Department of Agriculture Automated Multiple-Pass Method reduces bias in the collection of energy intakes. *Am J Clin Nutr*. 2008;88(2):324–332.
9. 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.
10. Prentice RL, Mossavar-Rahmani Y, Huang Y, et al. Evaluation and comparison of food records, recalls, and frequencies for energy and protein assessment by using recovery biomarkers. *Am J Epidemiol*. 2011;174(5):591–603.
11. Subar AF, Thompson FE, Kipnis V, et al. Comparative validation of the Block, Willett, and National Cancer Institute food frequency questionnaires: the Eating at America's Table Study. *Am J Epidemiol*. 2001;154(12):1089–1099.
12. 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.
13. Women's Health Initiative. Food questionnaire. <https://www.whi.org/studydoc/WHI%20Forms/F060%20v1.6.pdf>. Published 1993. Updated 2003. Accessed April 14, 2014.
14. United States Department of Agriculture. *Food and Nutrient Database for Dietary Studies: Foods, Portions/Weights, Nutrients for Analyzing Dietary Data*. <http://www.ars.usda.gov/services/docs.htm?docID=12089>. Accessed June 2004. Updated October 23, 2013.
15. Women's Health Initiative. Section 10: Dietary Assessment. In: WHI Procedure Manual. Vol. 2. (<https://www.whi.org/studydoc/WHI%20and%20ES1%20Manual%20of%20Operations/1993-2005%20WHI%20CT%20and%20OS/Vol%202,%2010%20-%20Dietary%20Assessment.pdf>). Published 1993. Updated 1997. Accessed April 11, 2014.
16. 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(1):348S–354S.
17. Bingham SA, Cummings JH. Urine nitrogen as an independent validity measure of dietary intake: a study of nitrogen balance in individuals consuming their normal diet. *Am J Clin Nutr*. 1985;42(6):1276–1289.
18. Subar AF, Midthune D, Tasevska N, et al. Checking for completeness of 24-h urine collection using para-amino benzoic acid not necessary in the Observing Protein and Energy Nutrition study. *Eur J Clin Nutr*. 2013;67(8):863–867.
19. Verbeke G, Molenberghs G. *Linear Mixed Models for Longitudinal Data*. New York, NY: Springer-Verlag; 2000.
20. Kaaks R, Riboli E, van Staveren W. Calibration of dietary intake measurements in prospective cohort studies. *Am J Epidemiol*. 1995;142(5):548–556.
21. Kipnis V, Izmirlian G. The impact of categorization of continuous exposure measured with error [abstract]. *Am J Epidemiol*. 2002;155:S28.
22. Rosner B, Willett WC. Interval estimates for correlation coefficients corrected for within-person variation: implications for study design and hypothesis testing. *Am J Epidemiol*. 1988;127(2):377–386.
23. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med*. 2002;21(11):1539–1558.
24. SAS Institute Inc. Statistical Analysis System (SAS) Software, Version 9.2. Cary, NC: SAS Institute Inc; 2008.
25. Block G, Subar AF. Estimates of nutrient intake from a food frequency questionnaire: the 1987 National Health Interview Survey. *J Am Diet Assoc*. 1992;92(8):969–977.
26. Burke LE, Dunbar-Jacob J, Orchard TJ, et al. Improving adherence to a cholesterol-lowering diet: a behavioral intervention study. *Patient Educ Couns*. 2005;57(1):134–142.
27. Fraser GE, Yan R. A multivariate method for measurement error correction using pairs of concentration biomarkers. *Ann Epidemiol*. 2007;17(1):64–73.
28. Carroll RJ, Midthune D, Subar AF, et al. Taking advantage of the strengths of 2 different dietary assessment instruments to improve intake estimates for nutritional epidemiology. *Am J Epidemiol*. 2012;175(4):340–347.
29. Prentice RL. Measurement error and results from analytic epidemiology: dietary fat and breast cancer. *J Natl Cancer Inst*. 1996;88(23):1738–1747.
30. Xie SX, Wang CY, Prentice RL. A risk set calibration method for failure time regression by using a covariate reliability sample. *J R Stat Soc Series B Stat Methodol*. 2001;63(4):855–870.
31. Liao X, Zucker DM, Li Y, et al. Survival analysis with error-prone time-varying covariates: a risk set calibration approach. *Biometrics*. 2011;67(1):50–58.
32. Tinker LF, Sarto GE, Howard BV, et al. Biomarker-calibrated dietary energy and protein intake associations with diabetes risk among postmenopausal women from the Women's Health Initiative. *Am J Clin Nutr*. 2011;94(6):1600–1606.
33. Prentice RL. Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika*. 1982;69(2):331–342.
34. Carroll RJ, Ruppert D, Stefanski LA, et al. *Measurement Error in Nonlinear Models: A Modern Perspective*. 2nd ed. Boca Raton, FL: Chapman and Hall; 2006.

35. Rosner B, Willett WC, Spiegelman D. Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. *Stat Med.* 1989;8(9):1051–1069.
36. Kipnis V, Midthune D, Buckman DW, et al. Modeling data with excess zeros and measurement error: application to evaluating relationships between episodically consumed foods and health outcomes. *Biometrics.* 2009;65(4):1003–1010.
37. Rimm EB, Giovannucci EL, Stampfer MJ, et al. Reproducibility and validity of an expanded self-administered semiquantitative food frequency questionnaire among male health professionals. *Am J Epidemiol.* 1992;135(10):1114–1126.
38. Heitmann BL, Lissner L, Osler M. Do we eat less fat, or just report so? *Int J Obes Relat Metab Disord.* 2000;24(4):435–442.

Appendix Table 1. Modifiers of Attenuation: P Value and Direction of Effect of the Interaction Terms Between Self-Report and Covariates in Calibration Equations in the OPEN (1999–2000), Energetics (2006–2009), AMPM (2002–2004), NBS (2004–2005), and NPAAS (2007–2009) Studies

Covariate by Intake Type	FFQ P Value		Single 24-Hour Recall P Value	
	Men	Women	Men	Women
Energy				
Age	0.6	0.8	0.5	0.2
BMI ^a	0.7	0.04 ^b	0.2	0.9
Race (black vs. other ^c)	0.8	0.01 ^b	0.3	0.2
Educational level (high school vs. college vs. postgraduate)	0.9	0.8	0.2	0.9
Protein				
Age	0.8	0.3	0.03 ^b	0.6
BMI ^a	0.9	0.01 ^b	0.4	0.6
Race (black vs. other ^c)	0.7	0.5	0.9	0.05 ^b
Educational level (high school vs. college vs. postgraduate)	0.7	0.03 ^d	0.2	0.13
Protein density				
Age	0.3	0.5	0.4	0.4
BMI ^a	0.7	0.2	0.8	0.07
Race (black vs. other ^c)	0.4	0.0002 ^b	0.7	0.14
Educational level (high school vs. college vs. postgraduate)	0.3	0.11	0.5	0.5

Abbreviations: AMPM, Automated Multiple Pass Method; BMI, body mass index; FFQ, food frequency questionnaire; NBS, Nutrition Biomarker Study; NPAAS, Nutrition and Physical Activity Assessment Study; OPEN, Observing Protein and Energy.

^a Measured as weight (kg)/height (m)².

^b The attenuation factor tends to decrease toward 0 with increasing values of the variable (for race, other < black; for educational level, high school < college < postgraduate). For example, for FFQ-reported protein density intake, the FFQ attenuation factor is smaller (closer to 0) for black women than for other women.

^c Other includes non-Hispanic whites.

^d The attenuation factor tends to increase toward 1 with increasing values of the variable.

Appendix Table 2. Calibration Equations for Usual Intakes of Energy, Protein, and Protein Density, Multiple Correlations, and Squared Correlation for the Instrument Only in Men in the OPEN (1999–2000), Energetics (2006–2009), and AMPM (2002–2004) Studies

Covariate by Instrument	Energy Intake				Protein Intake				Protein Density Intake			
	Regression Coefficient	P Value	R ^{2a}	Instrument R ^{2b}	Regression Coefficient	P Value	R ^{2a}	Instrument R ^{2b}	Regression Coefficient	P Value	R ^{2a}	Instrument R ^{2b}
<i>Single 24-Hour Recall</i>												
Study		0.043 ^c				0.007 ^c				0.60 ^c		
OPEN	0.155		0.46	0.13	0.166		0.29	0.13	0.264		0.19	0.16
Energetics	0.064		0.40	0.08	0.096		0.27	0.07	0.303		0.27	0.20
AMPM	0.088		0.34	0.05	0.273		0.42	0.24	0.342		0.17	0.15
Age, years		<0.001				0.25				0.26		
<40 vs. 50–59	0.035				0.037				0.005			
40–49 vs. 50–59	0.029				0.022				0.008			
60–69 vs. 50–59	–0.074				–0.019				0.048			
BMI ^d (log-transformed)	0.534	<0.001			0.599	<0.001			0.026	0.69		
Black race vs. other ^e	–0.044	0.052			–0.159	<0.001			–0.103	0.006		
Educational level		0.27				0.036				0.21		
High school vs. college	0.042				0.050				0.044			
Postgraduate vs. college	0.007				0.048				0.032			
<i>Food Frequency Questionnaire</i>												
Study		0.051 ^c				0.67 ^c				0.88 ^c		
OPEN	0.065		0.38	0.03	0.163		0.31	0.10	0.424		0.21	0.18
Energetics	0.074		0.37	0.04	0.122		0.26	0.06	0.343		0.21	0.14
AMPM	–0.024		0.33	0.01	0.193		0.31	0.07	0.417		0.14	0.11
Age, years		<0.001				0.19				0.31		
<40 vs. 50–59	0.043	<0.001			0.041	0.19			0.014	0.31		
40–49 vs. 50–59	0.032				0.007				0.002			
60–69 vs. 50–59	–0.075				–0.031				0.044			
BMI ^d (log-transformed)	0.570	<0.001			0.675	<0.001			0.045	0.49		
Black race vs. other ^e	–0.044	0.069			–0.180	<0.001			–0.128	0.002		
Educational level		0.70				0.08				0.22		
High school vs. college	0.021				0.055				0.056			
Postgraduate vs. college	0.006				0.042				0.027			

Abbreviations: AMPM, Automated Multiple Pass Method; BMI, body mass index; OPEN, Observing Protein and Energy.

^a R² for model with self-report instrument and covariates.

^b R² for model with self-report instrument only.

^c P value for heterogeneity of adjusted attenuation coefficient across studies.

^d Measured as weight (kg)/height (m)².

^e Other includes non-Hispanic whites.

Appendix Table 3. Calibration Equations for Usual Intakes of Energy, Protein, and Protein Density, Multiple Correlations, and Squared Correlation for the Instrument Only in Women in the OPEN (1999–2000), Energetics (2006–2009), AMPM (2002–2004), NBS (2004–2005), and NPAAS (2007–2009) Studies

Covariate by Instrument	Energy				Protein				Protein Density			
	Regression Coefficient	P Value	R ^{2a}	Instrument R ^{2b}	Regression Coefficient	P Value	R ^{2a}	Instrument R ^{2b}	Regression Coefficient	P Value	R ^{2a}	Instrument R ^{2b}
<i>Single 24-Hour Recall</i>												
Study		>0.99 ^c				0.002 ^c				0.003 ^c		
OPEN	0.072		0.39	0.05	0.112		0.18	0.07	0.067		0.04	0.01
Energetics	0.071		0.38	0.06	0.159		0.18	0.09	0.197		0.13	0.08
AMPM	0.075		0.65	0.12	0.318		0.33	0.27	0.403		0.24	0.22
NBS	0.069		0.36	0.01	0.216		0.48	0.25	0.187		0.22	0.13
NPAAS	0.070		0.62	0.04	0.265		0.33	0.18	0.211		0.16	0.09
Age, years		<0.001				<0.001				0.046		
<40 vs. 50–59	0.040				0.033				0.014			
40–49 vs. 50–59	0.049				0.011				–0.042			
60–69 vs. 50–59	–0.024				0.008				0.020			
70–79 vs. 50–59	–0.091				–0.039				0.030			
>80 vs. 50–59	–0.148				–0.166				–0.066			
BMI ^d (log-transformed)	0.409	<0.001			0.345	<0.001			–0.051	0.19		
Black race vs. other ^e	–0.034	0.003			–0.087	<0.001			–0.062	0.003		
Educational level		0.61				0.15				0.047		
High school vs. college	–0.011				–0.033				–0.042			
Postgraduate vs. college	–0.004				0.007				0.014			
<i>Food Frequency Questionnaire</i>												
Study		0.18 ^c				0.007 ^c				0.84 ^c		
OPEN	0.014		0.34	0.01	0.117		0.20	0.08	0.316		0.14	0.11
Energetics	0.072		0.39	0.11	0.004		0.08	0.01	0.372		0.25	0.22
AMPM	0.038		0.59	0.05	0.140		0.13	0.06	0.442		0.11	0.09
NBS	0.049		0.45	0.02	0.201		0.36	0.15	0.409		0.28	0.23
NPAAS	0.079		0.69	0.11	0.150		0.28	0.13	0.350		0.33	0.26
Age, years		<0.001				<0.001				<0.001		
<40 vs. 50–59	0.039				0.054				0.007			
40–49 vs. 50–59	0.048				0.018				–0.049			
60–69 vs. 50–59	–0.032				0.013				0.041			
70–79 vs. 50–59	–0.086				–0.055				0.037			
>80 vs. 50–59	–0.139				–0.195				–0.064			

Table continues

Appendix Table 3. Continued

Covariate by Instrument	Energy				Protein				Protein Density			
	Regression Coefficient	P Value	R ^{2a}	Instrument R ^{2b}	Regression Coefficient	P Value	R ^{2a}	Instrument R ^{2b}	Regression Coefficient	P Value	R ^{2a}	Instrument R ^{2b}
BMI ^d (log-transformed)	0.391	<0.001			0.338	<0.001			-0.066	0.052		
Black race vs. other ^e	-0.027	0.008			-0.089	<0.001			-0.028	0.14		
Educational level		0.34				0.38				0.88		
High school vs. college	-0.009				-0.016				-0.005			
Postgraduate vs. college	0.005				0.010				0.004			

Abbreviations: AMPM, Automated Multiple Pass Method; BMI, body mass index; NBS, Nutrition Biomarker Study; NPAAS, Nutrition and Physical Activity Assessment Study; OPEN, Observing Protein and Energy.

^a R² for model with self-report instrument and covariates.

^b R² for model with self-report instrument only.

^c P value for heterogeneity of adjusted attenuation coefficient across studies.

^d Measured as weight (kg)/height (m)².

^e Other includes non-Hispanic whites.