American Journal of Epidemiology
Vol. 173, No. 6
DOI: 10.1093/aje/kwq415
Advance Access publication:
February 22, 2011

Practice of Epidemiology

Application of a Repeat-Measure Biomarker Measurement Error Model to 2 Validation Studies: Examination of the Effect of Within-Person Variation in Biomarker Measurements

Sarah Rosner Preis, Donna Spiegelman*, Barbara Bojuan Zhao, Alanna Moshfegh, David J. Baer, and Walter C. Willett

* Correspondence to Dr. Donna Spiegelman, Departments of Epidemiology and Biostatistics, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115 (e-mail: stdls@hsph.harvard.edu).

Initially submitted December 11, 2009; accepted for publication October 29, 2010.

Repeat-biomarker measurement error models accounting for systematic correlated within-person error can be used to estimate the correlation coefficient (ρ) and deattenuation factor (λ), used in measurement error correction. These models account for correlated errors in the food frequency questionnaire (FFQ) and the 24-hour diet recall and random within-person variation in the biomarkers. Failure to account for within-person variation in biomarkers can exaggerate correlated errors between FFQs and 24-hour diet recalls. For 2 validation studies, ρ and λ were calculated for total energy and protein density. In the Automated Multiple-Pass Method Validation Study ($n = 471$), doubly labeled water (DLW) and urinary nitrogen (UN) were measured twice in 52 adults approximately 16 months apart (2002–2003), yielding intraclass correlation coefficients of 0.43 for energy (DLW) and 0.54 for protein density (UN/DLW). The deattenuated correlation coefficient for protein density was 0.51 for correlation between the FFQ and the 24-hour diet recall and 0.49 for correlation between the FFQ and the biomarker. Use of repeat-biomarker measurement error models resulted in a ρ of 0.42. These models were similarly applied to the Observing Protein and Energy Nutrition Study (1999–2000). In conclusion, within-person variation in biomarkers can be substantial, and to adequately assess the impact of correlated subject-specific error, this variation should be assessed in validation studies of FFQs.

bias (epidemiology); biological markers; data collection; energy intake; nutrition assessment; proteins; validation studies

Abbreviations: AMPM, Automated Multiple-Pass Method; CI, confidence interval; DLW, doubly labeled water; FFQ, food frequency questionnaire; ICC, intraclass correlation coefficient; OPEN, Observing Protein and Energy Nutrition; UN, urinary nitrogen.

The use of the food frequency questionnaire (FFQ) to measure diet in epidemiologic studies has been a subject of recent debate (1–4). The validity of the FFQ has been evaluated by correlations with diet records or 24-hour diet recalls, and these data have also been used to obtain the deattenuation factor (λ), used to correct relative risk estimates for bias due to measurement error (5). However, diet records and 24-hour diet recalls are also subject to measurement error, and this error may be correlated with the error in the FFQ. This would imply that the use of diet records or 24-hour diet recalls in validation studies would lead to overestimation of the FFQ's validity in measuring diet as estimated by the correlation between the methods (6, 7). Indeed, due to concern about correlated errors between the 24-hour diet recall and the FFQ, because they both rely on memory and perceptions of serving sizes, some investigators have chosen diet records, which rely on neither, as the preferred comparison method for validation studies (8–17). However, the diet record is also limited because it requires high motivation on the part of the participant to obtain an accurate diet measure.

The Observing Protein and Energy Nutrition (OPEN) Study was designed to investigate the extent of correlated errors between the FFQ and the 24-hour diet recall (18, 19).

The OPEN Study measured total energy and protein intakes using 3 instruments: the FFQ, the 24-hour diet recall, and biomarkers—doubly labeled water (DLW) for total energy and urinary nitrogen (UN) for protein. Although low correlations between the FFQ and biomarkers were found for total energy intake and absolute protein intake, the problem appeared to be substantially mitigated for the relative measure of protein intake, protein density (UN/DLW), which is consistent with earlier suggestions that adjustment for energy intake improves validity by cancelling correlated errors between nutrients assessed by the FFQ (16). A limitation of the OPEN Study was the availability of replicate biomarker measures of total energy intake for 25 subjects, which were taken only 2 weeks after the baseline measurement. Because the aim of the FFQ is to assess average intake over the past year, without a replicate of the biomarker data collected many months after the first measurement, it is unknown whether the 2-week interval is sufficient to measure within-person variation in the biomarker. In this paper, we demonstrate that under certain measurement error models, failure to adequately account for within-person variability in the assessment method assumed to be unbiased can falsely lead to the appearance of correlated errors and to underestimation of the FFQ's validity.

The US Department of Agriculture Automated Multiple-Pass Method (AMPM) Validation Study was designed to use total energy intake and protein biomarkers to validate diet questionnaires. Our objective in the current analysis was to estimate the deattenuation factor (λ) for total energy intake, protein, and protein density in the AMPM and OPEN studies, using repeat-biomarker measurement error models which account for both correlated errors in the FFQ and the 24-hour diet recall and random within-person variation in the biomarkers. Additionally, we reanalyzed data from the OPEN Study with a measurement error model that accounts for within-person variability in DLW and UN, using estimates of within-person variation in the DLW biomarker derived from the AMPM Study.

MATERIALS AND METHODS

The OPEN Study

As was described previously (18), the National Cancer Institute conducted the OPEN Study between September 1999 and March 2000 among 484 participants (261 men, 223 women) aged 40–69 years from Montgomery County, Maryland. Three months apart, the participants completed 2 24-hour diet recalls and 2 FFQs. Total energy expenditure was measured using DLW. The participants took their first DLW dose at visit 1 and returned 2 weeks later to complete the protocol. DLW assessment was repeated in a substudy of 25 participants who received a second DLW dose at visit 2 and then returned 2 weeks later to complete the protocol. Dietary protein intake was measured through UN. The participants gave 2 24-hour urine specimens approximately 14 days apart. Twenty-four-hour protein intake was calculated by dividing UN by 0.81 and then multiplying by 6.25. Figure 1 gives the timeline for the OPEN Study. Additional details about the OPEN Study methods can be found in the Web Appendix, which is posted on the *Journal*'s Web site (http://aje.oxfordjournals.org/).

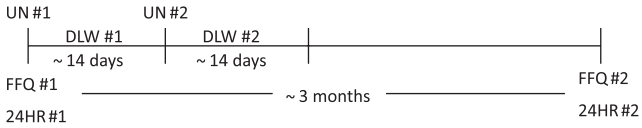

Figure 1. Timing of dietary measurements for the Observing Protein and Energy Nutrition Study, Montgomery County, Maryland, 1999–2000. The number of subjects with replicate measures was 479 for the food frequency questionnaire (FFQ), 482 for the 24-hour diet recall (24HR), 24 for doubly labeled water (DLW), and 297 for urinary nitrogen (UN).

The AMPM Validation Study

In the AMPM Study, 262 men and 262 women aged 30–69 years residing in the Baltimore, Maryland–Washington, DC, area completed 3 24-hour diet recalls over a 14-day study period in 2002–2003 using the US Department of Agriculture-developed computer-assisted multiple-pass 24-hour recall method (20). The first 24-hour diet recall was completed at the baseline visit, and the subsequent 24-hour diet recalls were completed via telephone. At the baseline visit, the participants received a dose of DLW, and over the next 2 weeks participants collected daily spot urine samples. During this 2-week period, participants completed 2 24-hour urine collections, which were averaged in the analysis, for measurement of UN. A substudy involving 52 participants was conducted approximately 16 months after the baseline examination to measure within-person variation in the biomarkers. The participants underwent a second DLW dosing, gave 2 additional 24-hour urine specimens for UN, and completed 3 additional 24-hour diet recalls over a 2-week period. The first FFQ was administered by mail several months after the baseline visit and was returned by 488 subjects. The second FFQ was given to participants in the substudy at the end of the 2-week period, completed at home, and returned by mail. The mean time elapsed between the first and second FFQs was 8.4 months (standard deviation, 0.6; range, 6.1–9.3 months). Figure 2 gives the timeline for the AMPM Study. Additional details about the AMPM Study methods can be found in the Web Appendix.

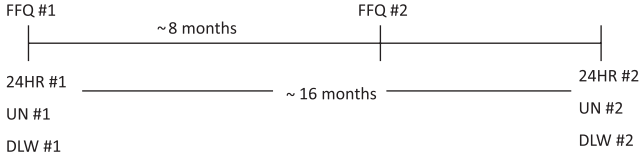

Figure 2. Timing of dietary measurements for the US Department of Agriculture Automated Multiple-Pass Method Study, Baltimore, Maryland–Washington, DC, 2002–2003. Each 24-hour diet recall (24HR) indicated on the figure represents the average of 3 24-hour diet recalls taken over a 2-week period. The number of subjects with replicate measures was 51 for the food frequency questionnaire (FFQ), 51 for the 24-hour diet recall, 50 for doubly labeled water (DLW), and 51 for urinary nitrogen (UN). UN #2 represents the average of 2 measurements taken over a 2-week period.

Statistical analysis

Mean values and standard deviations were calculated for total energy intake and protein for the FFQs, 24-hour diet recalls, and biomarker measurements, as applicable to each study design. Dietary protein was adjusted for total energy intake using the protein density (percentage of energy derived from protein) measure, calculated by multiplying protein intake (g/day) by 4 kcal/g and dividing by the participants' total energy intake. Data for all variables were natural log (ln)-transformed.

To assess reproducibility, we calculated intraclass correlation coefficients (ICCs) and 95% confidence intervals for correlations between repeat measures of the FFQ, 24-hour diet recall, or biomarker, as applicable to each study design, adjusted for sex. Sex-adjusted Pearson correlation coefficients and 95% confidence intervals were calculated for correlations between the second FFQ and the average of 2 24-hour diet recalls or biomarkers. The average of 3 24-hour diet recalls was used for the AMPM Study. Deattenuated Pearson correlation coefficients were calculated using the within- and between-person components of variation in 24-hour diet recall or biomarker measures. We calculated 95% confidence intervals for the deattenuated correlation coefficient as previously described (21, 22). The deattenuation factor (λ) was calculated from a linear regression of the assessment method that was assumed to be unbiased (either 24-hour diet recall or biomarker) on the surrogate measure (FFQ or 24-hour diet recall), adjusting for sex.

Spiegelman et al. (23) considered estimation of the deattenuation factor (λ) in the setting of a $\{Z^2X^2W^2\}$ design, where at least some validation study subjects have 2 or more measurements of the FFQ (Z), the diet records or 24-hour diet recall (X), and a biomarker (W). This development assumed that the diet records or, when reasonable, the 24-hour diet recall was unbiased, while the biomarker was biased, since concentration biomarkers rather than recovery biomarkers are typically available to most nutritional epidemiologists as follows:

$$\begin{aligned} Z_{ij} &= a + bx_i + r_i + \varepsilon_{Zij}, \quad j = 1,2 \\ X_{ij} &= x_i + s_i + \varepsilon_{Xij}, \quad j = 1,2 \\ W_{ij} &= e + fx_i + \varepsilon_{Wij}, \quad j = 1,2, \end{aligned} \tag{1}$$

where r is the systematic error in the FFQ, s is the systematic error in the 24-hour diet recall or diet records, $\mathrm{Var}(x) = \sigma_x^2$, $\mathrm{Var}(r) = \sigma_r^2$, $\mathrm{Var}(s) = \sigma_s^2$, $\mathrm{Var}(\varepsilon_Z) = \sigma_{\varepsilon_Z}^2$, $\mathrm{Var}(\varepsilon_X) = \sigma_{\varepsilon_X}^2$, $\mathrm{Var}(\varepsilon_W) = \sigma_{\varepsilon_W}^2$, $\mathrm{Cov}(\varepsilon_{Z_{ij}}, \varepsilon_{X_{ij}}) = \sigma_{\varepsilon_Z\varepsilon_X}$,$\mathrm{Cov}(r,s) = \sigma_{rs}$, and all other pairwise correlations between random terms are assumed to be zero, $j \geq 2$, and there are $i = 1, \ldots, n$ participants in the validation study. Two other groups of investigators, those from the OPEN Study (19) and those from the Women's Health Initiative (24), have made the assumption that the diet record or the 24-hour diet recall is biased but the biomarker is unbiased, leading to the following assumed model:

$$\begin{aligned} Z_{ij} &= a + bx_i + r_i + \varepsilon_{Zij}, \quad j = 1,2 \\ X_{ij} &= c + dx_i + s_i + \varepsilon_{Xij}, \quad j = 1,2 \\ W_{ij} &= x_i + \varepsilon_{Wij}, \quad j = 1,2, \end{aligned}$$

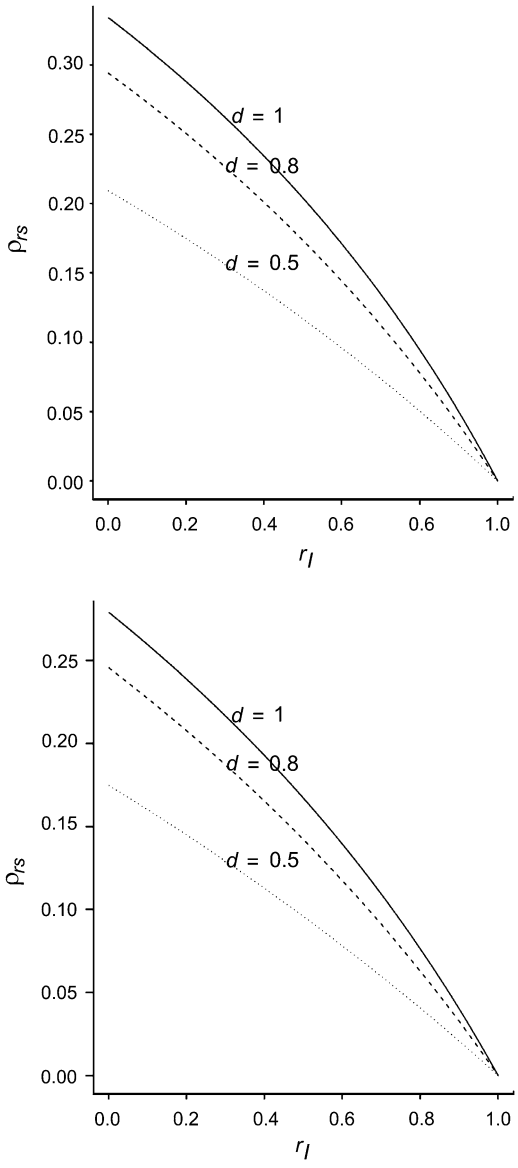

Figure 3. A plot of the expected value of $\widehat{\mathrm{Corr}}(r,s)$ ($\hat{\rho}_{rs}$) as a function of r_I, under the assumption that the true $\mathrm{Corr}(r,s) = 0$, in a hypothetical scenario. Top, $b = 1$; bottom, $b = 0.8$. The parameters b and d are the scale bias parameters for Z and X defined in the model.

where all terms are as defined above. Note that the standard model includes the first 2 lines of equation 1 or the first and third lines of equation 2, with $r = s = 0$ assumed for all subjects. Then, ordinary linear regression of X on Z will provide an estimate of the deattenuation factor under the assumption that there are no systematic within-person errors or, at least, that these errors are uncorrelated.

An expression for the convergent value of the estimated correlation between r and s, denoted $\mathrm{Corr}(r,s)$, is derived following the methods developed by Spiegelman et al. (23) when the true model is model 2 (with $\mathrm{Var}(\varepsilon_w) > 0$), but $\mathrm{Corr}(r,s)$ is estimated under the assumption that

Table 1. Baseline Characteristics of Participants in the Observing Protein and Energy Nutrition Study (1999–2000) and the US Department of Agriculture Automated Multiple-Pass Method Study (2002–2003)

	OPEN Study[a]						AMPM Study[b]					
	Validation Study			Substudy			Validation Study			Substudy		
	No.	%	Mean (SD)	No.	%	Mean (SD)	No.	%	Mean (SD)	No.	%	Mean (SD)
No. of participants	484			24			471			52		
Female sex	223	46		11	46		240	51		26	50	
Age, years			54 (8)			56 (10)			50 (11)			50 (11)
Body mass index[c]			27.8 (5.3)			28.2 (4.7)			26.6 (4.5)			27.2 (5.1)
Race												
White	411	85		21	88		382	81		44	85	
Black	30	6		0	0		58	12		2	10	
Other	43	9		3	12		31	7		3	6	
Hispanic ethnicity	19	4		1	4		13	3		2	4	

Abbreviations: AMPM, Automated Multiple-Pass Method; OPEN, Observing Protein and Energy Nutrition; SD, standard deviation.
[a] Baseline visit, 1999–2000.
[b] Baseline visit, 2002–2003.
[c] Weight (kg)/height (m)^2.

$\mathrm{Var}(\varepsilon_{W_{ij}}) = 0$. Setting $\theta_5 = \sigma_x^2$ and deleting θ_{11}, since it is not estimable in a $\{Z^2X^2W^1\}$ design, we obtain

$$\widehat{\mathrm{Corr}}(r,s) \xrightarrow{P}$$

$$\frac{\sigma_{rs} + bd\sigma_x^2(1-\rho_I^W)}{\sqrt{\sigma_s^2 + d^2\sigma_x^2(1-\rho_I^W)}\sqrt{\sigma_r^2 + b^2\sigma_x^2(1-\rho_I^W)}},$$

where b is the scale bias in the FFQ and d is the scale bias in the 24-hour diet recall as defined in model 2 above. Estimates of each of the 13 parameters in model 2 except ρ_I^W are presented in the article by Kipnis et al. (19), where ρ_I^W is the ICC of repeated measures of the biomarker over the course of a year. Values for ρ_I^W can be derived using data from Kipnis et al.'s Table 2 and their appendix (19). Under the assumption that the true $\mathrm{Corr}(r,s) = 0$, Figure 3 shows a plot of the expected value of the estimated correlation between the subject-specific error terms, $\widehat{\mathrm{Corr}}(r,s)$ (ρ_{rs}), as a function of ρ_I^W. Figure 3 suggests that if the within-person variation present in the biomarkers over a year was underestimated by the measurements taken 2 weeks apart, the OPEN Study investigators may have observed higher correlations between the subject-specific errors of the FFQ and the 24-hour diet recall than there truly were.

The methods of Spiegelman et al. (23) were used to evaluate the validity of the FFQ as a measure of total energy intake and protein as compared with DLW and UN biomarkers, assuming models 1 and 2. One feature of this method is that all of the data available for each measurement method are used to estimate the model parameters. Using the AMPM data, which have the study design $\{Z^2X^2W^2\}$ (23), indicating that some of the participants have 2 measurements for Z, X, and W, we estimated the deattenuation factor (λ), the correlation between the unobserved true exposure x and the FFQ ($\mathrm{Corr}(x,Z)$), and the ICC for correlation between repeated measures of each method. For these analyses, all nutrient variables were adjusted for sex using the residual method (25).

The OPEN Study has replicate measurements of DLW or UN taken approximately 14 days apart. It is unknown whether this is a long enough time period to adequately assess random within-person variation in these biomarkers. Under the design $\{Z^2X^2W^1\}$, models 1 and 2 cannot be fitted from the OPEN Study data alone. Instead, we used estimates of the ICC of the DLW and UN biomarkers from AMPM to obtain the expected values of the within- and between-person variance components in the OPEN Study as follows:

$$\hat{\sigma}^2_{e_W}(\mathrm{OPEN}) \approx (1 - \hat{\rho}_{\mathrm{I}}(\mathrm{AMPM})) \times \widehat{\mathrm{Var}}(W)(\mathrm{OPEN}).$$

These quantities were then used to estimate the deattenuation factor (λ) and $\mathrm{Corr}(x,Z)$ to be expected in the OPEN data, following the methods of Spiegelman et al. (23).

In the Appendix, we prove that the correlation between the unobserved true exposure and the FFQ estimated from models 1 and 2 is the same regardless of whether the biomarker or the 24-hour diet recall is assumed to be an unbiased method of assessment, even when the errors in this measure are assumed to be correlated with the FFQs. In addition, in the Appendix we derive the least restrictive models for both the biomarker and the diet recalls under which the correlation coefficient for correlation between the truth and the FFQ can be validly estimated.

RESULTS

Baseline characteristics

Table 1 shows the baseline characteristics of the study populations. Overall, in the AMPM Study, participants were

Table 2. Mean Nutrient Intakes for Each Dietary Measure in the Observing Protein and Energy Nutrition Study (1999–2000) and the US Department of Agriculture Automated Multiple-Pass Method Study (2002–2003)

	OPEN Study[a]		AMPM Study[a,b]	
	No. of Participants	Mean (SD)	No. of Participants	Mean (SD)
Calories, kcal/day				
FFQ no. 1	482	1,905 (870)	469	1,885 (651)
FFQ no. 2	480	1,771 (808)	51	1,705 (488)
24-hour diet recall no. 1	484	2,360 (818)	51	2,469 (819)
24-hour diet recall no. 2	482	2,273 (853)	52	2,173 (600)
Biomarker no. 1	451	2,627 (556)	462	2,555 (586)
Biomarker no. 2	24	2,621 (598)	50	2,734 (647)
Protein, g/day				
FFQ no. 1	482	72 (33)	469	81 (29)
FFQ no. 2	480	68 (33)	51	75 (24)
24-hour diet recall no. 1	484	88 (36)	51	97 (41)
24-hour diet recall no. 2	482	88 (42)	52	86 (31)
Biomarker no. 1	366	95 (31)	470	86 (33)
Biomarker no. 2	352	96 (30)	51	91 (30)
Protein density, % of energy				
FFQ no. 1	482	15.3 (3.2)	469	17.5 (3.3)
FFQ no. 2	480	15.4 (3.0)	51	17.7 (3.9)
24-hour diet recall no. 1	484	15.1 (4.3)	51	15.6 (3.0)
24-hour diet recall no. 2	482	15.6 (4.7)	52	15.8 (3.3)
Biomarker no. 1	340	14.7 (3.8)	461	13.8 (5.1)
Biomarker no. 2	329[c]	14.6 (3.6)	49	13.8 (4.1)

Abbreviations: AMPM, Automated Multiple-Pass Method; FFQ, food frequency questionnaire; OPEN, Observing Protein and Energy Nutrition; SD, standard deviation.

[a] For the OPEN Study, the sample size was 484 for the main validation study and 24 for the substudy. For the AMPM Study, the sample size was 471 for the main validation study and 52 for the substudy. Slight variations from these numbers in the tables are due to missing data.

[b] For the AMPM Study, there were 6 24-hour diet recalls in total (24-hour diet recall no. 1 represents the mean of 3 24-hour diet recalls taken over a 2-week period, and 24-hour diet recall no. 2 represents the mean of 3 24-hour diet recalls taken over a second 2-week period). Each urinary nitrogen measurement is the mean of 2 24-hour urinary nitrogen measurements taken over a 2-week period.

[c] For participants with only 1 doubly labeled water measurement, we used the first doubly labeled water measurement to adjust the first and second urinary nitrogen measurements to create 2 replicated measures of protein density.

slightly younger and more likely to be female and nonwhite than participants in the OPEN Study. The substudy participants were, in general, similar to the participants in the main validation study for both OPEN and AMPM.

Distribution of nutrient intakes

The mean values and standard deviations for the nutrient variables are presented in Table 2 by method of measurement. Mean numbers of calories decreased between the first and second FFQs and between the first and second 24-hour diet recalls for both studies. Mean protein intake decreased between the first and second FFQs for both studies and between the first and second 24-hour diet recalls for the AMPM Study. For protein density, the differences between the methods of measurement were small.

Intraclass correlation coefficients

The ICCs (ρ_I) (with a superscript sometimes being added to indicate which dietary assessment method is being referred to) for the repeated dietary measurements are presented in Table 3. For calories, ρ_I was similar for both the 24-hour diet recall and the FFQ for OPEN (1 24-hour diet recall at an interval of 3 months and a 3-month interval for the FFQ) and for AMPM (mean of 3 24-hour diet recalls and a 1-year interval for the FFQ and 24-hour diet recalls). The ρ_I^W for calories measured by the biomarker was much higher

Table 3. Intraclass Correlation Coefficients for Each Dietary Measure in the Observing Protein and Energy Nutrition Study (1999–2000) and the Automated Multiple-Pass Method Study (2002–2003)

	OPEN Study[a]				AMPM Study[a]			
	No. of Participants	Time Interval[b]	ρ_I	95% CI	No. of Participants	Time Interval[b]	ρ_I	95% CI
Calories, kcal/day[c]								
FFQ	479	3 months	0.76	0.73, 0.80	51	8 months	0.68	0.54, 0.80
24-hour diet recall	482	3 months	0.40	0.33, 0.48	51	16 months	0.46	0.27, 0.67
Biomarker	24	~14 days	0.91	0.84, 0.95	50	16 months	0.43	0.25, 0.64
Protein, g/day[c]								
FFQ	479	3 months	0.76	0.72, 0.80	51	8 months	0.59	0.42, 0.74
24-hour diet recall	482	3 months	0.32	0.25, 0.41	51	16 months	0.55	0.37, 0.73
Biomarker	297	~14 days	0.60	0.53, 0.67	51	16 months	0.73	0.60, 0.82
Protein density, % of energy[c]								
FFQ	479	3 months	0.69	0.64, 0.74	51	8 months	0.54	0.37, 0.70
24-hour diet recall	482	3 months	0.28	0.20, 0.36	51	16 months	0.41	0.21, 0.64
Biomarker	277[d]	~14 days	0.54	0.46, 0.62	49	16 months	0.54	0.36, 0.72

Abbreviations: AMPM, Automated Multiple-Pass Method; CI, confidence interval; FFQ, food frequency questionnaire; OPEN, Observing Protein and Energy Nutrition.

[a] For the OPEN Study, the sample size was 484 for the main validation study and 24 for the substudy. For the AMPM Study, the sample size was 471 for the main validation study and 52 for the substudy. Slight variations from these numbers in the tables are due to missing data.

[b] Time interval between the first and second measurements taken in the same person.

[c] All nutrient values were log (ln)-transformed. All results were adjusted for sex.

[d] For participants with only 1 doubly labeled water measurement, we used the first doubly labeled water measurement to adjust the first and second urinary nitrogen measurements to create 2 replicated measures of protein density.

for OPEN ($\rho_I^W = 0.91$), with a short time between measurements, than for AMPM ($\rho_I^W = 0.43$), over an interval of approximately 16 months. For protein density, the ρ_I for the FFQ was higher for OPEN ($\rho_I = 0.69$) than for AMPM ($\rho_I = 0.54$). For the 24-hour diet recall, the ICC was higher for AMPM ($\rho_I = 0.41$) than for OPEN ($\rho_I = 0.28$), and for the biomarker, ρ_I^W was the same between the 2 study populations ($\rho_I^W = 0.54$ for both). In OPEN, within-person variation in protein density could be observed only in the measurement of protein, except among the 24 OPEN participants who had a second DLW measured 2 weeks after the first.

Comparison of results from the standard measurement error model versus the new models (models 1 and 2)

Table 4 presents a comparison of the validity correlation coefficients for correlation between the FFQ and the 24-hour diet recall and the unobserved true values, Corr(x,Z), estimated from the 2 models for AMPM and OPEN (23, 26). In the AMPM Study, the deattenuated FFQ-versus-24-hour diet recall correlation ($\rho = 0.51$) for protein density was the same as the deattenuated FFQ-versus-biomarker correlation ($\rho = 0.49$). In the OPEN Study, using the estimates of within-person variation in DLW from the AMPM ICC as described above, the deattenuated correlation between the FFQ and the 24-hour diet recall for protein density was somewhat higher ($\rho = 0.55$) than the correlation between the FFQ and the biomarker ($\rho = 0.37$). The correlation coefficients of the FFQ and the unobserved truth, calculated using models 1 and 2 with DLW variance components obtained using the AMPM DLW ICCs, were similar to those calculated as deattenuated Pearson correlation coefficients. This is not surprising, since the results given in the table use the estimator derived in the Appendix under model 2, an alternate consistent estimator of the same quantity as the deattenuated Pearson correlation coefficient.

Table 5 presents a comparison of the deattenuation factors (λ) calculated from the 2 models in AMPM and OPEN. Overall, the results parallel those for the correlation coefficients presented in Table 4. When comparing results from model 1, which assumes that the 24-hour diet recall is the unbiased method of assessment and the 2 surrogates are the FFQ and the biomarker, with results from model 2, which assumes that the biomarker is the unbiased method of assessment and the 2 surrogates are the FFQ and the 24-hour diet recall, the deattenuation factors for calories and protein were similar. However, for protein density, the deattenuation coefficients were higher for the model which assumed that the biomarker was the gold standard.

Correlation of errors

In the OPEN Study, using the estimates of within-person variation in DLW from the AMPM ICC as described above, the correlation between subject-specific errors in the FFQ and the 24-hour diet recall/diet records, Corr(r,s), was 0.31 (95% confidence interval (CI): 0.13, 0.47) for calories, 0.25 (95% CI: 0.09, 0.40) for protein, and 0.59 (95% CI: 0.34, 0.76) for protein density. In AMPM, the correlation between

Table 4. Comparison of Correlation Coefficients With Food Frequency Questionnaire No. 2 Across Models, Observing Protein and Energy Nutrition Study (1999–2000) and US Department of Agriculture Automated Multiple-Pass Method Study (2002–2003)

Nutrient and Method	Assessment Method Assumed To Be Unbiased	OPEN Study[a]			AMPM Study[a]		
		No. of Participants	ρ	95% CI	No. of Participants	ρ	95% CI
Calories, kcal/day[b]							
Pearson	24-hour diet recall	480	0.28	0.20, 0.36	50	0.21	−0.07, 0.47
Pearson, deattenuated (21)	24-hour diet recall	480	0.36	0.25, 0.46	50	0.27	−0.08, 0.56
Pearson	Biomarker	24	0.39	−0.02, 0.69	49	−0.20	−0.46, 0.09
Pearson, deattenuated (21)	Biomarker	24	0.42	0.005, 0.71	49	−0.26	−0.55, 0.10
Model[c]	Either measure[d]	482	0.25	0.10, 0.38	469	−0.02	−0.19, 0.15
Protein, g/day[b]							
Pearson	24-hour diet recall	480	0.23	0.14, 0.31	50	0.37	0.10, 0.59
Pearson, deattenuated (21)	24-hour diet recall	480	0.30	0.18, 0.41	50	0.45	0.11, 0.69
Pearson	Biomarker	295	0.29	0.18, 0.39	50	0.35	0.08, 0.58
Pearson, deattenuated (21)	Biomarker	295	0.34	0.21, 0.46	50	0.40	0.09, 0.64
Model[c]	Either measure[d]	482	0.30	0.20, 0.39	470	0.27	0.14, 0.38
Protein density, % of energy[b,e]							
Pearson	24-hour diet recall	480	0.41	0.33, 0.48	50	0.40	0.14, 0.61
Pearson, deattenuated (21)	24-hour diet recall	480	0.55	0.44, 0.65	50	0.51	0.15, 0.75
Pearson	Biomarker	276	0.30	0.19, 0.41	48	0.40	0.13, 0.62
Pearson, deattenuated (21)	Biomarker	276	0.37	0.23, 0.49	48	0.49	0.15, 0.72
Model[c]	Either measure[d]	482	0.36	0.24, 0.47	469	0.42	0.25, 0.56

Abbreviations: AMPM, Automated Multiple-Pass Method; CI, confidence interval; OPEN, Observing Protein and Energy Nutrition.

[a] For the OPEN Study, the sample size was 484 for the main validation study and 24 for the substudy. For the AMPM Study, the sample size was 471 for the main validation study and 52 for the substudy. Slight variations from these numbers in the tables are due to missing data.

[b] All nutrient values were log (ln)-transformed. All results were adjusted for sex.

[c] Results for the OPEN Study used estimates of the within-person variability for each biomarker based on the intraclass correlation coefficient obtained from the AMPM Study. Models 1 and 2 used all available data on each measurement.

[d] Results were identical regardless of whether the 24-hour diet recall or the biomarker was considered the gold standard (see Appendix).

[e] For participants with only 1 doubly labeled water measurement, we used the first doubly labeled water measurement to adjust the first and second urinary nitrogen measurements to create 2 replicated measures of protein density.

subject-specific errors in the FFQ and the 24-hour diet recall was 0.38 (95% CI: −0.04, 0.69) for calories and 0.35 (95% CI: −0.29, 0.77) for protein. The variance of the subject-specific error term for protein density from the 24-hour diet recall was estimated to be zero, suggesting no subject-specific errors in protein density from the 24-hour diet recall in this study. Hence, there could be no correlation between these error terms.

DISCUSSION

The objective of this study was to assess the impact of random and correlated errors on the estimated validity correlations between FFQs and 24-hour diet recalls and on the estimated deattenuation factor (λ), which is used for measurement error correction. The findings of our study are 4-fold. First, substantial within-person variation existed in measures of DLW and UN over a period of approximately 16 months. Second, we proved mathematically that failure to adequately measure within-person variation in biomarkers can lead to overestimation of the correlation between the errors of the FFQ and the 24-hour diet recall. Third, we found in AMPM that the estimates of FFQ validity and of the deattenuation factor for energy-adjusted protein intake were similar regardless of whether the 24-hour diet recall or the biomarker was assumed to be the unbiased method of assessment, and regardless of whether these more detailed statistical methods were used or the standard regression calibration method was applied. It is unclear whether this observation was supported in the OPEN data because of the wide confidence intervals; larger validation studies will be needed to answer this question. Fourth, we show in the Appendix that when accounting for correlated errors, the correlation coefficient, $\mathrm{Corr}(x,Z)$, is the same regardless of whether the 24-hour diet recall or the biomarker is assumed to be the gold standard.

Two recent papers have derived deattenuation factors for measurement error models similar to those considered in this paper, under various assumptions (19, 24). As expected, estimates of the deattenuation factor improved after adjusting for total energy intake using the protein density measure, compared with results previously reported for total energy intake and protein intake themselves. Since energy-adjusted

Table 5. Comparison of Deattenuation Factors (λ) Across Models in the Observing Protein and Energy Nutrition Study (1999–2000) and the US Department of Agriculture Automated Multiple-Pass Method Study (2002–2003)

Nutrient and Method	Assessment Method Assumed To Be Unbiased	Surrogate Measure	OPEN Study		AMPM Study	
			λ	95% CI	λ	95% CI
Calories, kcal/day[a]						
Regression of assessment method assumed to be unbiased on the surrogate measure	24-hour diet recall	FFQ no. 2	0.19	0.14, 0.25	0.16	−0.05, 0.37
	Biomarker	FFQ no. 2	0.16	−0.01, 0.34	−0.09	−0.23, 0.04
	Biomarker	24-hour diet recall	0.25	−0.07, 0.57	0.18	0.003, 0.35
Model[b]	24-hour diet recall	FFQ no. 2	0.08	0.02, 0.14	−0.01	−0.05, 0.04
Model[b]	Biomarker	FFQ no. 2	0.07	0.03, 0.11	−0.01	−0.06, 0.05
Protein, g/day[a]						
Regression of assessment method assumed to be unbiased on the surrogate measure	24-hour diet recall	FFQ no. 2	0.18	0.11, 0.25	0.31	0.08, 0.53
	Biomarker	FFQ no. 2	0.18	0.11, 0.24	0.30	0.07, 0.53
	Biomarker	24-hour diet recall	0.25	0.18, 0.33	0.50	0.25, 0.75
Model[b]	24-hour diet recall	FFQ no. 2	0.09	0.04, 0.13	0.12	0.03, 0.20
Model[b]	Biomarker	FFQ no. 2	0.16	0.10, 0.21	0.20	0.10, 0.29
Protein density, % of energy[a,c]						
Regression of assessment method assumed to be unbiased on the surrogate measure	24-hour diet recall	FFQ no. 2	0.49	0.39, 0.59	0.32	0.11, 0.54
	Biomarker	FFQ no. 2	0.36	0.23, 0.50	0.54	0.17, 0.90
	Biomarker	24-hour diet recall	0.35	0.24, 0.47	1.11	0.74, 1.47
Model[b]	24-hour diet recall	FFQ no. 2	0.15	0.07, 0.24	0.30	0.16, 0.44
Model[b]	Biomarker	FFQ no. 2	0.33	0.21, 0.45	0.52	0.35, 0.68

Abbreviations: AMPM, Automated Multiple-Pass Method; CI, confidence interval; FFQ, food frequency questionnaire; OPEN, Observing Protein and Energy Nutrition.

[a] All nutrient values were log (ln)-transformed. All results were adjusted for sex.

[b] Results for the OPEN Study used estimates of the within-person variability for each biomarker based on the intraclass correlation coefficient obtained from the AMPM Study.

[c] For participants with only 1 doubly labeled water measurement, we used the first doubly labeled water measurement to adjust the first and second urinary nitrogen measurements to create 2 replicated measures of protein density.

nutrient data are of the greatest interest in epidemiologic research, we focused on protein density here as well (25). In the original OPEN Study analysis (19), under the assumption that the 24-hour diet recall was the unbiased measure, the deattenuation factor for protein density was 0.41 for men and 0.50 for women. When assuming that the biomarker was the unbiased measure, the original OPEN Study analysis found deattenuation factors for protein density of 0.40 for men and 0.32 for women (19). In the Women's Health Initiative, assuming that the biomarker was the unbiased method of assessment, the deattenuation factor for protein was 0.44 (24). In comparison, in our reanalysis of the OPEN data together with data on within-person variation from AMPM, we obtained an estimate of 0.15 when the 24-hour diet recall was assumed to be the unbiased measure and 0.33 when the biomarker was assumed to be the unbiased measure. For the AMPM Study, we obtained estimates of 0.42 when the 24-hour diet recall was assumed to be the unbiased measure and 0.52 when the biomarker was assumed to be the unbiased measure. The deattenuation factors obtained from the simple regression of the FFQ on the 24-hour diet recall for energy were 0.49 for OPEN and 0.32 for AMPM. The differences in estimates between the original OPEN Study analysis and our reanalysis are due not only to the use of within-person variation estimates from AMPM but also to the use of different measurement error models. Although the biomarkers for energy and protein have been validated in small feeding studies, it is presently unknown how well the constants developed for the conversion equations derived from these experiments apply more broadly to older, overweight, and nonwhite adults (27, 28). Nevertheless, if the correlations are as uniformly high as may appear from these data, an explanation for the difference in the results obtained from model 1 versus model 2 is that model 1 is incorrect.

The correlation in errors between the FFQ and the 24-hour diet recall appeared to be greater for the OPEN Study. It is notable that the FFQs used in the 2 studies were different; in the OPEN Study, the FFQ used questions on serving sizes for each food, whereas the FFQ in the AMPM Study used prespecified units. Concern has been raised that the similarity in assessments of serving sizes for the OPEN FFQ and the 24-hour diet recall might lead to greater correlations in errors and thus tend to overstate the validity of that FFQ (29); this may have accounted for the differences in findings between the OPEN and AMPM studies. Although both of the validation studies considered in this paper used the 24-hour diet recall as the comparison

measure of intake, in many validation studies weighed diet records have been used instead to minimize correlated errors, because the cognitive processes involved in the collection of weighted diet records are different from those used to complete FFQs (16). Thus, the comparability of results from the OPEN and AMPM studies with results that would be obtained from validation studies using weighed diet records is unknown.

In the AMPM Study, 8%–15% of the para-amino benzoic acid recovery was greater than 110%. Problems with apparent underrecovery in 5% of the 24-hour urine samples were previously observed in the Women's Health Initiative (24), where it was found that the primary findings reported were not sensitive to the occurrence of possibly problematic collections.

One of the major limitations of this analysis is that, in the OPEN Study, the interval between replicates of DLW may have been unrealistic for assessing within-person variation and the number of participants with replicate measurements was small because of the very high cost of DLW measurements, making the estimates unstable. Another limitation is that we needed to apply the measure of within-person variation in the biomarker from the AMPM Study to the OPEN Study population. The assumption that the proportion of within-person variation observed in the biomarkers in AMPM was the same as that which would be observed in OPEN may not strictly hold, and we were unable to validate this assumption. Additionally, we could not determine whether differences in estimation of the correlation coefficient for protein density when within-person variation in the biomarker was taken into account were due to true correction of the variation via adjustment for repeated measures of DLW from AMPM or to inadequate use of the parameters estimated from the AMPM Study. Overall, our sample sizes were small, especially for persons who completed the repeat measurements of diet. For the AMPM Study, we only had data on approximately 50 persons who completed the diet recalls.

A further limitation is that only 1 energy-adjusted nutrient was studied; it will be important in the future to conduct similar validation studies with other nutrients for which biomarkers exist. Because the magnitudes of within-person variation differ among nutrients, the findings related to protein may not be generalizable to other dietary factors. As is shown by model 1 in the article by Spiegelman et al. (23) and in the Appendix, the biomarkers need not be unbiased for much useful information to be obtained about λ and Corr(x,Z). Additionally, in the OPEN Study, only 1 24-hour diet recall was completed, while in the AMPM Study, 3 24-hour diet recalls were completed at each time point. However, since the 3 24-hour diet recalls in the AMPM Study were completed over a narrow time period of 2 weeks, it is unlikely that they represented a significantly better measure of long-term diet than the single 24-hour diet recall measure in the OPEN Study. The time period between the repeat measures of 24-hour diet recalls and FFQs was different for the OPEN (approximately 3 months) and AMPM (approximately 16 months) studies, and this should be taken into account when interpreting the results. Theoretically, this should have led to higher correlations in the OPEN Study than in AMPM, although empirically this was not generally the case.

In conclusion, in this analysis, substantial within-person variation in biomarker measurements of protein and calories existed over a period of approximately 16 months. When accounting for within-person variation in the assessment method that is assumed to be unbiased, the validity of energy-adjusted protein intake from the FFQ does not seem to be seriously overestimated by comparison with the 24-hour diet recall instead of the biomarker. Estimates of validity for energy-adjusted protein assuming that the biomarkers are the unbiased method of assessment and allowing for correlated errors in the same FFQ and 24-hour diet recalls are similar to those previously reported using diet records as the unbiased measure (17, 22). It appears that analyses that adjust for measurement error using standard methods will often be substantially less biased than analyses that ignore measurement error, even when biomarkers do not exist or are otherwise infeasible to include in validation studies.

ACKNOWLEDGMENTS

Author affiliations: Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts (Donna Spiegelman, Sarah Rosner Preis, Walter C. Willett); Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts (Donna Spiegelman); Tianjin University of Finance and Economics, Tianjin, China (Barbara Bojuan Zhao); Beltsville Human Nutrition Research Center, US Department of Agriculture, Beltsville, Maryland (David J. Baer, Alanna Moshfegh); Department of Nutrition, Harvard School of Public Health, Boston, Massachusetts (Walter C. Willett); and Channing Laboratory, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts (Walter C. Willett).

This work was supported by National Institutes of Health grant R01 CA50597.

Conflict of interest: none declared.

REFERENCES

1. Byers T. Food frequency dietary assessment: how bad is good enough? *Am J Epidemiol*. 2001;154(12):1087–1088.
2. Kristal AR, Peters U, Potter JD. Is it time to abandon the food frequency questionnaire? *Cancer Epidemiol Biomarkers Prev*. 2005;14(12):2826–2828.
3. Schatzkin A, Kipnis V. Could exposure assessment problems give us wrong answers to nutrition and cancer questions? *J Natl Cancer Inst*. 2004;96(21):1564–1565.
4. Willett WC, Hu FB. Not the time to abandon the food frequency questionnaire: point. *Cancer Epidemiol Biomarkers Prev*. 2006;15(10):1757–1758.
5. Willett W. Reproducibility and validity of food-frequency questionnaires. In: *Nutritional Epidemiology*. New York, NY: Oxford University Press; 1998:101–147.
6. 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(6):642–651.
7. Spiegelman D, Schneeweiss S, McDermott A. Measurement error correction for logistic regression models with an ''alloyed gold standard.'' *Am J Epidemiol*. 1997;145(2):184–196.
8. Kumanyika SK, Mauger D, Mitchell DC, et al. Relative validity of food frequency questionnaire nutrient estimates in the Black Women's Health Study. *Ann Epidemiol*. 2003;13(2):111–118.
9. Salvini S, Hunter DJ, Sampson L, et al. Food-based validation of a dietary questionnaire: the effects of week-to-week variation in food consumption. *Int J Epidemiol*. 1989;18(4):858–867.
10. Feskanich D, Marshall J, Rimm EB, et al. Simulated validation of a brief food frequency questionnaire. *Ann Epidemiol*. 1994;4(3):181–187.
11. Feskanich D, Rimm EB, Giovannucci EL, et al. Reproducibility and validity of food intake measurements from a semiquantitative food frequency questionnaire. *J Am Diet Assoc.* 1993;93(7):790–796.
12. Goldbohm RA, van den Brandt PA, Brants HA, et al. Validation of a dietary questionnaire used in a large-scale prospective cohort study on diet and cancer. *Eur J Clin Nutr.* 1994; 48(4):253–265.
13. Pietinen P, Hartman AM, Haapa E, et al. Reproducibility and validity of dietary assessment instruments. II. A qualitative food frequency questionnaire. *Am J Epidemiol*. 1988;128(3):667–676.
14. Tsubono Y, Kobayashi M, Sasaki S, et al. Validity and reproducibility of a self-administered food frequency questionnaire used in the baseline survey of the JPHC Study Cohort I. *J Epidemiol*. 2003;13(suppl 1):S125–S133.
15. Tsugane S, Kobayashi M, Sasaki S. Validity of the self-administered food frequency questionnaire used in the 5-year follow-up survey of the JPHC Study Cohort I: comparison with dietary records for main nutrients. *J Epidemiol*. 2003; 13(suppl 1):S51–S56.
16. Willett W. *Nutritional Epidemiology.* 2nd ed. New York, NY: Oxford University Press; 1998.
17. 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.
18. 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.
19. Kipnis V, Subar AF, Midthune D, et al. Structure of dietary measurement error: results of the OPEN biomarker study. *Am J Epidemiol*. 2003;158(1):14–21.
20. 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.
21. 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.
22. 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.
23. Spiegelman D, Zhao B, Kim J. Correlated errors in biased surrogates: study designs and methods for measurement error correction. *Stat Med*. 2005;24(11):1657–1682.
24. 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.
25. Willett W, Stampfer MJ. Total energy intake: implications for epidemiologic analyses. *Am J Epidemiol*. 1986;124(1):17–27.
26. 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.
27. Schoeller DA. Measurement of energy expenditure in free-living humans by using doubly labeled water. *J Nutr.* 1988; 118(11):1278–1289.
28. Bingham SA, Cummings JH. Urine nitrogen as an independent validatory measure of dietary intake: a study of nitrogen balance in individuals consuming their normal diet. *Am J Clin Nutr.* 1985;42(6):1276–1289.
29. Willett W. Invited commentary: OPEN questions. *Am J Epidemiol*. 2003;158(1):22–24.
30. Jain MG, Rohan TE, Soskolne CL, et al. Calibration of the dietary questionnaire for the Canadian Study of Diet, Lifestyle and Health cohort. *Public Health Nutr.* 2003;6(1): 79–86.
31. Munger RG, Folsom AR, Kushi LH, et al. Dietary assessment of older Iowa women with a food frequency questionnaire: nutrient intake, reproducibility, and comparison with 24-hour dietary recall interviews. *Am J Epidemiol*. 1992;136(2): 192–200.
32. Shu XO, Yang G, Jin F, et al. Validity and reproducibility of the food frequency questionnaire used in the Shanghai Women's Health Study. *Eur J Clin Nutr.* 2004;58(1):17–23.
33. Abbey DE, Andress M, Fraser G, et al. Validity and reliability of alternative nutrient indices based on a food frequency questionnaire [abstract]. *Am J Epidemiol*. 1988; 128(suppl):934.
34. Flagg EW, Coates RJ, Calle EE, et al. Validation of the American Cancer Society Cancer Prevention Study II Nutrition Survey cohort food frequency questionnaire. *Epidemiology.* 2000;11(4):462–468.
35. Messerer M, Johansson SE, Wolk A. The validity of questionnaire-based micronutrient intake estimates is increased by including dietary supplement use in Swedish men. *J Nutr.* 2004;134(7):1800–1805.
36. Stram DO, Hankin JH, Wilkens LR, et al. Calibration of the dietary questionnaire for a multiethnic cohort in Hawaii and Los Angeles. *Am J Epidemiol*. 2000;151(4):358–370.
37. 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.

APPENDIX

Below we prove that Corr(x,Z) is the same regardless of whether the biomarker or the diet record or diet recall is assumed to be unbiased, even when the diet record or diet recall errors are correlated with the food frequency questionnaire's.

Instead of assuming, as given in model 1,

$$\begin{aligned} Z_{ij} &= a + bx_i + r_i + \varepsilon_{Zij}, \qquad j = 1,2 \\ X_{ij} &= x_i + s_i + \varepsilon_{Xij}, \qquad j = 1,2 \\ W_{ij} &= c + dx_i + \varepsilon_{Wij}, \qquad j = 1,2, \end{aligned}$$

we now assume that (model 2)

$$Z_{ij} = a + bx_i + r_i + \varepsilon_{ij}, \quad j = 1,2$$
$$X_{ij} = c + dx_i + s_i + \varepsilon_{Xij}, \quad j = 1,2$$
$$W_{ij} = x_i + \varepsilon_{Wij}, \quad j = 1,2,$$

where, in both model 1 and model 2, $\mathrm{Corr}(\varepsilon_{Zij}, \varepsilon_{Xij}) \neq 0$ and X is the 24-hour diet recall, Z is the food frequency questionnaire, and W is the biomarker. Model 2 assumes that the 24-hour diet recall or diet record is biased and the biomarker is unbiased, while model 1 assumes that the biomarker is biased and the diet record or 24-hour diet recall is unbiased. Following the methods of Spiegelman et al. (23), Appendix Table 1 shows the estimable first and second central moments that are available in the $\{Z^2X^2W^2\}$ design, where there are 2 measurements available for at least some of each of the 3 types of measurements, when either model 1 or model 2 is assumed to hold.

Because there are 13 uniquely estimable moments, the 13 parameters of models 1 and 2 can be estimated by substituting estimates of each of the 13 moments for their true values according to the relations given in Appendix Table 2, which were obtained by solving for the relations between the parameters and the moments given in Appendix Table 1.

From Appendix Tables 1 and 2, using basic algebra, it can be derived that, assuming model 1,

$$\mathrm{Corr}(x,Z) = \frac{\mathrm{Cov}(x,Z)}{\sqrt{\mathrm{Var}(x)}\sqrt{\mathrm{Var}(Z)}} = \frac{b\sigma_x}{\sqrt{Var(Z)}} = \theta_{11}/\sqrt{\theta_{13}\theta_1},$$

and the same result is obtained under the assumptions of model 2. The estimate for Corr(x,Z) is obtained by substituting the estimated variances and covariances of which it is a function in the formula given above. Similar results are obtained for Corr(x,X) and Corr(x,W) for the 2 models.

Note that although the 2 models give identical estimates for Corr(x,Z), they do not give identical estimates for the deattenuation factor, λ. When model 1 is assumed to hold, it can be seen that

$$\lambda = \frac{\mathrm{Cov}(x,Z)}{\mathrm{Var}(Z)} = \frac{b\sigma_x^2}{\mathrm{Var}(Z)} = \frac{\frac{\theta_{11}\theta_{12}^2}{\theta_{12}\theta_{13}}}{\theta_1} = \frac{\theta_{11}\theta_{12}}{\theta_{13}\theta_1},$$

while, when model 2 is assumed to hold,

$$\lambda = \frac{\mathrm{Cov}(x,Z)}{\mathrm{Var}(Z)} = \frac{b\sigma_x^2}{\mathrm{Var}(Z)} = \frac{\frac{\theta_{11}}{\theta_{13}}\theta_{13}}{\theta_1} = \frac{\theta_{11}}{\theta_1}.$$

To allow for random within-person variation in the biomarker in our analysis of the OPEN Study, we estimated $\sigma^2_{e_W}$ for the OPEN Study from the intraclass correlation for the biomarker found in the AMPM Study and the overall biomarker variance obtained in the OPEN Study. We then treated this parameter, $\sigma^2_{e_W}$, as fixed. Under model 1,

$$\lambda = \frac{\theta_{11}\theta_{12}}{(\theta_5 - \hat{\sigma}^2_{eW})\theta_1}$$

and

$$\rho_{xZ} = \theta_{11}/\sqrt{(\theta_5 - \hat{\sigma}^2_{eW})\theta_1},$$

and under model 2,

$$\lambda = \frac{\theta_{11}}{\theta_1}$$

as above and

$$\rho_{xZ} = \theta_{11}/\sqrt{(\theta_5 - \hat{\sigma}^2_{eW})\theta_1}$$

Appendix Table 1. Correspondence Between Estimable Moments and Model Parameters in the $\{Z^2X^2W^2\}$ Design

Parameter	Moment	Expression	
		Model 1	Model 2
θ_1	$\mathrm{Var}(Z_{ij})$	$b^2\sigma_x^2 + \sigma_r^2 + \sigma_{e_Z}^2$	$b^2\sigma_x^2 + \sigma_r^2 + \sigma_{e_Z}^2$
θ_2	$\mathrm{Cov}(Z_{ij},Z_{ik})$, $j \neq k$	$b^2\sigma_x^2 + \sigma_r^2$	$b^2\sigma_x^2 + \sigma_r^2$
θ_3	$\mathrm{Var}(X_{ij})$	$\sigma_x^2 + \sigma_s^2 + \sigma_{e_X}^2$	$d^2\sigma_x^2 + \sigma_s^2 + \sigma_{e_X}^2$
θ_4	$\mathrm{Cov}(X_{ij},X_{ik})$, $j \neq k$	$\sigma_x^2 + \sigma_s^2$	$d^2\sigma_x^2 + \sigma_s^2$
θ_5	$\mathrm{Var}(W_{ij})$	$d^2\sigma_x^2 + \sigma_{e_W}^2$	$\sigma_x^2 + \sigma_{e_W}^2$
θ_6	$\mathrm{Cov}(Z_{ij},X_{ij})$	$b\sigma_x^2 + \sigma_{rs} + \sigma_{e_Z,e_X}$	$bd\sigma_x^2 + \sigma_{rs} + \sigma_{e_Z,e_X}$
θ_7	$\mathrm{Cov}(Z_{ij},X_{ik})$, $j \neq k$	$b\sigma_x^2 + \sigma_{rs} + \sigma_{e_Z,e_X}$	$bd\sigma_x^2 + \sigma_{rs}$
θ_8	$E(Z_{ij})$	$a + b\mu_x$	$a + b\mu_x$
θ_9	$E(X_{ij})$	μ_x	$c + d\mu_x$
θ_{10}	$E(W_{ij})$	$c + d\mu_x$	μ_x
θ_{11}	$\mathrm{Cov}(Z_{ij}, W_{ik})\forall j,k$	$bd\sigma_x^2$	$b\sigma_x^2$
θ_{12}	$\mathrm{Cov}(X_{ij}, W_{ik})\forall j,k$	$d\sigma_x^2$	$d\sigma_x^2$
θ_{13}	$\mathrm{Cov}(W_{ij},W_{ik})$, $j \neq k$	$d^2\sigma_x^2$	σ_x^2

Appendix Table 2. Functional Relations Between Estimable Moments and Model Parameters

Model Parameter	Moment Parameters	
	Model 1	Model 2
$\sigma^2_{e_Z}$	$\theta_1 - \theta_2$	$\theta_1 - \theta_2$
$\sigma^2_{e_X}$	$\theta_3 - \theta_4$	$\theta_3 - \theta_4$
$\sigma_{e_Z e_X}$	$\theta_6 - \theta_7$	$\theta_6 - \theta_7$
$\sigma^2_{e_W}$	$\theta_5 - \theta_{13}$	$\theta_5 - \theta_{13}$
b	θ_{11}/θ_{12}	θ_{11}/θ_{13}
σ^2_x	$\theta^2_{12}/\theta_{13}$	θ_{13}
σ^2_s	$\theta_4 - \theta^2_{12}/\theta_{13}$	$\theta_4 - \theta^2_{12}/\theta_{13}$
μ_x	θ_9	θ_{10}
a	$\theta_8 - \theta_{11}\theta_9/\theta_{12}$	$\theta_8 - \theta_{11}\theta_{10}/\theta_{13}$
c	$\theta_{10} - \theta_{13}\theta_9/\theta_{12}$	$\theta_9 - \theta_{12}\theta_{10}/\theta_{13}$
d	θ_{13}/θ_{12}	θ_{12}/θ_{13}
σ^2_r	$\theta_2 - \theta^2_{11}/\theta_{13}$	$\theta_2 - \theta^2_{11}/\theta_{13}$
σ_{rs}	$\theta_7 - \theta_{11}\theta_{12}/\theta_{13}$	$\theta_7 - \theta_{11}\theta_{12}/\theta_{13}$

as in model 1. This explains why, by assuming model 2, Subar et al. (18) did not need replicate biomarker measures to obtain a valid estimate of λ, but under model 1 assumptions, as in the article by Spiegelman et al. (23) and as assumed by many other investigators (8–15, 17, 30–37), replicates are needed.

Interestingly, these results can be extended to a number of variations of the assumptions given by models 1 and 2 about the relation of the 24-hour diet recall (X) and the biomarker (W) to the underlying true exposure. There are 16 possible linear models which can be assumed to describe the relation between x and X, Z, and W, incorporating possible location and scale bias for each measure, with a maximum of 15 parameters. These models are given below, in a schematic manner, as members of the class of models represented most generally by

$$\begin{aligned} Z_{ij} &= a + bx_i + r_i + \varepsilon_{Zij}, \qquad j = 1,2 \\ X_{ij} &= c + dx_i + s_i + \varepsilon_{Xij}, \qquad j = 1,2 \\ W_{ij} &= e + fx_i + \varepsilon_{Wij}, \qquad j = 1,2. \end{aligned}$$

In all that follows, it is assumed as in models 1 and 2 that $Z_{ij} = a + bx_i + r_i + \varepsilon_{Zij},\ j = 1,2$. Under the $\{Z^2X^2W^2\}$ design, 13 unique first and second moments are available. Appendix Table 3 shows which of the 15 parameters are identifiable, given the 16 possible models, including models 1 and 2. It can be seen that the estimates of the parameters σ^2_r, σ^2_s, σ_{rs}, $\sigma^2_{e_Z}$, $\sigma^2_{e_X}$, $\sigma^2_{e_W}$, and $\sigma_{e_Z e_X}$ are the same functions of the moments for all models (models 1–16). Most importantly, it can be seen that ρ_{xZ} is identifiable in all 16 models, even when X, Z, and W all have location and scale bias. In contrast, λ is identifiable only when either d or f is equal to 1—that is, when either X or W has no scale bias. We showed above that if $d = 1$, as in models 1, 3–5, and 10–12,

$$\lambda = \frac{\theta_{11}\theta_{12}}{\theta_{13}\theta_1},$$

and when $f = 1$, as in models 2 and 10–16,

$$\lambda = \frac{\theta_{11}}{\theta_1}.$$

Appendix Table 3. Parameter Identifiability in a General Class of Measurement Error Models With Location and/or Scale Bias

	Model															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Model assumptions																
$c =$	0	c	0	c	c	0	0	c	c	0	0	c	c	0	0	c
$d =$	1	d	1	1	1	d	d	d	d	1	1	1	1	d	d	d
$e =$	e	0	0	0	e	e	0	0	e	e	0	0	e	e	0	e
$f =$	f	1	f	f	f	f	f	f	f	1	1	1	1	1	1	1
Parameter identifiability																
λ	√	√	√	√	√	×	×	×	×	√	√	√	√	√	√	√
ρ_{xZ}, $\rho_{e_X e_Z}$, ρ_{rs}, R_{rZ}, R_{sX}	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
$\sigma^2_{e_Z}$, $\sigma^2_{e_X}$, $\sigma^2_{e_W}$, $\sigma_{e_Z e_X}$, σ^2_r, σ^2_s, σ_{rs}	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
Location μ_x, a, c, e	√	√	√	√	×	×	×	×	×	√	√	√	×	√	√	×
Scale σ^2_x, b, d, f	√	√	√	√	√	×	×	×	×	√	√	√	√	√	√	√
Fully identifiable	√	√	√	√	×	×	×	×	×	√	√	√	×	√	√	×