

## Empirical Evidence of Correlated Biases in Dietary Assessment Instruments and Its Implications

Victor Kipnis,<sup>1</sup> Douglas Midthune,<sup>1</sup> Laurence S. Freedman,<sup>2</sup> Sheila Bingham,<sup>3</sup> Arthur Schatzkin,<sup>4</sup> Amy Subar,<sup>5</sup> and Raymond J. Carroll<sup>6</sup>

Multiple-day food records or 24-hour recalls are currently used as “reference” instruments to calibrate food frequency questionnaires (FFQs) and to adjust findings from nutritional epidemiologic studies for measurement error. The common adjustment is based on the critical requirements that errors in the reference instrument be independent of those in the FFQ and of true intake. When data on urinary nitrogen level, a valid reference biomarker for nitrogen intake, are used, evidence suggests that a dietary report reference instrument does not meet these requirements. In this paper, the authors introduce a new model that includes, for both the FFQ and the dietary report reference instrument, group-specific biases related to true intake and correlated person-specific biases. Data were obtained from a dietary assessment validation study carried out among 160 women at the Dunn Clinical Nutrition Center, Cambridge, United Kingdom, in 1988–1990. Using the biomarker measurements and dietary report measurements from this study, the authors compare the new model with alternative measurement error models proposed in the literature and demonstrate that it provides the best fit to the data. The new model suggests that, for these data, measurement error in the FFQ could lead to a 51% greater attenuation of true nutrient effect and the need for a 2.3 times larger study than would be estimated by the standard approach. The implications of the results for the ability of FFQ-based epidemiologic studies to detect important diet-disease associations are discussed. *Am J Epidemiol* 2001;153:394–403.

biological markers; dietary assessment methods; epidemiologic methods; measurement error; models, statistical; model selection; regression analysis; research design

Scientists have long sought a connection between diet and cancer. A number of large prospective studies have now challenged conventional wisdom, which was derived in large part from international correlation studies and animal experiments, in reporting no association between dietary fat and breast cancer (1) and, most recently, no association between dietary fiber and colorectal cancer (2). These null epidemiologic findings may ultimately be shown to reflect

the truth about these diet-cancer hypotheses. Alternatively, however, the studies themselves may have serious methodological deficiencies.

Usually, in large studies, a relatively inexpensive method of measurement, such as a food frequency questionnaire (FFQ), is employed. Investigators now recognize that errors in the values reported on FFQs can profoundly affect the results and interpretation of nutritional epidemiologic studies (3–5). Dietary measurement error often attenuates (biases toward 1) the estimated disease relative risk and reduces the statistical power to detect an effect. An important relation between diet and disease may therefore be obscured.

Realization of this problem has prompted the integration into large epidemiologic investigations of calibration sub-studies that involve a more intensive but presumably more accurate dietary reporting method, called the “reference” instrument. Typically, the instruments chosen for reference measurements have been multiple-day food records, sometimes with weighed quantities instead of estimated portion sizes, or multiple 24-hour recalls. FFQs have been “validated” against such instruments, and correlations between FFQs and reference instruments, sometimes adjusted for within-person random error in the reference instrument, have been quoted as evidence of FFQ validity (6, 7). Additionally, on the basis of such studies, statistical methods have been employed to adjust FFQ-based relative risks

Received for publication January 31, 2000, and accepted for publication November 28, 2000.

Abbreviations: AIC, Akaike Information Criterion; BIC, Bayes Information Criterion; FFQ, food frequency questionnaire.

<sup>1</sup>Biometry Research Group, Division of Cancer Prevention, National Cancer Institute, Bethesda, MD.

<sup>2</sup>Department of Mathematics, Statistics and Computer Science, Bar Ilan University, Ramat Gan, Israel.

<sup>3</sup>Dunn Human Nutrition Unit, Medical Research Council, Cambridge, United Kingdom.

<sup>4</sup>Nutritional Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD.

<sup>5</sup>Applied Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, MD.

<sup>6</sup>Department of Statistics, Texas A&M University, College Station, TX.

Reprint requests to Dr. Victor Kipnis, National Cancer Institute, Executive Plaza North, Room 344, 6130 Executive Blvd., MSC 7354, Bethesda, MD 20892-7354 (e-mail: victor\_kipnis@nih.gov).

for measurement error (8), using the regression calibration approach.

The correct application of the regression calibration approach relies on the assumptions that errors in the reference instrument are uncorrelated with 1) true intake and 2) errors in the FFQ (9). Throughout this paper, we take these two conditions as requirements for a valid reference instrument.

Recent evidence suggests that these assumptions may be unwarranted for dietary report reference instruments. Studies involving biomarkers, such as doubly labeled water for measuring energy intake and urinary nitrogen for measuring protein intake (10–16), suggest that reports using food records or recalls are biased (on average, towards underreporting) and that individuals may systematically differ in their reporting accuracy. This could mean that all dietary report instruments involve bias at the individual level, although direct evidence for individual macronutrients other than protein is not yet available. Part of the bias may depend on true intake (which manifests itself in what we call group-specific bias), therefore violating the first assumption for a reference instrument. Part of the bias may also be person-specific (defined below in detail) and may correlate with its counterpart in the FFQ, thereby violating the second assumption.

For this reason, Kipnis et al. (9) proposed a new measurement error model that allows for person-specific bias in the dietary report reference instrument as well as in the FFQ. Using sensitivity analysis, they showed that if the correlation between person-specific biases in the FFQ and the reference instrument was 0.3 or greater, the usual adjustment for measurement error in the FFQ would be seriously incorrect. However, the paper presented no empirical evidence that such correlations exist.

In this paper, we present results of a reanalysis of a calibration study conducted in Cambridge, United Kingdom (17–19) that employed urinary nitrogen excretion as a biomarker for assessing nitrogen intake (20) in addition to the conventional dietary instruments. The biomarker measurements allowed us to generalize the model by Kipnis et al. (9) and further explore the structure of measurement error in dietary assessment instruments and its implications for nutritional epidemiology.

## MODELS AND METHODS

### Effect of measurement error

Consider the disease model

$$R(D|T) = \alpha_0 + \alpha_1 T, \quad (1)$$

where  $R(D|T)$  denotes the risk of disease  $D$  on an appropriate scale (e.g., logistic) and  $T$  is the true long term usual intake of a given nutrient, also measured on an appropriate scale. In this analysis, all nutrients were measured on the logarithmic scale. The slope  $\alpha_1$  represents an association between nutrient intake and disease. Let  $Q = T + e_Q$  denote the nutrient intake obtained from an FFQ (also on a loga-

rithmic scale), where the difference between the reported and true intakes,  $e_Q$ , defines measurement error. Note that short term variation in diet is included in  $e_Q$ , as well as systematic and/or random error components resulting from the instrument itself. We assume throughout that error  $e_Q$  is non-differential with respect to disease  $D$ ; that is, reported intake contributes no additional information about disease risk beyond that provided by true intake.

Fitting model 1 to observed intake  $Q$  instead of true intake  $T$  yields a biased estimate  $\tilde{\alpha}_1$  of the exposure effect. To an excellent approximation (21), the expected observed effect is expressed as

$$E(\tilde{\alpha}_1) = \lambda_1 \alpha_1, \quad (2)$$

where the bias factor  $\lambda_1$  is the slope in the linear regression calibration model

$$T = \lambda_0 + \lambda_1 Q + \xi, \quad (3)$$

where  $\xi$  denotes random error.

Although, in principle, when measurement error  $e_Q$  is correlated with true exposure  $T$ ,  $\lambda_1$  could be negative or greater than 1 in magnitude, in nutritional studies  $\lambda_1$  usually lies between 0 and 1 (22) and can be thought of as an attenuation of the true effect  $\alpha_1$ .

Measurement error also leads to loss of statistical power for testing the significance of the disease-exposure association. Assuming that the exposure is approximately normally distributed, the sample size required to reach the requested statistical power for a given exposure effect is proportional to (22)

$$N \propto 1/[\rho^2(Q,T)\sigma_T^2] = 1/(\lambda_1^2\sigma_Q^2), \quad (4)$$

where  $\rho(Q,T)$  is the correlation between the reported and true intakes,  $\sigma_Q^2$  is the variance of the questionnaire-reported intake, and  $\sigma_T^2$  is the variance of true intake. Thus, the asymptotic relative efficiency of the “naive” significance test, compared with one based on true intake, is equal to the squared correlation coefficient  $\rho^2(Q,T)$ .

### Commonly used measurement error adjustment

Following equations 2 and 3, the unbiased (adjusted) effect can be calculated as  $\hat{\lambda}_1^{-1}\tilde{\alpha}_1$ , where  $\hat{\lambda}_1$  is the estimated attenuation factor. Estimation of  $\lambda_1$  usually requires simultaneous evaluation of additional dietary intake measurements made by the reference instrument in a calibration substudy. The common approach in nutritional epidemiology, introduced and made popular by Rosner et al. (8), uses food records/recalls as reference measurements ( $F$ ), assuming that they are unbiased instruments for true long term nutrient intake at the personal level. For person  $i$  and repeat measurement  $j$ , the common model can be expressed as

$$Q_i = T_i + e_{Q_i}, \quad (5)$$

$$F_{ij} = T_i + e_{Fij}, \quad (6)$$

where it is assumed that errors  $e_{Qi}$  and  $e_{Fij}$  satisfy

$$E(e_{Fij}|T_i) = 0, \quad (7)$$

$$\text{Cov}(e_{Fij}, e_{Fij'}) = 0, j \neq j', \quad (8)$$

$$\text{Cov}(e_{Fij}, e_{Qi}) = 0. \quad (9)$$

Note that the assumption in equation 7 assures that  $\text{Cov}(e_{Fij}, T_i) = 0$ .

### The calibration data

The data were obtained from a dietary assessment validation study carried out at the Medical Research Council's Dunn Clinical Nutrition Center, Cambridge, United Kingdom (17). One hundred and sixty women aged 50–65 years were recruited through two general medical practices in Cambridge. Subjects from practice 1 (group 1) were studied from October 1988 to September 1989, and those from practice 2 (group 2) were studied from October 1989 to September 1990. The principal measures for this study were a 4-day weighed food record and two 24-hour urine collections obtained on each of four occasions (seasons) over the course of 1 year. Season 1 was October–January; season 2, February–March; season 3, April–June; and season 4, July–September.

The weighed food record was the primary dietary report instrument of interest. The weighed records were obtained using portable electronic tape-recorded automatic scales that automatically record verbal descriptions and weights of food without revealing the weight to the subject. Each 4-day period included different days chosen to ensure that all days of the week were studied over the year, with an appropriate ratio of weekend days to weekdays.

Urine specimens were checked for completeness with *p*-aminobenzoic acid and were used to calculate urinary nitrogen excretion (23). Since it is estimated that approximately 81 percent of nitrogen intake is excreted through the urine (20), the urinary nitrogen values were adjusted, dividing by 81 percent, to estimate the total nitrogen intake of each individual. Subjects were asked to collect the first 24-hour urine sample on the third or fourth day of their food record procedure and the second sample 3–4 days later.

In this analysis, we studied nitrogen intake (g/day) and analyzed the Oxford FFQ, which is based on the widely used FFQ of Willett et al. (24), modified to accommodate the characteristics of a British diet. Nitrogen in foods is analyzed directly and then converted to dietary protein content using established factors of 5.18–6.38 (25). The FFQ was administered 1 day before the start of the weighed food record in season 3. We used the weighed food record as the dietary report reference instrument and the adjusted urinary nitrogen measurements as the biomarker. Urinary nitrogen has long been used as a critical measure of protein nutriture in nitrogen balance studies (20, 26–39), and adjusted urinary nitrogen appears to provide a marker for nitrogen intake that

is valid as a reference instrument, as defined in the Introduction. (See the Appendix for more details.)

Note that both weighed food records and urinary nitrogen measure intake over a short period of time, while the FFQ assesses diet during the previous year. Therefore, errors in weighed food records and urinary nitrogen may reflect seasonal patterns in food consumption, but FFQ errors should not, in principle, contain seasonality.

In all of our analyses, we applied logarithmic transformation to the data to better approximate normality. Table 1 lists the mean values and variances of the transformed data according to instrument and season.

### Check of standard reference instrument assumptions

As we noted above, it is a requirement that the reference instrument in a calibration study contain only error that is unrelated to true nutrient intake and is independent of error in the FFQ. Here we demonstrate an indirect check of these assumptions for the weighed food record in the Medical Research Council data. A critical assumption in our analysis is that adjusted urinary nitrogen meets the above requirements of a reference instrument for nitrogen intake.

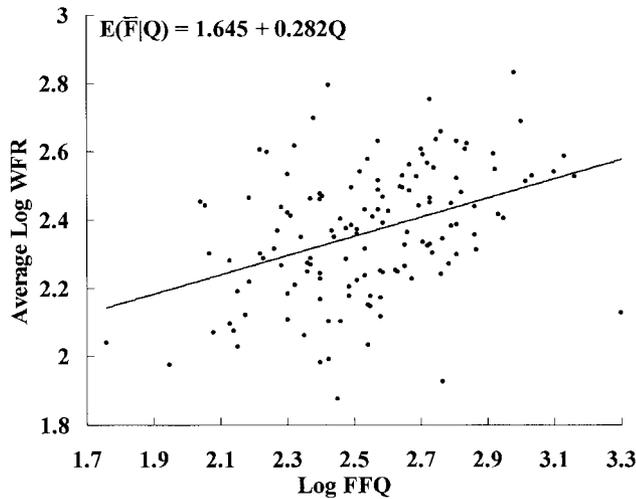
Suppose that the common assumptions (equations 5–9) for a reference instrument hold for the weighed food record. We would then expect that using the common approach (8) with the weighed food record as the reference instrument should lead to nearly the same estimated attenuation as using the urinary nitrogen as the reference instru-

**TABLE 1. Numbers of individuals, mean values, and variances of log-transformed nitrogen intake measurements in the Medical Research Council study\***

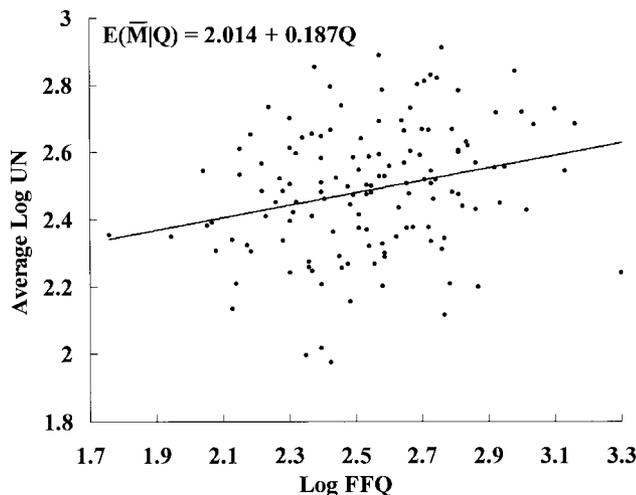
Instrument	No.	Mean	Variance
Food frequency questionnaire	137	2.544	0.0709
Weighed food record			
Season 1	160	2.371	0.0584
Season 2	160	2.380	0.0490
Season 3	160	2.354	0.0502
Season 4	156	2.321	0.0450
Adjusted urinary nitrogen level			
Season 1			
First measurement	117	2.497	0.0547
Second measurement	112	2.476	0.0606
Season 2			
First measurement	112	2.538	0.0425
Second measurement	111	2.523	0.0466
Season 3			
First measurement	116	2.507	0.0517
Second measurement	110	2.483	0.0515
Season 4			
First measurement	122	2.446	0.0469
Second measurement	116	2.446	0.0530

\* Data were obtained from a dietary assessment validation study (17) carried out at the Dunn Human Nutrition Unit, Cambridge, United Kingdom, 1988–1990.

ment. Figures 1 and 2 display scatterplots of averaged weighed food record data versus FFQ data and averaged urinary nitrogen data versus FFQ data, respectively; the slopes of the regression lines give the estimates of the respective attenuation factors. The former method yielded an estimated attenuation factor of 0.282, while the latter estimated it as 0.187; using a statistical test based on their bootstrap distributions, the difference between these two estimates is statistically significant ( $p = 0.022$ ). This



**FIGURE 1.** Scatterplot of log nitrogen intake as measured by averaged values from a dietary report reference instrument  $F$  (weighed food record (WFR)) versus a food frequency questionnaire (FFQ)  $Q$ , with an estimated linear regression line. Data were obtained from a dietary assessment validation study (17) carried out at the Dunn Clinical Nutrition Center, Cambridge, United Kingdom, 1988–1990.



**FIGURE 2.** Scatterplot of log nitrogen intake as measured by the averaged biomarker  $M$  (adjusted urinary nitrogen (UN) excretion) versus a food frequency questionnaire (FFQ)  $Q$ , with an estimated linear regression line. Data were obtained from a dietary assessment validation study (17) carried out at the Dunn Clinical Nutrition Center, Cambridge, United Kingdom, 1988–1990.

important finding means that the attenuation caused by measurement error in the FFQ is in fact more severe than it would appear when using the weighed food record as the reference instrument. If we accept the previously stated assumptions concerning urinary nitrogen, this result suggests that the weighed food record does not satisfy at least one of the two major requirements for a reference instrument—namely, that its error be unrelated to true intake and independent of error in the FFQ.

### A new dietary measurement error model

*Model for the FFQ.* The error in an FFQ is thought likely to include a systematic within-person bias  $b$  that may depend on the individual's true intake  $T$ , as well as within-person variation  $\varepsilon$  (19, 21, 40), so that

$$Q = T + e_Q = T + b + \varepsilon.$$

We approximate the relation between bias  $b$  and true intake  $T$  as the linear regression

$$b = \beta_{Q0} + \beta_{Q1}^* T + r,$$

where  $r$  has zero mean and variance  $\sigma_r^2$  and is independent of  $T$ .  $T$  itself has mean  $\mu_T$  and variance  $\sigma_T^2$ . The component  $\beta_{Q0} + \beta_{Q1}^* T$  is common to all persons with the same true intake and may be called group-specific bias. The second term  $\beta_{Q1}^* T$  can be thought of as arising from correlation between error and true intake. For example, given the social/cultural pressure to follow the “correct” dietary pattern, persons with a low intake of supposedly healthy food may be tempted to overreport their intake, and those with a high intake of supposedly unhealthy food may be tempted to underreport. In this case, as in many other instances,  $\beta_{Q1}^*$  is negative, giving rise to the flattened slope phenomenon in the regression of reported intake on true intake  $E(Q|T) = \beta_{Q0} + (\beta_{Q1}^* + 1)T$ .

The difference  $r$  between within-person bias and its group-specific component varies from person to person and may be determined by personality characteristics such as susceptibility to social/cultural influences. We will call it person-specific bias. Note that this error component is part of within-person systematic error and will be reproduced in repeated measurements on the same individual.

Gathering all of the error components together, we model the FFQ intake  $Q_{ij}$  for individual  $i$  and repeat measurement (season)  $j$  as

$$Q_{ij} = \mu_{Qj} + \beta_{Q0} + \beta_{Q1} T_i + r_i + \varepsilon_{ij}, \quad (10)$$

where  $\beta_{Q1} = \beta_{Q1}^* + 1$ . The term  $\mu_{Qj}$  represents a possible seasonal effect at the population level, a factor that usually improves model fit (41). Similarly, below we use the symbols  $\mu_{Fj}$  and  $\mu_{Mj}$  to represent seasonal effects in reference instrument reports and in marker levels, respectively. Within-person random error  $\varepsilon_{ij}$  has variance  $\sigma_\varepsilon^2$  and is independent of other terms in model (equation) 10.

*Model for the dietary report reference instrument.* As we have argued, we need to allow for systematic group-specific and person-specific biases in dietary report reference instruments. Thus, we now make the same assumptions regarding the error structure for the reference instrument as for the FFQ and use a model which is analogous to that of model 10.

In the Medical Research Council study, each individual  $i$  was requested to provide the weighed food record in each ( $j$ ) of the four seasons. We model these data as

$$F_{ij} = \mu_{Fj} + \beta_{F0} + \beta_{F1}T_i + s_i + u_{ij},$$

$$i = 1, \dots, n, j = 1, 2, 3, 4, \tag{11}$$

where  $\beta_{F0} + \beta_{F1}T_i$  represents group-specific bias and where  $s_i$  and  $u_{ij}$  denote person-specific bias and within-person random error, with variances  $\sigma_s^2$  and  $\sigma_u^2$ , respectively, and are assumed to be independent of each other and of true intake  $T_i$ . As before,  $\mu_{Fj}$  represents a seasonal effect at the population level.

Note that the term  $s_i$  in equation 11 is parallel to the term  $r_i$  in equation 10 for the FFQ. Since the same personality traits can influence both person-specific biases, one may anticipate that the two will have a nonzero correlation  $\rho(r,s)$ .

Because there was only one application of the FFQ in the Medical Research Council study (17), we cannot estimate  $\sigma_\epsilon^2$  and  $\sigma_r^2$  separately, only their sum. Thus, we can estimate the covariance between  $r$  and  $s$  and the correlation between  $r + \epsilon$  and  $s$ , but not the correlation  $\rho(r,s)$ . The correlation between  $r + \epsilon$  and  $s$  will be smaller than  $\rho(r,s)$ , because  $\epsilon$  is independent of  $s$ .

*Model for the biomarker.* As we mentioned above, it is reasonable to assume that adjusted urinary nitrogen has errors that are unrelated to true intake and to errors in dietary assessment instruments. The Medical Research Council study included two repeat urinary nitrogen measurements in each of the four seasons. Letting  $j$  denote season ( $j = 1, 2, 3, 4$ ), as before, and  $k$  denote the repeat measurement within the season ( $k = 1, 2$ ), we write this model as

$$M_{ijk} = \mu_{Mj} + T_i + w_i + v_{ijk}, \tag{12}$$

where 1)  $M_{ijk}$  denotes the  $k$ th repeat of the urinary nitrogen measurement of person  $i$  in season  $j$ ; 2)  $w_i$  and  $v_{ijk}$  denote person-specific bias and within-person random error, with variances  $\sigma_w^2$  and  $\sigma_v^2$ , respectively, and are assumed to be independent of each other and of true intake  $T_i$ ; and 3)  $\mu_{Mj}$  represents a seasonal effect at the population level. It is critical that  $w_i$  is independent of true intake  $T_i$  and of all error components in the dietary report instruments  $Q$  and  $F$ .

As we explain in the Appendix, external evidence suggests that the variance of the person-specific bias,  $w_i$ , is very small relative to the variance of other terms in the model. Therefore, we assume in our main analysis that its variance is actually zero, and we show in the Appendix that our results do not change appreciably when other reasonable values of the variance are used.

Unlike model 10–11 for dietary assessment methods, which is not identifiable without biomarker data (9), model

12 with a specified value for the variance of  $w_i$ , such as zero, is identifiable on its own. Fitting it to the Medical Research Council data supports the assumption that the within-person random errors  $v_{ijk}$  are mutually independent (i.e., they are not correlated within season) and have constant variances within seasons but not between seasons. In particular, season 2 has a different error variance than the other three seasons, which have similar variances, so that, denoting the variance of  $v_{ijk}$  by  $\sigma_{vj}^2$ ,  $\sigma_{v1}^2 = \sigma_{v3}^2 = \sigma_{v4}^2 \neq \sigma_{v2}^2$ .

In contrast, the variances of  $\epsilon_{ij}$  and  $u_{ij}$  are assumed to be constant for all  $i$  and  $j$ ; this assumption is supported by examination of plots of residuals after fitting model 10–12 to the data. The within-person random errors  $\epsilon_{ij}$ ,  $u_{ij}$ , and  $v_{ijk}$  are assumed to be mutually independent, except when the instruments are administered in the same season, in which case seasonal fluctuations in diet are assumed to produce nonzero correlation between  $u_{ij}$  and  $v_{ijk}$ . To verify that FFQ errors were not affected by seasonality, we initially allowed for nonzero correlations between  $\epsilon_{ij}$  and each of the errors  $u_{ij}$  and  $v_{ijk}$  in season 3. As we expected, these correlations were found to be very small and statistically nonsignificant, and we did not include them in the final model.

Model 10–12 involves 20 unknown parameters. From the data, we can estimate 19 unique variances and covariances. These, together with an assumed value for the variance of  $w_i$ , allow us to estimate all of the parameters of the model. In practice, we use the method of maximum likelihood for estimation, which increases efficiency when there are missing values in the data.

**Alternative measurement error models**

Several alternatives to measurement model 10–12 have been proposed in the literature. In table 2, we list six models that are special cases of (and nested within) the more general model 10–12. These include the common model of

**TABLE 2. Six alternative models that are special cases of the new model, model 10–12**

Model	Parameter restrictions
Common model (Rosner et al. (8))	$\beta_{F1} = 1; \sigma_s^2 = 0; \rho(r,s) = 0; \rho(\epsilon,u) = 0$
Freedman et al. (42)	$\beta_{F1} = 1; \sigma_s^2 = 0; \rho(r,s) = 0; \rho(\epsilon,u)$ is nonzero for contemporaneous measures
Kaaks et al. (40)	$\beta_{F1} = 1; \rho(r,s) = 0$
Spiegelman et al. (43)	$\beta_{F1} = 1; \sigma_r^2 = \sigma_s^2 = 0; \rho(\epsilon,u)$ is nonzero for all measures
Kipnis et al. (9)	$\beta_{F1} = 1; \rho(\epsilon,u)$ is nonzero for contemporaneous measures
New (restricted)	$\rho(r,s) = 0; \rho(\epsilon,u)$ is nonzero for contemporaneous measures

Rosner et al. (8) and models proposed by Freedman et al. (42), Kaaks et al. (40), Spiegelman et al. (43), and Kipnis et al. (9). The defining manner in which each model departs from model 10–12 is given in the table. To test the significance of the correlation between person-specific biases in the FFQ and the weighed food record, we also included in the comparison a version of model 10–12 with  $\rho(r,s) = 0$ .

For comparison purposes, we slightly modified the literature-based models by adding the term  $\mu_{F_j}$  to represent a possible seasonal effect in the weighed food record. We also included the urinary nitrogen measurements that were modeled by equation 12.

Plummer and Clayton (19) suggest a quite general model (their model II(c)) that includes our model as a special case. They do not consider person-specific biases but allow group-specific biases to vary in repeat administrations of the same instrument. In addition, within-person random errors are assumed to be correlated, both across repeat administrations of the same instrument and across instruments, with the exception of errors in the biomarker. These are assumed to be correlated across repeat administrations within the same season and with errors in dietary report instruments in the same season but to be independent of measurements taken in different seasons. Moreover, all of the correlations and variances that are assumed to exist are allowed to differ from one another.

Prentice (44) suggested a model similar to that presented by Kipnis et al. (9), except that he explicitly assumed that  $\rho(r,s) = \sigma_s/\sigma_r$  (9). However, all model parameters are allowed to depend on body mass index, and we do not include his model in this comparison.

## MODEL COMPARISON USING MEDICAL RESEARCH COUNCIL DATA

### Model comparison criteria

All models mentioned above were fitted to the Medical Research Council data by the method of maximum likelihood under multivariate normality—a reasonable assumption after the logarithmic transformation—and compared

using three criteria. We first tested the models' goodness of fit by comparing each model with the unstructured (i.e., fully saturated) model using the likelihood ratio  $\chi^2$  test. A model that fits the data should produce a nonsignificant  $p$  value, thereby indicating that it does not explain the data significantly worse than the most general model possible. We also applied the likelihood ratio test to evaluate differences in model fit for nested models. In addition, all models were compared using two standard model selection criteria, the Akaike Information Criterion (AIC) and the Bayes Information Criterion (BIC) (29). These are defined as

$$\text{AIC} = \log(\text{likelihood}) - d$$

and

$$\text{BIC} = \log(\text{likelihood}) - \log(n) \times d/2,$$

where  $d$  is the number of parameters and  $n$  is the sample size. Larger values of AIC and BIC are desirable. Both AIC and BIC penalize more complex models: The "best" models chosen by the BIC tend to be simpler than those chosen by the AIC.

### Model comparison results

The results of model comparison are given in table 3. Ideally, one aims to find a model that passes the goodness-of-fit test, is not significantly different from any more complex model, provides a significantly better fit than all models nested within it, and has the highest AIC and BIC scores among all models. For the Medical Research Council data, model 10–12 emerges as best by these criteria. First, it is one of only four models, together with its two simplified versions and the model of Plummer and Clayton (19), to pass the goodness-of-fit test. Second, it does not fit the data significantly worse than the more general model of Plummer and Clayton. The likelihood ratio  $\chi^2$  statistic comparing the two models is 38.8 ( $38.8 = 1,173.2 - 1,134.4$ ), based on 37 degrees of freedom ( $37 = 56 - 19$ ) ( $p = 0.39$ ). Third, model 10–12 provides a significantly better fit ( $p \leq 0.0011$ ) than

TABLE 3. Results of a model comparison using the Medical Research Council data\*

Model	$-2 \times$ log likelihood	Degrees of freedom†	$p$ value‡	Akaike Information Criterion	Bayes Information Criterion
Unstructured (fully saturated)	-1,222.3	104		507.2	224.7
Plummer and Clayton (19) model II(c)	-1,173.2	56	0.426	530.6	378.5
New model (equations 10–12)	-1,134.4	19	0.393	548.2	496.6
New model restricted to $\rho(r,s) = 0$	-1,123.7	18	0.167	543.9	495.0
Kipnis et al. (9)	-1,122.2	18	0.142	543.1	494.2
Kaaks et al. (40)	-1,112.4	17	0.049	539.2	493.0
Spiegelman et al. (43)	-1,058.0	17	<0.001	512.0	465.8
Freedman et al. (42)	-1,050.1	16	<0.001	509.1	465.6
Common model (Rosner et al. (8))	-1,050.1	15	<0.001	510.1	469.3

\* Data were obtained from a dietary assessment validation study (17) carried out at the Dunn Human Nutrition Unit, Cambridge, United Kingdom, 1988–1990.

† Number of parameters.

‡  $p$  value for goodness-of-fit test relative to the unstructured model.

any model nested in it. For example, comparing it with its version with uncorrelated person-specific biases, the likelihood ratio  $\chi^2$  statistic is 10.7 ( $10.7 = 1,134.4 - 1,123.7$ ), based on 1 degree of freedom ( $1 = 19 - 18$ ), with a  $p$  value of 0.0011.

These results suggest that group- and person-specific biases exist in both the FFQ and the weighed food record, and that these person-specific biases are indeed correlated. With only one FFQ measurement, this correlation cannot be estimated directly, but it is at least 0.35 (the low bound for  $\rho(r,s)$  corresponding to  $\sigma_\varepsilon^2 = 0$ ) and may be considerably higher. For example, if the variance of the person-specific bias is the same for the FFQ and the weighed food record, this correlation is estimated as 0.81.

#### Attenuation of estimated effect and statistical power

Table 4 displays the estimates of the most interesting parameters for model 10–12 and the common model. They include the attenuation factor  $\lambda_1$ , the variance of true intake  $\sigma_T^2$ , the correlation  $\rho(Q,T)$  between the FFQ and true usual intake, and the slopes  $\beta_{Q1}$  and  $\beta_{F1}$  that represent group-specific biases in the FFQ and weighed food record, respectively. For all parameters, except  $\sigma_T^2$ , there are major differences between model 10–12 and the common approach. First, the slope of the regression of the weighed food record on true intake,  $\beta_{F1}$ , assumed to be 1 in the common approach, is estimated as 0.766 in our model, thereby demonstrating the flattened slope phenomenon in the reference instrument. In addition, the common approach suggests that the slope in the regression of the FFQ on true intake,  $\beta_{Q1}$ , is 0.661 and the correlation  $\rho(Q,T)$  between the FFQ and true usual intake is 0.432, while our model estimates them as 0.430 and 0.284, respectively, indicating much less accuracy.

The major parameter controlling the ability to detect disease-nutrient relations using an FFQ is the attenuation factor  $\lambda_1$ . The common approach yields the attenuation factor of 0.282, while our model estimates it as 0.187. Since the true effect of an exposure is calculated as the observed effect divided by the attenuation factor, our model suggests that the true effect would be 51 percent greater than the one estimated by the common approach. There is also a much greater impact on the design of epidemiologic studies. As follows from equation 4, for any two models, the ratio of the sample sizes for the same required statistical power is the same as the squared ratio of their attenuation factors. Thus, our model suggests that the study size based on the common

model should be increased by the factor 2.3 ( $(0.282/0.187)^2 = 2.3$ ); that is, studies would have to be more than twice as large as suggested by the common model in order to maintain nominal power.

#### DISCUSSION

Our purpose has been to propose a statistical framework (model 10–12) for evaluating common dietary assessment reference instruments (multiple-day food records, 24-hour recalls) and to employ this framework to evaluate the weighed food record as a reference instrument for nitrogen intake (which is essentially equivalent to protein intake) using data from the Medical Research Council study (17). We have demonstrated that our model produces the best fit to these data when compared with several other models proposed in the literature:

- Its fit is not significantly different from that of the more complex models that we studied.
- It provides a significantly better fit than the simpler models, which are special cases of it.
- It has the highest values of AIC and BIC, two numerical measures of model fit.

Our statistical framework allows evaluation of two major common assumptions about a dietary report reference instrument: 1) there is no correlation between its measurement error and true intake; and 2) there is no correlation between its measurement error and that of the FFQ. Our results using the Medical Research Council data suggest that both assumptions are violated because of the presence of both group- and person-specific biases in the weighed food record and the correlation of the person-specific bias with that in the FFQ.

The statistical model we used rests on the requirement that the urinary nitrogen marker for nitrogen intake does itself satisfy assumptions 1 and 2 above. Assumption 1 is supported by several studies, documented in the Appendix, that have examined urinary nitrogen under various controlled feeding situations. Assumption 2 is based on the strong intuition that discrepancies between this biomarker measurement and true intake are caused by physiologic factors and therefore will be unrelated to errors in a dietary report instrument.

We have thus demonstrated that, at least for these data, the weighed food record may well be a flawed reference instrument. There still remains the question, Do these flaws

**TABLE 4. Estimated parameters for the new and common models using the Medical Research Council data\***

Model	Attenuation factor ( $\lambda_1$ )	$\sigma_T^2$	$\rho(Q,T)$	$\beta_{Q1}$	$\beta_{F1}$
New	0.187 (0.056)†	0.031 (0.004)	0.284 (0.082)	0.430 (0.129)	0.766 (0.066)
Common	0.282 (0.054)	0.030 (0.004)	0.432 (0.076)	0.661 (0.131)	1

\* Data were obtained from a dietary assessment validation study (17) carried out at the Dunn Human Nutrition Unit, Cambridge, United Kingdom, 1988–1990.

† Numbers in parentheses, standard error.

translate into anything of importance? We believe that they do. As was shown above, using the common approach yields the estimated attenuation factor of 0.282, but it is estimated as 0.187 when using the new model. In addition, the estimated correlation between the FFQ-based nitrogen intake and true intake is 0.432 by the common approach but only 0.284 by the new model. This correlation is used as a measure of the FFQ validity, and its squared value represents the loss in statistical power to test the significance of a disease-exposure association. Thus, for these data, the real effect of measurement error in the FFQ is a greater attenuation (51 percent) and a greater loss of power (52 percent) for testing the true effect than would be estimated by the common approach.

Our estimates of the attenuation factor also indicate that the common approach may lead to unexpectedly underpowered studies. For the Medical Research Council data, our model suggests the need for a study 2.3 times larger than would have been designed had the common approach been used.

In summary, our results suggest that the impact of measurement error in dietary assessment instruments on the design, analysis, and interpretation of nutritional studies may be much greater than has been previously suspected, at least regarding protein intake. Both the attenuation of diet effect and the loss of statistical power in FFQ-based epidemiologic studies may be greater than previously estimated, because of the use of dietary reporting methods as reference instruments. This means that current and past studies may be underpowered and may explain some of the null results that have been found in nutritional epidemiology. There is a need to confirm our results by conducting further studies with biomarkers.

Our paper covers only the analysis of protein intake unadjusted for total energy intake. Further work is needed on the effects of measurement error on the analysis of protein density or energy-adjusted protein intake (6), an approach that is often used in nutrition analyses. This will require simultaneous consideration of both energy intake, using a biomarker such as doubly labeled water (10), and protein intake, using urinary nitrogen excretion. Black et al. (16) reported results from a small study with such data that supported a correlation between underreporting of protein and underreporting of energy, but also higher rates of underreporting of energy than of protein. As was reported previously (45), the effect of measurement error in energy-adjusted models can be more complex than in univariate analysis. Therefore, further studies are needed in which data from questionnaires, dietary report reference instruments, and biomarkers for protein and energy intakes are all collected and analyzed simultaneously to investigate the effects of measurement error on protein density or energy-adjusted protein intake.

## ACKNOWLEDGMENTS

Dr. Raymond Carroll's research was supported by a grant from the National Cancer Institute (CA-57030) and by the

Texas A&M Center for Environmental and Rural Health via a grant from the National Institute of Environmental Health Sciences (P30-E509106).

## REFERENCES

- Hunter DJ, Spiegelman D, Adami H-O, et al. Cohort studies of fat intake and the risk of breast cancer—a pooled analysis. *N Engl J Med* 1996;334:356–61.
- Fuchs CS, Giovannucci EL, Colditz GA, et al. Dietary fiber and the risk of colorectal cancer and adenoma in women. *N Engl J Med* 1999;340:169–76.
- Beaton GH, Milner J, Corey P, et al. Sources of variance in 24-hour dietary recall data: implications for nutrition study design and interpretation. *Am J Clin Nutr* 1979;32:2546–59.
- Freudenheim JL, Marshall JR. The problem of profound mis-measurement and the power of epidemiologic studies of diet and cancer. *Nutr Cancer* 1988;11:243–50.
- Freedman LS, Schatzkin A, Wax J. The impact of dietary measurement error on planning a sample size required in a cohort study. *Am J Epidemiol* 1990;132:1185–95.
- Willett WC. *Nutritional epidemiology*. New York, NY: Oxford University Press, 1990:69–91.
- 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;127:377–86.
- 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;8:1051–69.
- Kipnis V, Carroll RJ, Freedman LS, et al. A new dietary measurement error model and its implications for the estimation of relative risk: application to four calibration studies. *Am J Epidemiol* 1999;150:642–51.
- Bandini LG, Schoeller DA, Cyr HN, et al. Validity of reported energy intake in obese and nonobese adolescents. *Am J Clin Nutr* 1990;52:421–5.
- Livingstone MB, Prentice AM, Strain JJ, et al. Accuracy of weighed dietary records in studies of diet and health. *Br Med J* 1990;300:708–12.
- Heitmann BL. The influence of fatness, weight change, slimming history and other lifestyle variables on diet reporting in Danish men and women aged 35–65 years. *Int J Obes* 1993;17:329–36.
- Heitmann BL, Lissner L. Dietary underreporting by obese individuals—is it specific or non-specific? *Br Med J* 1995;311:986–9.
- Martin LJ, Su W, Jones PJ, et al. Comparison of energy intakes determined by food records and doubly labeled water in women participating in a dietary intervention trial. *Am J Clin Nutr* 1996;63:483–90.
- Sawaya AL, Tucker K, Tsay R, et al. Evaluation of four methods for determining energy intake in young and older women: comparison with doubly labeled water measurements of total energy expenditure. *Am J Clin Nutr* 1996;63:491–9.
- Black AE, Bingham SA, Johansson G, et al. Validation of dietary intakes of protein and energy against 24 urinary N and DLW energy expenditure in middle-aged women, retired men and post-obese subjects: comparisons with validation against presumed energy requirements. *Eur J Clin Nutr* 1997;51:405–13.
- Bingham SA, Gill C, Welch A, et al. Comparison of dietary assessment methods in nutritional epidemiology: weighed food records v. 24 h recalls, food frequency questionnaires and estimated diet records. *Br J Nutr* 1994;72:619–43.
- Bingham SA, Cassidy A, Cole TJ, et al. Validation of weighed records and other methods of dietary assessment using the 24 h nitrogen technique and other biological markers. *Br J Nutr* 1995;73:531–50.

19. Plummer M, Clayton D. Measurement error in dietary assessment: an investigation using covariance structure models. (Parts I and II). *Stat Med* 1993;12:925–48.
20. 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;42:1276–89.
21. Carroll RJ, Ruppert D, Stefanski LA. *Measurement error in nonlinear models*. London, United Kingdom: Chapman and Hall Ltd, 1995.
22. Kaaks R, Riboli E, van Staveren W. Calibration of dietary intake measurements in prospective cohort studies. *Am J Epidemiol* 1995;142:548–56.
23. Bingham S, Cummings JH. The use of 4-aminobenzoic acid as a marker to validate the completeness of 24 h urine collections in man. *Clin Sci* 1983;64:629–35.
24. Willett WC, Sampson L, Stampfer MJ, et al. Reproducibility and validity of a semiquantitative food frequency questionnaire. *Am J Epidemiol* 1985;122:51–65.
25. Matthews DE. Proteins and amino acids. In: Shils ME, Olson JA, Shike M, et al, eds. *Modern nutrition in health and disease*. 9th ed. Baltimore, MD: Williams and Wilkins Company, 1999: 11–48.
26. Campbell WW, Crim MC, Dallal GE, et al. Increased protein requirements in elderly people: new data and retrospective reassessments. *Am J Clin Nutr* 1994;60:501–9.
27. Zanni E, Calloway DH, Zezulka AY. Protein requirements of elderly men. *J Nutr* 1979;109:513–24.
28. Oddoye EA, Margen S. Nitrogen balance studies in humans: long term effect of high nitrogen intake and nitrogen accretion. *J Nutr* 1979;109:363–77.
29. Weller LA, Calloway DH, Margen S. Nitrogen balance of men fed amino acid mixtures based on Rose's requirements, egg white protein, and serum free amino acid patterns. *J Nutr* 1971; 101:1499–508.
30. Bunker VW, Lawson MS, Stansfield MF, et al. Nitrogen balance studies in apparently healthy elderly people and those who are housebound. *Br J Nutr* 1987;57:211–21.
31. Uauy R, Scrimshaw NS, Young VR. Human protein requirements: nitrogen balance response to graded levels of egg protein in elderly men and women. *Am J Clin Nutr* 1978;31:779–85.
32. Castaneda C, Charnley JM, Evans WJ, et al. Elderly women accommodate to a low-protein diet with losses of body cell mass, muscle function, and immune response. *Am J Clin Nutr* 1995;62:30–9.
33. Cheng AH, Gomez A, Bergan JG, et al. Comparative nitrogen balance study between young and aged adults using three levels of protein intake from a combination wheat-soy-milk mixture. *Am J Clin Nutr* 1978;31:12–22.
34. Atinmo T, Mbofung CM, Egun G, et al. Nitrogen balance study in young Nigerian adult males using four levels of protein intake. *Br J Nutr* 1988;60:451–8.
35. Rand WM, Scrimshaw NS, Young VR. Retrospective analysis of data from five long term, metabolic balance studies: implications for understanding dietary nitrogen and energy utilization. *Am J Clin Nutr* 1985;42:1339–50.
36. Tarnopolsky MA, Atkinson SA, MacDougall JD, et al. Evaluation of protein requirements for trained strength athletes. *J Appl Physiol* 1992;73:1986–95.
37. Pannemans DL, Wagenmakers AJ, Westerterp KR, et al. Effect of protein source and quantity metabolism in elderly women. *Am J Clin Nutr* 1998;68:1228–35.
38. Wayler A, Queiroz E, Scrimshaw NS, et al. Nitrogen balance studies in young men to assess the protein quality of an isolated soy protein in relation to meat proteins. *J Nutr* 1983;113: 2485–91.
39. Young VR, Wayler A, Garza C, et al. A long term metabolic balance study in young men to assess the nutritional quality of an isolated soy protein and beef proteins. *Am J Clin Nutr* 1984; 39:8–15.
40. Kaaks R, Riboli E, Esteve J, et al. Estimating the accuracy of dietary questionnaire assessments: validation in terms of structural equation models. *Stat Med* 1994;13:127–42.
41. Landin R, Carroll RJ, Freedman LS. Adjusting for time trends when estimating the relationship between dietary intake obtained from a food frequency questionnaire and true average intake. *Biometrics* 1995;51:169–81.
42. Freedman LS, Carroll RJ, Wax Y. Estimating the relationship between dietary intake obtained from a food frequency questionnaire and true average intake. *Am J Epidemiol* 1991;134: 310–20.
43. Spiegelman D, Schneeweiss S, McDermott A. Measurement error correction for logistic regression models with an alloyed gold standard. *Am J Epidemiol* 1997;145:184–96.
44. Prentice R. Measurement error and results from analytic epidemiology: dietary fat and breast cancer. *J Natl Cancer Inst* 1996;88:1738–47.
45. Kipnis V, Freedman LS, Brown CC, et al. Effect of measurement error on energy-adjustment models in nutritional epidemiology. *Am J Epidemiol* 1997;146:842–55.
46. Millward DJ. Urine nitrogen as an independent validity measure of dietary intake: potential errors due to variation in magnitude and type of protein intake. *Br J Nutr* 1997;77:141–3.
47. Bingham SA. Urine nitrogen as an independent validity measure of protein intake. *Br J Nutr* 1997;77:144–8.
48. Millward DJ, Bowtell JL, Pacy P, et al. Physical activity, protein metabolism and protein requirements. *Proc Nutr Soc* 1994;53:223–40.
49. Millward DJ. The nutritional value of plant-based diets in relation to human amino acid and protein requirements. *Proc Nutr Soc* 1999;58:249–60.

---

## APPENDIX

Nitrogen balance studies require known levels of protein/nitrogen intakes and complete urine collections in addition to either estimation or collection of fecal, sweat, or other miscellaneous losses in order to be valid, and this has been done with varying levels of rigor and oversight. Generally, the goals of such studies have been to assess protein requirements and protein sources.

Among studies with varying levels of controlled conditions in which protein intakes were provided at levels necessary to maintain a positive nitrogen balance (a near-given in diets in developed countries), the long term ratio of urinary nitrogen to dietary nitrogen among individuals is generally within a range of 70–90 percent (20, 26–39). Bingham and Cummings (20) specifically addressed the question of nitrogen output and validation of dietary intakes in a rigorously controlled feeding study of eight adults adhering to their regular diets and found that the mean ratio of urinary nitrogen to dietary nitrogen was 81 percent, with a standard error of 2 percent (range, 78–83 percent). In other well-controlled studies, group means have ranged from 77 percent to 88 percent (26–32). Generally, urinary nitrogen is robust in free-living adults, except when there is inadequate total energy and/or protein intake, inadequate essential amino acid intake, a very high fiber intake, or profuse sweating (46–49). None of these conditions are prevalent in adequately nourished populations, and a range of 70–90 percent represents a realistic range for biologic variability in the ratio of urinary nitrogen to dietary nitrogen that does not depend on age, gender, and source of protein, as long as subjects maintain a positive nitrogen balance. This is supported by different studies that measured this

range in old and young participants and in men and women with soy, egg, meat, or mixed sources of protein in their diets (20, 26–39).

Nevertheless, the ratio of urinary nitrogen to dietary nitrogen does not represent an exact biologic constant and may still include interperson variability, or person-specific bias. Three studies described by Bingham and Cummings (20), Oddoye and Margen (28), and Castaneda et al. (32) and two studies described by Young et al. (39) provided information on within-person variation in the ratio ( $R$ ) of urinary nitrogen to dietary nitrogen and therefore can be analyzed by analysis of variance to estimate and/or test the presence of person-specific bias in the urinary nitrogen biomarker. These five studies represent a valuable subsample of the controlled feeding studies and include both men (20, 28, 39) and women (32), young (28, 39), middle-aged (20), and elderly (32) participants, and a variety of protein sources, including soy protein (39), meat-free protein (32), formula diets (26), beef protein (39), and usual diet (20).

We carried out a meta-analysis of these five studies using a random effects model for ratio  $R$  that included both a random study effect and, nested in it, a random person effect (person-specific bias). The study effect  $\eta$  was very small (variance  $\sigma_{\eta}^2 = 0.0006$ ) and not statistically significant ( $p = 0.21$ ), while the person effect  $w$  was also relatively small (variance  $\sigma_w^2 = 0.0021$ ) but highly statistically significant ( $p = 0.0008$ ). These results provide some evidence that although ratio  $R$  does not seem to depend on age, gender, and source of protein intake, it does contain a small person-specific bias. After we pooled all of the participants from the five studies and fitted a random effects model with a random effect representing person-specific bias, the variance of this bias was estimated as 0.0027 (standard deviation 5.2 percent). The mean long term ratio of urinary nitrogen to dietary nitrogen was estimated as 83.5 percent (standard error 2.3 percent), which agrees well with the cal-

ibration constant of 81 percent suggested by Bingham and Cummings (20). The mean ratio (83.5 percent) and the standard deviation of its person-specific bias (5.2 percent) agree well with the general observation that individual ratios fall between 70 percent and 90 percent.

These results suggest that urinary nitrogen level satisfies both requirements for a reference instrument. The stability of the urinary nitrogen:dietary nitrogen ratio and the relatively low person-specific bias support the essential absence of correlation between errors in adjusted urinary nitrogen and true nitrogen. The relatively low person-specific bias and the fact that the bias is probably physiologically based rather than psychologically based also support the essential absence of correlation between errors in adjusted urinary nitrogen and errors in dietary report instruments.

It is interesting to note that the estimated variation due to person-specific bias in the urinary biomarker for protein intake constitutes only about 10 percent of the estimated variation of true protein intake. Nevertheless, to investigate how this person-specific bias might change the result of our model fit, we conducted a sensitivity analysis by including person-specific bias in the biomarker model and changing its value from  $\sigma_w^2 = 0$  (the value assumed in the main text) to  $\sigma_w^2 = 0.0027$  (the value estimated in this appendix). The results are reported in appendix table 1. The estimated attenuation factor was not affected by the presence of person-specific bias in the biomarker, since this bias does not violate the two major requirements for the reference instrument. Other parameters in the model changed slightly. The estimated variance of true intake was reduced by the variation due to person-specific bias. The estimated correlation between true intake and its FFQ measure was increased by 4.5 percent, and the estimated slopes in the regressions of FFQ and weighed food record on true intake were increased by approximately 10 percent each. However, the general conclusions reached in the paper remain the same.

**APPENDIX TABLE 1. Estimated parameters for the new model, with and without person-specific biases in the urinary biomarker, using the Medical Research Council data\***

Model	Attenuation factor ( $\lambda_1$ )	$\sigma_T^2$	$\rho(Q, T)$	$\beta_{Q1}$	$\beta_{F1}$
$\sigma_w^2 = 0$	0.187 (0.056)†	0.031 (0.004)	0.284 (0.082)	0.430 (0.129)	0.766 (0.066)
$\sigma_w^2 = 0.0027$	0.187 (0.056)	0.028 (0.004)	0.297 (0.085)	0.472 (0.142)	0.839 (0.074)

\* Data were obtained from a dietary assessment validation study (17) carried out at the Dunn Human Nutrition Unit, Cambridge, United Kingdom, 1988–1990.

† Numbers in parentheses, standard error.