 International Journal of Epidemiology 2001;**30**:309–317

Epidemiological assessment of diet: a comparison of a 7-day diary with a food frequency questionnaire using urinary markers of nitrogen, potassium and sodium

NE Day,[a] N McKeown,[b] MY Wong,[c] A Welch[a] and S Bingham[a]

Background Validation studies of dietary instruments developed for epidemiological studies have typically used some form of diet record as the standard for comparison. Recent work suggests that comparison with diet record may overestimate the ability of the epidemiological instrument to measure habitual dietary intake, due to lack of independence of the measurement errors. The degree of regression dilution in estimating diet-disease association may therefore have been correspondingly underestimated. Use of biochemical measures of intake may mitigate the problem. In this paper, we report on the use of urinary measures of intakes of nitrogen, potassium and sodium to compare the performance of a semi-quantitative food frequency questionnaire (FFQ) and a 7-day diet diary (7DD) to estimate average intake of these nutrients over one year.

Methods In all, 179 individuals were asked to complete an FFQ and a 7DD on two occasions separated by approximately 12 months. The individuals were also asked to provide 24-hour urine samples on six occasions over a 6–9-month period, covering the time at which the record FFQ and 7DD were completed. The urine was assayed for nitrogen, potassium and sodium. The protocol was completed by 123 individuals. The data from these individuals were analysed to estimate the covariance structure of the measurement errors of the FFQ, the 7DD and a single 24-hour urine measurement, and to estimate the degree of regression dilution associated with the FFQ and 7DD.

Results The results demonstrated that: (1) the error variances for each of the three nutrients was more than twice as great with the FFQ than the 7DD; (2) there was substantial correlation (0.46–0.58) between the error of both the FFQ and the 7DD completed on different occasions; (3) there was moderate correlation (0.24–0.29) between the error in the FFQ and the error in the 7DD for each nutrient; (4) the correlation between errors in different nutrients was higher for the FFQ (0.77–0.80) than for the 7DD (0.52–0.70).

Conclusions The regression dilution with the FFQ is considerably greater than with the 7DD and also, for the nutrients considered, greater than would be inferred if validation studies were based solely on record or diary type instruments.

Keywords Biological markers; food diaries; nutrition assessment; questionnaires; validity; EPIC-Norfolk

Accepted 1 August 2000

[a] Strangeways Research Laboratory, Institute of Public Health, University of Cambridge, Cambridge CB1 8RN, UK. E-mail: nick.day@srl.cam.ac.uk

[b] Dunn Human Nutrition Unit, Cambridge Institute for Medical Research, Cambridge CB2 2XY, UK.

[c] Department of Mathematics, Hong Kong University of Science and Technology, Hong Kong.

The development of instruments to measure habitual dietary intake in the context of large epidemiological studies has been investigated extensively. It is now widely acknowledged that prospective studies give more reliable estimates of association between diet and disease than retrospective studies.[1] For most

disease endpoints of interest, such prospective studies need to be large with several tens if not hundreds of thousands of individuals recruited to give baseline information. Dietary instruments to be used in studies of this size are clearly constrained by resource considerations. Food frequency questionnaires (FFQ) have been largely the instrument of choice,[2] but increasingly the use of diet diaries is being proposed.[3] No dietary instrument can capture habitual diet with complete accuracy, and methods are available to correct observed diet-disease association for the bias induced by the imprecision of the dietary assessment, the so-called regression dilution bias. Regression dilution bias is an intrinsic aspect of modern quantitative epidemiology. However, the credibility of correcting for regression dilution depends on the magnitude of the correction required. In addition, estimation of the requisite correction factors becomes problematic if the instrument measures habitual diet poorly. Both the potential for bias and the imprecision of estimation increases rapidly as the association between estimated and true dietary intake decreases.[4–6] The performance of the FFQ has been extensively compared with other, more intensive, record-based methods, such as weighed diet records.[2,7] There is increasing evidence, however, that record- or recall-based methods do not satisfy the independence criteria required to act as validation methods.[8–11] Non-independence can lead to substantial overestimation of the capacity of the FFQ to assess habitual diet accurately.[8] Few studies have characterized the performance of FFQ against validated biochemical measures of intake. Twenty-four hour (24-h) urine collections, verified for completeness by the *para*-amino benzoic acid (PABA) method have been shown to give an unbiased, calibrated measure of intake for nitrogen (N),[3] potassium (K)[3] and sodium (Na).[12] In this paper, we report on the results of a validation study performed in association with the UK component based in Norfolk of the European Prospective Investigation of Cancer (EPIC).[13] This cohort study of approximately 25 000 adults, known as EPIC-Norfolk, used both a 7DD and a semi-quantitative FFQ.[14] This validation study compared the performance of these two dietary instruments with 24-h urine measurements of N, K and Na, verified for completeness using PABA. An underlying assumption of this study is that the errors of measurement associated with the biochemical assessment of intake are independent from the errors of measurement of the 7DD and FFQ methods of dietary assessment, provided the urine collection and the completion of the diary do not occur together. This assumption is considered later in Discussion.

Method

Study design

Over an 18-month period, 179 members of the EPIC-Norfolk cohort took part in a validation study. In brief, each individual was asked to complete an FFQ and a 7DD on two occasions: once on their entry into EPIC-Norfolk and on another occasion 18 months later (±3 months). Over a 12-month period which covered the time of completion of the second FFQ and 7DD, six 24-h urine collections were requested, with completeness to be verified by PABA. The time relationship between urine collections and completion of the FFQ and 7DD is displayed in Figure 1. In all, 123 individuals completed the full protocol, and the results presented here refer to these 123 individuals. Nitrogen, K and Na were assayed on all urine samples.[3]

Dietary assessment methods

The 7-day food diary

The participants were asked to record, in as much detail as possible, all food and beverages consumed over a 7-day period. The 7DD included coloured photographs of 17 foods, each with three different portion sizes. Participants could choose which photograph represented their portion size or indicate if they consumed more or less than the amount shown in the photograph. Participants were also allowed to describe their portion size in other measures, such as weights or household units.

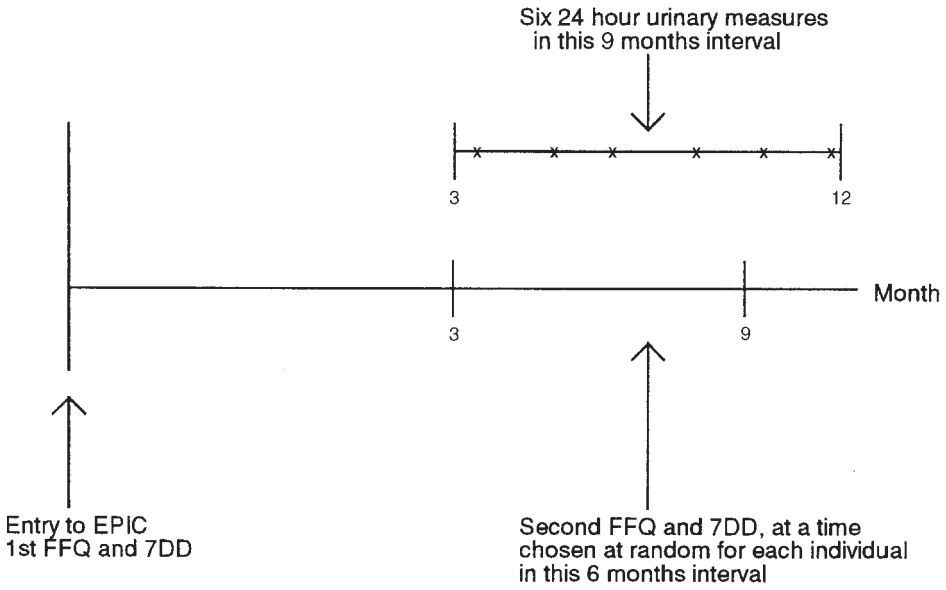

Figure 1 Diagram for the study design methods, illustrating the timing of the first and second 7-day diary (7DD) and food frequency questionnaire (FFQ) in relation to six 24-h urine collection

The food frequency questionnaire

The self-administered FFQ was designed to measure an individual's habitual food and nutrient intake during the past year. The questionnaire was a modified version of the FFQ in the US Nurses Health Study[15,16] with a food list that was adapted to include foods that were commonly consumed within the UK. The food list was compiled from national dietary intake data and was based on 130 main food items.[17] The FFQ in the present study was a revised version of the questionnaire previously used in validation studies.[18] For each food item, participants were asked to indicate their usual consumption from nine frequency categories, ranging from never or <1/month to ⩾6 times per day. The FFQ did not include specific questions on portion size but rather specified medium servings, defined by natural (e.g. apple, slice of bread) or household units (e.g. glass, cup, spoon). Calculation of nutrient intake for both instruments was based on published food composition tables[19] which are frequently updated with data on new food items and nutrients.

Urine collections

Participants received written and verbal instructions on the technique of collecting 24-h urine samples and the use of PABA tablets (PABA; PABA check, Laboratories for Applied Biology, London). On the first morning of the urine collection, participants were asked to discard their first urine specimen and from then on to collect all specimens for the next 24 hours, up to and including the first urine specimen of the next day. They were given three 80 mg PABA tablets to take at each meal on the day of the urine collection to verify completeness of the 24-h urine collection.[20]

The Norwich District Ethics Committee gave permission for both the main EPIC-Norfolk study and this validation study.

Statistical methods

We designate diary measures by R, the FFQ by Q and the urinary measures by M. We assume there is a true but unobservable intake we designate T, in this case comprising the average intake over a year. We designate the population variance of T by σ^2_T. For each nutrient, we take M to be a calibrated but imprecise measure of intake, so that

$$M = T + \varepsilon_M$$

where ε_M is an error term with variance σ^2_M, different for each of the three nutrients.

R and Q we take as biased and imprecise measures of intake, so that

$$R = \alpha_R + \beta_R T + \varepsilon_R \qquad E(\varepsilon_R) = 0,\ Var(\varepsilon_R) = \sigma^2_R$$
$$Q = \alpha_Q + \beta_Q T + \varepsilon_Q \qquad E(\varepsilon_Q) = 0,\ Var(\varepsilon_Q) = \sigma^2_Q$$

where β_R and β_Q can be regarded as scale factors and ε_R and ε_Q as measurement bias. As before, the parameters differ for each nutrient. We also denote, by ρ_{RR} and ρ_{QQ} respectively, the correlation between values of ε_R and ε_Q from repeated measures of R and Q, and by ρ_{RQ} the correlation between ε_R and ε_Q.

This model is similar to the one introduced by Kipnis *et al.*,[8] except that the error terms are parameterized differently, and a simplified version of that investigated by Plummer and Clayton.[10,11] Person-specific bias is included in the error term, and manifests itself as correlated error. Kipnis *et al.* introduced a separate term for individual bias.

We are interested primarily in estimating the parameters σ^2_T, σ^2_M, σ^2_R, σ^2_Q, β_R, β_Q, ρ_{RR}, ρ_{QQ} and ρ_{RQ}.

We assume that errors in M are independent of errors in R and Q, and also between repeated measures of M.

In particular, we are interested in estimating the regression dilution correction for R and Q, given by

$$[\text{Correction Factor for } R]^{-1} = \beta_R\sigma^2_T/(\beta^2_R\sigma^2_T + \sigma^2_R)$$
$$[\text{Correction Factor for } Q]^{-1} = \beta_Q\sigma^2_T/(\beta^2_Q\sigma^2_T + \sigma^2_Q)$$

when only one measure of R or Q has been used as the basis for diet-disease association estimates.

To estimate the required parameter, we have used the method of moments, equating each observed variance and covariance to its expected value.[5,6,10,11,21] We have

$$E(\bar{S}_{R_i,R_i}) = \beta^2_R\sigma^2_T + \sigma^2_R$$
$$E(\bar{S}_{R_i,R_j}) = \beta^2_R\sigma^2_T + \rho_{RR}\sigma^2_R \quad i \neq j$$
$$E(\bar{S}_{Q_i,Q_i}) = \beta^2_{QR}\sigma^2_T + \sigma^2_Q$$
$$E(\bar{S}_{Q_i,Q_j}) = \beta^2_R\sigma^2_T + \rho_{QQ}\sigma^2_Q \quad i \neq j$$
$$E(\bar{S}_{M_i,M_i}) = \sigma^2_T + \sigma^2_M$$
$$E(\bar{S}_{M_i,M_j}) = \sigma^2_T + \rho_{MM}\sigma^2_M \quad i \neq j$$
$$E(\bar{S}_{R_i,Q_j}) = \beta_R\beta_Q\sigma^2_T + \rho_{RQ}\sigma_R\sigma_Q$$
$$E(\bar{S}_{R_i,M_j}) = \beta_R\sigma^2_T$$
$$E(\bar{S}_{Q_i,M_j}) = \beta_Q\sigma^2_T$$

where for $X, Y \in \{R, Q, M\}$, $\bar{S}_{X_i,X_i}$ is the average of sample variances over the repeated measures on measurement method X, $\bar{S}_{X_i,X_j}$ is the average of sample covariances over all pairs i, j of measures on measurement method X, $i \neq j$, and $\bar{S}_{X_i,Y_j}$ is the average over all i, j of sample covariances for the ith measure on measurement method X and the jth measure on measurement method Y and where R_i, Q_i and M_i designate repeated measures of R, Q and M, respectively. Using the method of moments, no distribution assumptions on T, ε_R, ε_Q and ε_M are needed. We only assume that their variances and covariances exist and are finite. It is known that the moment estimators are identical to the maximum likelihood estimators if T, ε_R, ε_Q and ε_M are normally distributed and when the repeats are complete.

We have estimated the unknown parameters for N, K and Na separately, and the regression dilution correction factor, together with the variance of the estimate for each nutrient for both the 7DD and the FFQ. The expressions of the estimates for unknown parameters are given in the Appendix.

Results

Tables 1–4 give the main sample statistics. The FFQ gives higher mean intakes than the 7DD for N and K. The mean intakes from three measurement methods differ significantly ($P < 0.001$) except the FFQ and the 7DD sodium intakes. Both the FFQ and the 7DD underestimate sodium intake, compared to the urinary measure. For each nutrient, the observed variance of the 7DD measure was smallest. For Na, the urine measure had

Table 1 Sample mean, standard deviation, coefficient of variation and 95% CI of sample mean of average daily intake of nitrogen, potassium and sodium by each measurement method. *R*: 7-day diary, *Q*: food frequency questionnaire and *M*: 24-h urine

	Mean	Standard deviation	Coefficient of variation	95% CI
Nitrogen (g)				
R	11.87	2.85	0.24	(11.53–12.20)
Q	13.24	3.94	0.30	(12.77–13.72
M	10.91	2.95	0.39	(10.71–11.12)
Potassium (mmol)				
R	82.12	18.80	0.23	(79.90–84.34)
Q	97.19	24.13	0.25	(94.28–100.10)
M	73.16	22.29	0.27	(71.61–74.71)
Sodium (mmol)				
R	120.16	36.50	0.30	(115.85–124.47)
Q	120.24	46.71	0.39	(114.61–125.87)
M	145.01	56.37	0.39	(141.08–148.93)

Table 2 Sample variances and correlations (average over repeated measures) on each measurement method. *R*: 7-day diary, *Q*: food frequency questionnaire and *M*: 24-h urine

	Sample variance			Correlation between repeats		
	N	K	Na	N	K	Na
7-day diary (*R*)	8.14	357.34	1339.26	0.72	0.71	0.64
Food frequency questionnaire (*Q*)	15.48	584.07	2201.29	0.47	0.60	0.59
24-h urine (*M*)	8.71	499.61	3189.32	0.58	0.47	0.50

Table 3 Sample correlation between different methods of measurement (average over all pairs of measurements). *R*: 7-day diary, *Q*: food frequency questionnaire and *M*: 24-h urine

	N	K	Na
R and Q	0.33	0.41	0.29
R and M	0.49	0.38	0.36
Q and M	0.15	0.22	0.13

Table 4 Sample correlation between different nutrients on the same measurement method (average over all repeats)

	N & K	N & Na	K & Na
7-day diary (R)	0.69	0.68	0.53
Food frequency questionnaire (Q)	0.79	0.77	0.75
24-h urine (M)	0.55	0.61	0.47

substantially the largest variance (Table 2). It is striking from Table 3, that although the correlations between *R* and *M* and between *R* and *Q* are of similar magnitude for all three nutrients, the correlations between *Q* and *M* are markedly smaller. It is noticeable, in Table 4, that the correlation between the estimated intakes of the three nutrients are largest for the FFQ, and smallest for the urinary measure.

Figures 2(a), (b) and (c) display graphically the relationship between *R* and *M* and between *Q* and *M* for N, K and Na. For *R* and *Q* separately, the sample is divided into quintiles, and the mean value of *M* calculated for each quintile. For *R*, for each of the three nutrients, the mean value of *M* increases steadily with the increasing quintile of *R*. The estimated slope (with standard error) is 1.11 (0.16), 6.45 (0.90) and 14.00 (1.53) for N, K and Na, respectively. All three slopes are significant ($P < 0.01$). For *Q*, the picture is rather different, with the relationship between mean values of *M* and the quintiles of *Q* having smaller slopes than between *M* and *R*. The estimated slope (with standard error) is 0.42 (0.32), 3.77 (2.01) and 6.99 (2.61), respectively. All three slopes are not significantly different from zero. The difference between the two slopes for N, K and Na are significant at the 0.05, 0.22 and 0.02 level, respectively, with a combined significance level of approximately 0.004 ($\chi^2_1 = 8.03$). Figure 2 corroborates graphically the values in Table 3.

Tables 5–8 give estimates of the basic parameters in the model. From Table 5, it is clear that the error variance of *Q* is substantially larger than that of either *R* or *M*. The error variances of *R* and *M* are similar in magnitude to the estimated variance of the true underlying exposure *T*. Table 6 indicates that repeated measures of both *R* and *Q* have substantially correlated errors, and that errors in *R* and *Q* are also moderately correlated. The values in Table 6 indicate clearly the danger of basing validation studies just on record or questionnaire type instruments.

Table 7 gives the correlations between the 'true' intakes of N, K and Na, and between the errors of measurement of the three nutrients from the same 7DD, FFQ or urine sample. The correlations for *Q* are greater than the correlations for *R* ($P = 0.09$ for N, 0.05 for K and 0.03 for Na), which in turn are substantially greater than the correlations for *M*.

Table 8 displays the estimates of the scale factors for *R* and *Q*. They indicate, as does Table 5, that the FFQ is rather weakly associated with *T*, whereas the diary *R*, at least for N and K, relates more closely to true intake. The lower values seen for Na clearly reflect the underestimate of sodium intake by both the 7DD and the FFQ, as seen in Table 1.

The regression dilution correction factors for the 7DD and the FFQ, for each of the nutrients, is given in Table 9. The values

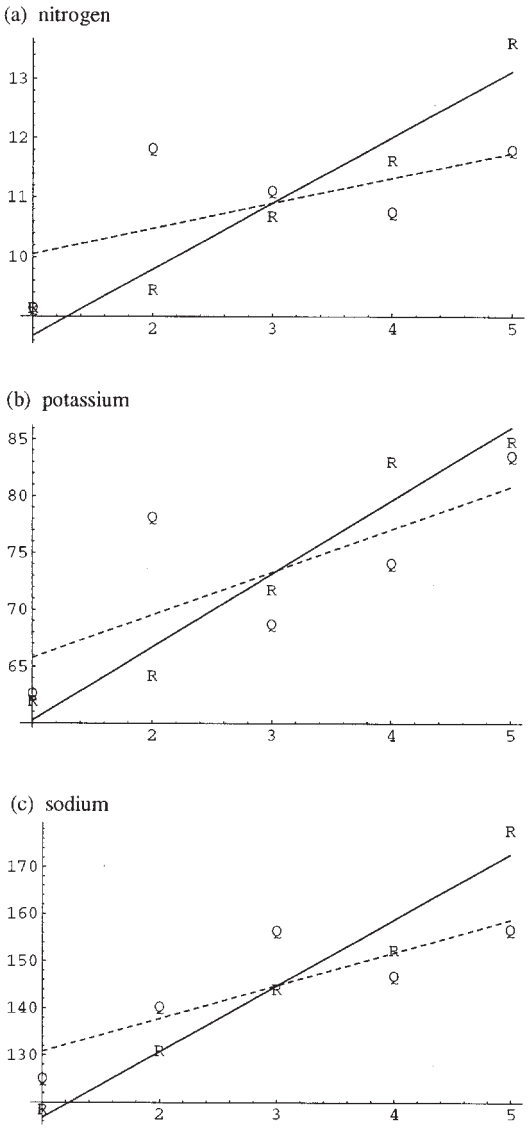

Figure 2 The regression of the mean of urinary values on the quintiles of *R* and *Q*. *x*-axis is the mean of urinary values and *y*-axis is the quintile

Table 5 Estimated variance of the true exposure and associated error variance of three dietary methods for each nutrient. *R*: 7-day diary, *Q*: food frequency questionnaire and *M*: 24-h urine

	N	K	Na
'True' exposure (*T*)	5.07	235.70	1572.68
Error of *R*	4.79	246.04	991.41
Error of *Q*	14.89	522.19	2131.19
Error of *M*	3.63	263.91	1616.64

Table 6 Estimated correlation in the error of repeated measures on each measurement method. *R*: 7-day diary, *Q*: food frequency questionnaire

Error correlation:	N	K	Na
Between repeated of *R*	0.52	0.58	0.52
Between repeated of *Q*	0.45	0.56	0.51
Between repeated of *R* and *Q*	0.27	0.29	0.24

Table 7 Estimated correlation between the 'true' intake of the different nutrients and between the errors in the estimated intakes by the three dietary methods. *R*: 7-day diary, *Q*: food frequency questionnaire and *M*: 24-h urine

	N & K	N & Na	K & Na
'True' exposure (T)	0.69	0.81	0.56
Error of *R*	0.70	0.63	0.52
Error of *Q*	0.80	0.77	0.77
Error of *M*	0.41	0.38	0.39

Table 8 Estimated scale factor of the 7-day diary (*R*) and the food frequency questionnaire (*Q*)

	N	K	Na
7-day diary (*R*)	0.81	0.69	0.47
Food frequency questionnaire (*Q*)	0.34	0.51	0.21

given refer to the correction needed both when only one measure of *R* or *Q* has been used to estimate diet-disease association, and in the situation when the diet-disease association has been estimated from two measures of *R* or *Q*. This latter correction incorporates the correlation in the error terms between repeats of *R* and *Q*. The reduction in the required corrections in going from one to two measures of *R* and *Q* is clearly less than it would be if the error terms were independent.

The implication of these results can be illustrated by considering two examples both of which are quantitative factors measured in the EPIC-Norfolk cohort and are predictors of subsequent mortality. One is lung function, as measured by FEV_1 (forced expired volume in one second), which is related to fruit and vegetable consumption and hence potassium intake.[22] The second is plasma vitamin C, also related to fruit and vegetable intake.[23] For the individuals on whom we have both a 7DD and the FFQ coded from the initial EPIC-Norfolk examination, and on whom we have FEV_1 (n = 2018) and plasma vitamin C (n = 1839) measured, linear regression of these two measures against dietary potassium was performed. Age and sex, with age-sex interaction, height, body mass index (BMI, [kg/m^2]) and cigarette smoking were also included in the model. The estimated regression coefficients, with their standard errors, for the 7DD and the FFQ are given in Table 10, together with the correction factor and the corrected regression coefficient. These latter values can be interpreted as the regression coefficients one would have obtained if average dietary potassium intake had been assessed without error by 24-h urinary measures (i.e. sufficient number of such assays to reduce random error to a negligible value) and so should be the same when derived from the 7DD or the FFQ. The corrected coefficients for the FFQ and 7DD are almost identical for both FEV_1 and plasma vitamin C, despite wide difference in the observed coefficients.

Table 9 The univariate correction factor to be applied to when there are *k* repeats in the main study

k	*R* only	*Q* only	The best combination of *R* and *Q*
1			
Nitrogen	1.97 (1.56–2.38)[a]	8.96 (1.73–16.19)	1.91 (1.50–2.32)
Potassium	2.21 (1.62–2.80)	4.84 (2.41–7.27)	2.03 (1.48–2.58)
Sodium	1.81 (1.28–2.34)	6.63 (0.51–12.75)	1.68 (1.17–2.19)
2			
Nitrogen	1.69 (1.24–2.14)	6.57 (1.20–11.94)	1.65 (1.22–2.08)
Potassium	1.88 (1.27–2.49)	3.87 (1.83–5.91)	1.78 (1.25–2.31)
Sodium	1.49 (0.98–2.00)	5.05 (0.35–9.75)	1.41 (0.94–1.88)

[a] 95% CI.

Table 10 Crude and adjusted regression coefficients of FEV_1 and plasma vitamin C on potassium intake as estimated by the 7-day diary or the food frequency questionnaire (FFQ)

	Crude regression coefficient	Correction factor	Adjusted regression coefficient
FEV_1			
7-day diary	0.0070 (0.0016)[a]	2.21	0.0154 (0.0041)
FFQ	0.0032 (0.0012)	4.84	0.0155 (0.0069)
Plasma vitamin C			
7-day diary	0.0067 (0.00067)	2.21	0.0149 (0.0025)
FFQ	0.0030 (0.00049)	4.84	0.0145 (0.0044)

[a] Standard error.

Discussion

The results presented here demonstrate that a 7DD provides a better estimate of average intake, as assessed by urinary measures, than does the EPIC-Norfolk FFQ, for each of the three nutrients we have considered. One can only speculate whether the same conclusion holds for other nutrients. Additional biomarkers for other nutrients, substantially correlated with 'true' intake, are required. The correlations between the diary estimates of intake of the three nutrients and the urinary measures are between 0.36 and 0.49, for the FFQ the correlations vary from 0.13 to 0.22. This comparison can be seen graphically in Figures 2(a), (b) and (c). From the values in Tables 5 and 8, one can calculate the correlations between the 7DD measures and average intake T and between the FFQ and *T*. For the 7DD, the values are 0.64, 0.56 and 0.59 for nitrogen, potassium and sodium, respectively, whereas the corresponding values for the FFQ are 0.20, 0.33 and 0.18, respectively. For the 24-h urine, the corresponding values are 0.76, 0.69 and 0.69. The latter two values correspond to estimates of 0.74 and 0.66 from the INTERSALT study.[24]

The correlations for the FFQ are much lower than those often reported for the FFQ from validation studies.[2] Usually, however, the validation is done in terms of other record-based instruments. As can be seen from Table 6, however, correlated errors between the 7DD and the FFQ will lead to overestimation of the correlation between the FFQ and underlying intake. The values in Table 3, comparing the correlations between intakes from the FFQ and the 7DD and between intakes from the FFQ and the 24-h urinary measure, illustrate the point. The lower values for the latter could have been due to excessive error variation in the 24-h urinary measure, but Table 6 demonstrates that this is not the case. The lower correlations are due to the substantial correlation between the errors in the FFQ and the 7DD and are of a similar magnitude for each of the three nutrients.

The correlation between the errors of the 7DD and the FFQ can be expressed in terms of the corresponding correlation between person-specific biases, following Kipnis *et al.*,[8] if one assumes that all the correlation between the two errors derives from correlation between person-specific biases. This correlation is then given by $\rho_{RQ}/(\rho_{RR}\,\rho_{QQ})^{1/2}$ with values of 0.56, 0.51 and 0.47 for N, K and Na, respectively. These values are at the top end of the values Kipnis *et al.* considered (Table 2 of ref.8).

The lower correlations between the FFQ and underlying intake leads to a larger degree of regression dilution than has often been assumed. With correction factors in the range 4 to 9, only large underlying relative risks will lead to appreciably elevated observed risks when using the FFQ, at least for the three nutrients under consideration. In addition to the correction factor estimates being large, for the FFQ, the associated standard errors for the estimated correction factor are large also, as can be seen in Figure 3. For a validation study to yield acceptable precision in the estimates of the correction factors for the FFQ, it would need to include several thousand individuals.

In contrast to the FFQ, the 7DD performs reasonably well. Correlation with the underlying intake is substantial, the correction factors are much smaller than for the FFQ, and the precision with which they are estimated is much greater. In the light of the values in Table 9, the 7DD should have a substantially enhanced capacity to identify diet-disease association over the FFQ.

The substantial correlation of the error terms for both the diary and the FFQ requires comment. It may be associated with

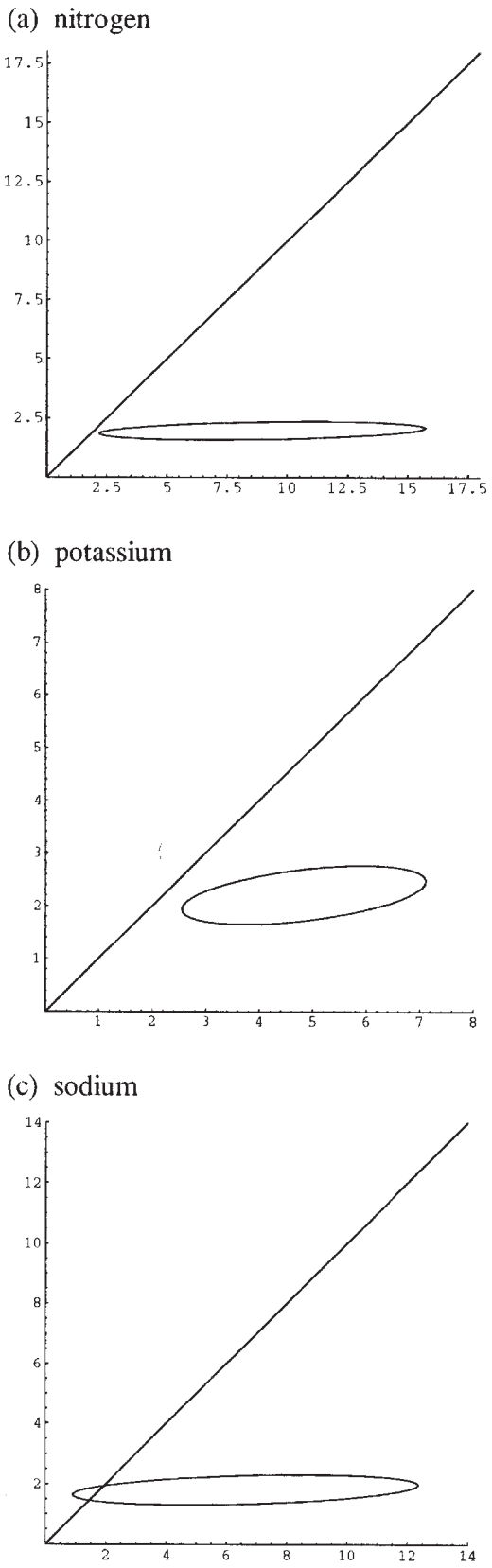

Figure 3 95% confidence region of the correction factors of *R* and *Q*. *x*-axis is the correction factor on *Q* and *y*-axis is the correction factor on *R*

the well-known phenomenon of underreporting, i.e. that there is considerable inter-individual variation in the way dietary records and questionnaires are completed.[4] One consequence for the design of epidemiological studies, however, is that there is limited value in obtaining several repeat diaries or FFQ. In particular, the poor performance of the FFQ cannot be overcome by simply repeating it on numerous occasions. Given the level of correlation in the error term, the most that can be achieved is the equivalent of two independent repeats.

Similar analyses have been published for nitrogen intake from an early validation study for the EPIC-Norfolk cohort.[10,11] There is, however, an essential difference in that the earlier publication did not use an open-ended diet diary, but a 7-day daily check list. The earlier work also uses two different FFQ and so although our methodology derives in part from the earlier work, the substantive results are not comparable.

The assumption that the error associated with a single urinary measure is independent from the errors of the repeated urinary measures, and also from the errors of the 7DD and FFQ needs some comment. The situation of dependence between errors in the repeated 24-h urine measures has been considered in an earlier paper.[5] Since the correlation of the mean of six urine measures with habitual intake over one year will be high (from Table 5), the effect of error correlation up to 0.4 will be relatively small.[3] With regard to possible correlations between errors in the urinary measures and the corresponding measures from the 7DD and the FFQ, the urinary measure is clearly of a different type, and physically independent from the questionnaire type instruments. The timing of the urine collection was also chosen not to coincide with completion of either the 7DD or the FFQ. It is conceivable, however, that any approach to study participants may change their behaviour, thereby inducing some level of correlation. It is impossible to estimate this correlation without a further independent measure which we are not aware of being available. We have undertaken sensitivity analysis to investigate the degree to which our estimates of the correction factors in Table 9 change if error in the FFQ or 7DD is correlated with error in the urinary measure. The sensitivity analysis focused on the nitrogen values. We define ρ_{RM} and ρ_{QM} as the correlation between error in the urinary measure, *M*, and error in the 7DD, *R*, and the FFQ, *Q*, respectively. The correction factor for the FFQ is highly sensitive to moderate values of ρ_{QM}, becoming infinite with $\rho_{QM} = 0.23$. The correction factor for the 7DD is relatively insensitive to moderate values of ρ_{RM}, equalling 3.09 with $\rho_{RM} = 0.3$. The conclusion thus is similar to that in our earlier paper,[5] that the sensitivity of the estimated correction factor to departures from the assumption of independence becomes greater the more weakly the measured intake (*R*, *Q* or *M*) is related to true intake (*T*). It is also worth noting that the values in Table 10 indicate that the relative magnitudes of the correction factors (for potassium) for the 7DD and the FFQ under the independence assumptions appear to be approximately correct since the adjusted regression coefficients are the same for the 7DD and the FFQ for both FEV_1 and plasma vitamin C.

An additional feature of the results presented in this paper is the high level of correlation between the errors in estimating the three different nutrients (Table 7). These correlations are greater for the FFQ than for the 7DD and will lead to greater apparent confounding in the univariate situation, but their

main effect will be in the multivariate situation. If several dietary factors are being examined simultaneously, these correlations will contribute to the multivariate regression dilution. In the multivariate situation the effect of measurement error is not simply dilution of each parameter, since there is an additive component, as discussed by Kipnis *et al.*[25] These authors label this additive component regression contamination. The greater the correlation between the errors of different nutrients, the greater the resulting contamination is likely to be.[6]

As a final point, the large values for the regression correction factor seen for the FFQ derive both from the error variances of the FFQ and the underlying between-individual variation in the study population. If the latter is increased, the correction factor will become smaller, which underlines the importance of variation across the study population. This forms an important part of the design rationale of EPIC.

Acknowledgements

The EPIC-Norfolk Study is funded by the Cancer Research Campaign, the Medical Research Council, the British Heart Foundation, the Ministry of Agriculture, Fisheries and Food, and the Europe against Cancer Programme of the Commission of the European Communities. We are grateful to all the participants and general practitioners who have helped with this study, and to the nurses, technicians and the staff of the EPIC co-ordinating centre. Undergraduate students from Coleraine, Surrey, Newcastle, Leeds Metropolitan and Wageningen Universities assisted with the collection and interpretation of the fieldwork. The third author was funded by the British Council to do this collaborative work. The authors would like to thank the referees for their helpful comments.

References

[1] COMA Nutritional Aspects of the Development of Cancer. *Committee on Medical Aspects of Food and Nutrition Policy Report on Health and Social Subjects* **48**, London: Department of Health, The Stationery Office, 1998.

[2] Willett W. *Nutritional Epidemiology*. New York: Oxford University Press, 1998.

[3] Bingham S, Gill C, Welch A *et al.* Validation of dietary assessment methods in the UK arm of EPIC. *Int J Epidemiol* 1997;**26:**S137–51.

[4] Bingham SA, Gill C, Welch A *et al.* Comparison of dietary assessment methods in nutritional epidemiology: weighed records v. 24 h recalls, food-frequency questionnaires and estimated-diet records. *Br J Nutr* 1994;**72:**619–43.

[5] Wong MY, Day NE, Bashir SA, Duffy SW. Measurement error in epidemiology: the design of validation studies I: univariate situation. *Stat Med* 1999;**18:**2815–29.

[6] Wong MY, Day NE, Wareham NJ. Measurement error in epidemiology: the design of validation studies II: bivariate situation. *Stat Med* 1999;**18:**2830–45.

[7] Kaaks R, Slimani N, Riboli E. Pilot phase studies on the accuracy of dietary intake measurements in the EPIC project: overall evaluation of results. European Prospective Investigation into Cancer and Nutrition. *Int J Epidemiol* 1997;**26(Suppl.1):**S26–36.

[8] Kipnis V, Carroll RJ, Freedman LS, Li L. Implications of a new dietary measurement error model for estimation of relative risk: application to four calibration studies. *Am J Epidemiol* 1999;**150:**642–51.

[9] Kroke A, Klipstein-Grobusch K, Voss S *et al.* Validation of a self-administered food-frequency questionnaire administered in the European Prospective Investigation into Cancer and Nutrition (EPIC) Study: comparison of energy, protein, and macronutrient intakes estimated with the doubly labeled water, urinary nitrogen, and repeated 24-h dietary recall methods. *Am J Clin Nutr* 1999;**70:**439–47.

[10] Plummer M, Clayton D. Measurement error in dietary assessment: an investigation using covariance structure models, Part I. *Stat Med* 1993;**12:**925–35.

[11] Plummer M, Clayton D. Measurement error in dietary assessment: an investigation using covariance structure models, Part II. *Stat Med* 1993;**12:**937–48.

[12] INTERSALT Cooperative Research Group. INTERSALT: an international study of electrolyte excretion and flood pressure: result for 24-hour urinary sodium and potassium excretion. *Br Med J* 1988; **297:**319–28.

[13] Riboli E. Nutrition and cancer: background and rationale of the European Prospective Investigation into Cancer and Nutrition (EPIC). *Ann Oncol* 1992;**3:**783–91.

[14] Day NE, Oakes S, Luben R *et al.* EPIC in Norfolk: study design and characteristics of the cohort. *Br J Cancer* 1999;**80(Suppl.1):**95–103.

[15] Willett WC, Sampson L, Stampfer MJ *et al.* Reproducibility and validity of a semiquantitative food frequency questionnaire. *Am J Epidemiol* 1985;**122:**51–65.

[16] Willett WC, Sampson L, Browne ML *et al.* The use of a self-administered questionnaire to assess diet four years in the past. *Am J Epidemiol* 1988;**127:**188–99.

[17] Committee NFS. *Annual Report of Household Food Consumption and Expenditure 1980.* London: Her Majesty's Stationery Office, 1982.

[18] Bingham SA, Gill C, Welch A *et al.* Comparison of dietary assessment methods in nutritional epidemiology: weighted records v. 24h recalls, food-frequency questionnaires and estimated-diet records. *Br J Nutr* 1994;**72:**619–43.

[19] Holland B, Welch AA, Unwin ID, Buss DH, Paul AA, Southgate DA. *McCance and Widdowson's The Composition of Foods*. Fifth revised and extended edition. The Royal Society of Chemistry and Ministry of Agriculture, Fisheries and Food, 1991 and Supplements to this edition.

[20] Bingham S, Cummings JH. The use of 4-aminobenzoic acid as a marker to validate the completeness of 24h urine collections in man. *Clin Sci* 1983;**64:**629–35.

[21] Kaaks R, Riboki E, Esteve J, van Kappel A, van Staveren W. Estimating the accuracy of dietary questionnaire assessments: validation in terms of structural equation models. *Stat Med* 1994;**13:**127–42.

[22] Hu G, Cassano PA. Antioxidant nutrients and pulmonary function: The Third National Health and Nutrition Examination Survey (NHANES III). *Am J Epidemiol* 2000;**151:**975–81.

[23] Price JF, Fowkes FGR. Antioxidant vitamins in the prevention of cardiovascular disease: the epidemiological evidence. *Eur Heart J* 1997;**18:**719–27.

[24] Dyer AR, Shirley M, Elliott P. Urinary electrolyte excretion in 24 h and blood pressure in the Intersalt study: estimate of reliability. *Am J Epidemiol* 1994;**139:**927–39.

[25] Kipnis V, Freedman LS, Brown CC, Hartman AM, Schatzkin A, Wacholder S. Effect of measurement error on energy-adjustment models in nutritional epidemiology. *Am J Epidemiol* 1997;**146:** 842–55.

Appendix

Under the assumption that the errors between successive determinations using biomarker method are uncorrelated (that is, $\rho_{MM} = 0$) and $\rho_{RM} = \rho_{QM} = 0$, if the validation study consists of repeated measures on R, Q and M, the estimators are obtained by equating the sample statistics to their expected values. We thus get

$$\hat{\beta}_R = \frac{\bar{S}_{R_i,M_j}}{\bar{S}_{M_i,M_j}},$$

$$\hat{\beta}_Q = \frac{\bar{S}_{Q_i,M_j}}{\bar{S}_{M_i,M_j}},$$

$$\hat{\sigma}_T^2 = \bar{S}_{M_i,M_j},$$

$$\hat{\sigma}_R^2 = \bar{S}_{R_i,R_i} - \frac{\bar{S}_{R_i,M_j}^2}{\bar{S}_{M_i,M_j}},$$

$$\hat{\sigma}_Q^2 = \bar{S}_{Q_i,Q_i} - \frac{\bar{S}_{Q_i,M_j}^2}{\bar{S}_{M_i,M_j}},$$

$$\hat{\sigma}_M^2 = \bar{S}_{M_i,M_i} - \bar{S}_{M_i,M_j},$$

$$\hat{\rho}_{RR} = \frac{\bar{S}_{R_i,R_j}\bar{S}_{M_i,M_j} - \bar{S}_{R_i,M_j}^2}{\bar{S}_{R_i,R_i}\bar{S}_{M_i,M_j} - \bar{S}_{R_i,M_j}^2}$$

$$\hat{\rho}_{QQ} = \frac{\bar{S}_{Q_i,Q_j}\bar{S}_{M_i,M_j} - \bar{S}_{Q_i,M_j}^2}{\bar{S}_{Q_i,Q_i}\bar{S}_{M_i,M_j} - \bar{S}_{Q_i,M_j}^2},$$

$$\hat{\rho}_{RQ} = \frac{\bar{S}_{R_i,Q_j}\bar{S}_{M_i,M_j} - \bar{S}_{R_i,M_j}\bar{S}_{Q_i,M_j}}{\sqrt{\bar{S}_{R_i,R_i}\bar{S}_{M_i,M_j} - \bar{S}_{R_i,M_j}^2}\sqrt{\bar{S}_{Q_i,Q_i}\bar{S}_{M_i,M_j} - \bar{S}_{R_i,M_j}^2}}.$$

The estimates of the univariate correction factor for R and Q are, thus, $\bar{S}_{R_i,M_j}/\bar{S}_{R_i,R_i}$ and $\bar{S}_{Q_i,M_j}/\bar{S}_{Q_i,Q_i}$, respectively.

Since the estimates of unknown parameters and univariate correction factors are a function of sample statistics from validation study, we obtain their asymptotic variances by the delta method.

The estimated asymptotic variances of the univariate correlation factor of R and Q are equal to

$$\left(\frac{\bar{S}_{R_i,M_j}}{\bar{S}_{R_i,R_i}}\right)^2\left[\frac{1}{m_r m_m \bar{S}_{R_i,M_j}^2}\left\{(\bar{S}_{R_i,R_i} + (m_r - 1)\bar{S}_{R_i,R_j})(\bar{S}_{M_i,M_i} + (m_m - 1)\bar{S}_{M_i,M_j}) + m_r m_m \bar{S}_{R_i,M_j}^2\right\} - \frac{2}{m_r \bar{S}_{R_i,R_i}^2}\left\{\bar{S}_{R_i,R_i}^2 + (m_r - 1)\bar{S}_{R_i,R_j}(2\bar{S}_{R_i,R_i} - \bar{S}_{R_i,R_j})\right\}\right],$$

and

$$\left(\frac{\bar{S}_{Q_i,M_j}}{\bar{S}_{Q_i,Q_i}}\right)^2\left[\frac{1}{m_q m_m \bar{S}_{Q_i,M_j}^2}\left\{(\bar{S}_{Q_i,Q_i} + (m_q - 1)\bar{S}_{Q_i,Q_j})(\bar{S}_{M_i,M_i} + (m_m - 1)\bar{S}_{M_i,M_j}) + m_q m_m \bar{S}_{Q_i,M_j}^2\right\} - \frac{2}{m_q \bar{S}_{Q_i,Q_i}^2}\left\{\bar{S}_{Q_i,Q_i}^2 + (m_q - 1)\bar{S}_{Q_i,Q_j}(2\bar{S}_{Q_i,Q_i} - \bar{S}_{Q_i,Q_j})\right\}\right],$$

where m_r, m_q and m_m are the number of repeated measures on R, Q and M, respectively.

International Journal of Epidemiology 2001;**30**:317–319

Commentary: Dietary diaries versus food frequency questionnaires—a case of undigestible data

Walter Willett

In conducting studies of diet and disease risk, methods of measuring diet with sufficient validity to detect important associations are essential. Cost is also a critical factor because prospective studies, which are necessarily large, are desirable to avoid problems of selection and recall bias. Most investigators have converged to use some form of a food frequency questionnaire (FFQ) for this purpose, and the validity of this approach has been documented repeatedly by comparisons with more detailed methods, correlations with biochemical indicators of dietary factors, and the ability to predict risk of future disease.[1,2] However, all methods of dietary assessment are imperfect, and

Departments of Nutrition and Epidemiology, Harvard School of Public Health, and the Channing Laboratory, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA.

Correspondence: Walter Willett, Department of Nutrition, Harvard School of Public Health, 655 Huntington Ave, Boston, MA 02115, USA.

quantification of measurement error is desirable both to help in the interpretation of findings from epidemiological studies, and to correct relative risks and confidence intervals for this source of error.

In studies of questionnaire validity, the comparison method should have errors that are minimally correlated with those of the questionnaire to avoid spurious overestimation of validity. For this reason, we and others have chosen weighed dietary records for comparisons because they do not depend on memory and allow quantitative measurements of the amounts of foods at the time they are actually consumed. Biochemical indicators of diet are also attractive because their errors should have little correlation with those of reported food intake; unfortunately for many nutrients of interest appropriate biochemical indicators do not exist.

In this issue, Day *et al.*[3] have combined data collected by an FFQ, diet diary, and biochemical measurements of urine to quantify measurement errors and to estimate the correlations of errors between different dietary assessment methods. Unfortunately, their data, although carefully collected, do more to obfuscate than enlighten. In particular, they have been unfair to their FFQ. The primary problems with this paper, which are inter-related, are that the authors have ignored the heterogeneity in their population, and have examined only absolute rather than energy-adjusted intakes.

As regards the first issue, the study population included both men and women of various ages and sizes. This heterogeneity spuriously increases the between-person variation in absolute intakes of nutrients because in any real epidemiological application we would normally control for age, sex, and body size. Any evaluation of questionnaire validity is unrealistic unless the major sources of variation that would not exist, or that could be controlled, in an epidemiological application are first removed.

The second problem is that they have only examined validity for absolute nutrient intake and have ignored the consensus among nutritional epidemiologists over the last decade that the energy-adjusted nutrient intake of a dietary factor, i.e. dietary composition, is appropriately the primary focus of nutritional epidemiology.[1,4] The principle reason for the focus on energy-adjusted intakes is that this is primarily what can be changed by individuals or populations. Individuals must increase or decrease their intake of nutrients by changing the composition of their diets because their total energy intake is constrained within a narrow range by their size and level of physical activity. Analogously, experimentalists evaluating the effects of specific nutrients routinely compare iso-caloric diets, otherwise changes in weight will confound the specific effects of the nutrient being evaluated. There are many important hypotheses relating the protein composition of the diet to risk of chronic disease, but unfortunately, the paper by Day *et al.*, which uses absolute urinary nitrogen excretion as a measure of protein intake, fails to inform us about the value of their methods for examining these issues.

A secondary benefit of adjusting for total energy intake, but not the primary reason for doing so, is that errors in measuring individual nutrients are strongly correlated with errors in measuring total energy intake because over- or under-reporting of individual foods leads to similar errors in all constituents. These correlated errors are strong, ranging from about 0.6 for total energy versus micronutrients to 0.95 for total energy versus macronutrients (unpublished data based on the Nurses' Health Study). Thus, adjustment for total energy 'cancels' a substantial amount of error. Food frequency questionnaires are directed primarily at dietary composition, which is largely determined by the mix of foods that are consumed. Diet records, which provide more precise quantification of foods consumed, will usually be relatively better for measuring absolute intake. However, a limited number of days of diet records will perform less well for dietary composition because within-person day-to-day variation is much greater for dietary composition than for absolute intake.[5]

For the reasons noted above, an examination of validity for absolute rather than energy-adjusted nutrient intake will tend to favour a one-week diet diary compared to an FFQ. Indeed, there is much empirical evidence that the results for energy-adjusted nutrients would be substantially different to those reported by Day *et al*. Quite consistently, the correlations between FFQ and both diet records and biochemical indicators of nutritional status increase after adjustment for energy intake.[1,6] Also, the correlations among different nutrients estimated by the same FFQ decrease greatly after energy adjustment and are much lower than those reported by Day *et al*.[3] In addition, the apparent correlations in errors between the FFQ and 'diet diary' method reported in their paper are likely to be larger than those using weighed diet records as the comparison because many of the cognitive processes were similar between the methods. A higher degree of correlated error is also likely in studies using 24-hour recalls as the comparison method. The issue of correlated errors between an FFQ and weighed diet records has been examined earlier by Spiegelman,[7] who found that correlations between errors were much lower than reported by Day and not sufficiently strong to affect seriously the estimates of validity from studies comparing FFQ with weighed diet records.

The value of FFQ for assessing dietary composition has already been documented objectively by correlations with biochemical indicators and the prediction of outcomes in prospective studies.[2] These questionnaires have great advantages over dietary records in cost and participant burden. These advantages are particularly important because they allow large populations to be enrolled in prospective studies and repeated assessments of diet during the follow-up period. Replicate assessments may not be of great value at a one-year interval as suggested by Day *et al*., although we do not know whether this is true for dietary composition. However, over a longer period, individual diets do change and repeated measures can be of great value.[8,9] Whether 7-day diet diaries or records add useful information above and beyond FFQ remains an open question, and I will look forward to further findings based on the data collected by Day *et al*. Hopefully, in the future they will be analysed and presented in a way that will be useful to epidemiologists.

References

[1] Willett WC. *Nutritional Epidemiology. 2nd Edn*. New York: Oxford University Press, 1998.

[2] Willett WC. Invited commentary: comparison of food frequency questionnaires [editorial; comment]. *Am J Epidemiol* 1998;**148:**1157–59; discussion 1162–65.

[3] Day NE, McKeown N, Wong MY, Welch A, Bingham S. Epidemiological assessment of diet: a comparison of a 7-day diary with a food

frequency questionnaire using urinary markers of nitrogen, potassium and sodium. *Int J Epidemiol* 2001;**30:**309–17.

[4] Willett WC, Howe GR, Kushi LH. Adjustment for total energy intake in epidemiologic studies. *Am J Clin Nutr* 1997;**65(Suppl.):**1220S–28S.

[5] 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:**2546–49.

[6] Stram DO, Hankin JH, Wilkens LR *et al*. Calibration of the dietary questionnaire for a multiethnic cohort in Hawaii and Los Angeles [see comments]. *Am J Epidemiol* 2000;**151:**358–70.

[7] 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.

[8] Hu FB, Stampfer MJ, Manson JE *et al*. Dietary fat intake and the risk of coronary heart disease in women. *N Engl J Med* 1997;**337:**1491–99.

[9] Hu FB, Sampson LA, Stampfer MJ, Rosner BA, Willett WC. A validation study of repeated measurement of diet through food frequency questionnaire in assessing long-term diet among female nurses (abstract). *Fourth International Conference on Dietary Assessment Methods, 17–20 September 2000.*