Validation of dietary intakes measured by diet history against 24 h urinary nitrogen excretion and energy expenditure measured by the doubly-labelled water method in middle-aged women

Alison E. Black*, Ailsa A. Welch† and Sheila A. Bingham‡

Dunn Clinical Nutrition Centre, Hills Road, Cambridge CB2 2DH, UK

(Received 22 October 1998 – Revised 13 August 1999 – Accepted 6 October 1999)

A diet history method for estimating energy and N intakes was validated against 24 h urinary N excretion and energy expenditure measured by the doubly-labelled water (DLW) method. Forty-eight women aged 50–65 years were studied over 1 year. Weighed diet records from 4 d and two 24 h urine collections, for measurement of urinary N excretion, were obtained in each of four seasons. At the end of the year, a diet history was obtained, BMR was measured by whole-body calorimetry, and, in sixteen women, total energy expenditure (EE) was measured by DLW. Energy intake (EI) and N intake (NI) were calculated using food tables. Using weighed records and diet history respectively mean NI were 11·21 (SD 2·09) g and 11·47 (SD 2·40) g (NS) and EI were 8·08 (SD 1·54) MJ and 8·20 (SD 1·86) MJ (NS). Mean urine N : NI and EI : BMR values indicated bias to under-reporting by weighed record and diet history techniques in some individuals, but there was no significant difference between these measures at the group level. The Pearson correlation coefficient (r) for urine N *v.* NI was 0·81 for the weighed record and 0·38 for the diet history. The correlation of EE *v.* EI was r 0·48 for weighed record and r 0·11 for diet history. In this study the diet history gave the same estimate of mean intake, but the weighed record appeared to perform better in ranking individuals.

Diet history: Diet record: Validation

The diet history method of assessing dietary intake is favoured for two reasons. First, as a retrospective questionnaire method, it places less burden on subjects than do prospective diet records. Second, it is believed to obtain a better measure of long-term habitual intake than diet records, which, of necessity, record intake on a limited number of days. The debate over the respective merits of the weighed diet record and the diet history is a long one and many studies over several decades have compared their relative validity. On balance the diet history tends to obtain higher mean energy intakes (EI) (see reviews by Bingham, 1987; Black *et al*. 1991). However, without external validation such studies have been unable to reach a firm conclusion as to which, if either, is valid in absolute terms.

A valid dietary assessment should give an estimate of mean intake that is close to the true intake and also be able to rank individuals correctly for intake within acceptable limits of precision. The two requirements do not necessarily go together. For example, a dataset of valid 24 h recalls would give a valid estimate of mean intake but could not rank individuals correctly for habitual intake. A diet record of sufficient length to measure habitual intake might rank individuals correctly but give an invalid estimate of mean intake due to systematic bias across all subjects. If random errors are large and if systematic bias is variable across individuals, then ranking will also be poor. Studies of relative validity indicate whether one method gives a higher intake than another, and comparison with expected energy requirements can help determine the validity of each. However, an independent marker of intake is essential to determine which method best ranks individuals according to true intake.

It is now recognized that there is widespread bias to underestimation of EI by self-reports of food intake. A review of forty-six studies (fifty-six groups) that included measurements of both EI and energy expenditure (EE) measured by the doubly-labelled water (DLW) method (Black, 1999) found a mean underestimation of EI

Abbreviations: DLW, doubly-labelled water; EE, energy expenditure; EI, enegy intake; FFQ, food-frequency questionnaire; NI, nitrogen intake; PABA, *p*-amino benzoic acid; PAL, physical activity level.

* **Present address:** 9 Birch Close, Cambridge CB4 1XN, UK.

† **Corresponding author:** Ailsa Welch, present address: EPIC, Strangeways Research Laboratory, Worts Causeway, Cambridge CB1 8RN, UK, fax +44 (0)1223 740 177, email ailsa.welch@srl.cam.ac.uk

‡ **Present address:** MRC Dunn Human Nutrition Unit, Hills Road, Cambridge CB2 2DH, UK.

compared with EE of 16 %. In 63 % of groups mean EI was more than 10 % below mean EE and in only one group was it more than 10 % above EE. Comparisons of EI expressed as a multiple of BMR (EI : BMR) with the physical activity level (PAL) for a sedentary lifestyle have confirmed the widespread tendency to underestimation of EI in large national dietary surveys from several countries (Heywood *et al*. 1993; Klesges *et al*. 1995; Ballard-Barbash *et al*. 1996; Fogelholm *et al*. 1996; Briefel *et al*. 1997; Lafay *et al*. 1997; Price *et al*. 1997; Pryer *et al*. 1997; Rothenberg *et al*. 1997; Braam *et al*. 1998; Gnardellis *et al*. 1998; Voss *et al*. 1998) and also many smaller studies. Mean reported EI : BMR values were predominantly in the range of 1·2–1·5, whereas DLW studies suggest that EE is greater than 1·55 × BMR in all age groups except those aged over 75 years (Black *et al*. 1996).

The number of studies with external validations of the diet history are few, since the majority of DLW studies have used diet records to measure EI and the large national surveys have predominantly used 24 h recalls or food-frequency questionnaires (FFQ). The earliest validations of the diet history method by urinary N excretion (Steen *et al*. 1977; Bingham *et al*. 1982; van Staveren *et al*. 1985; Hultén *et al*. 1990) did not include validation for completeness of the 24 h urine collections by the PABACHEK method (Bingham & Cummings, 1983) which uses *p*-amino benzoic acid (PABA) as a marker. It is known that a substantial proportion of collections may be incomplete; Heitmann (1993) rejected 17 %, Bingham *et al.* (1995) 25 %, and SD Poppitt, G Keogh and AA Black (personal communication) 40 %. The ratio urine N : N intake (NI) is used as the estimate of validity and incomplete collection produces a bias towards showing the reported intake to be valid. Among the externally validated studies of the diet history are two validated by urinary N excretion (Petersen *et al*. 1992; Heitman, 1993), two validated by urinary N excretion without PABA and by EE measured by physical activity questionnaire (Lindroos *et al*. 1993; Visser *et al*. 1995), two validated by DLW EE (Livingstone *et al*. 1992; Rothenberg *et al*. 1998), and one validated by EE measured by physical activity questionnaire (Körtzinger *et al*. 1997). Only the present study has included both DLW EE and PABA-validated urinary N excretion.

The aim of the present study was to determine the validity of the diet history, compared with that of 16 d weighed records, with respect to both mean intake and ranking of individuals using the independent markers of urinary N excretion and DLW EE. The results were compared also with findings from the validation study of other dietary assessment methods, including the 7 d estimated diet record, the Oxford FFQ and an unstructured 24 h recall, undertaken for the Cambridge arm of the European Prospective Investigation into Cancer (Bingham *et al.* 1994).

Methods

Outline of the study

EI and NI were assessed in forty-eight middle-aged women by 4 d weighed records in each of four seasons (16 d in total). Urinary 24 h N excretion (2 d in each of the four seasons, eight collections in total) was also measured. These data were obtained in a study to determine the best dietary assessment method to use in the European Prospective Investigation into Cancer (Bingham *et al*. 1994, 1995) conducted between 1989 and 1990. The present study additionally assessed EI and NI in the fourth season by diet history and, in sixteen subjects, measured EE by DLW and BMR by whole-body calorimetry. Twenty-four diet histories were obtained in autumn 1989, the remaining diet histories and the DLW measurements in autumn 1990.

Recruitment

For the main study, all women aged 50–65 years from the lists of two general practices in Cambridge were contacted by post. Those expressing interest in a detailed study of diet were contacted by telephone and visited at home and, if still willing, were entered into the study. In all, 160 women completed the main study, about 15 % of the total number contacted initially. No exclusions were made on the grounds of ill health. The study was conducted in two groups. For the present study, after completion of the fourth season, subjects from both groups were asked if they were willing to be visited at home by a dietitian to obtain a diet history. Those willing to participate were ranked according to their reported EI from the 16 d weighed record. When a subject agreed to the diet history, an attempt was made to recruit a ‘matched’ subject with similar reported EI. The aim was to obtain paired groups with the same mean EI in order to test observer bias in the technique of the diet history. One member of each pair was interviewed by A. W. and the other by A. B. This procedure was followed until twenty-four pairs (forty-eight subjects) had been recruited.

Subjects in the second group were also approached before the fourth season for measurement of metabolic rate using DLW and calorimetry. Forty-five out of eighty-four subjects were willing to participate. Final selection of subjects for DLW measurement was confined to those who provided at least five valid 24 h urine collections and included subjects from the full range of urine N : NI values. The numbers eventually obtained from each fifth of the distribution of urine N : NI were 4 (lowest), 5, 2, 4 and 3 (highest). The intention was for the DLW measurement to span the 4 d of diet records. Unfortunately, owing to a scarcity of water enriched with ^{18}O at that time and a consequent unanticipated price increase, the study was limited to sixteen subjects and delayed in some subjects until after completion of the final diet records. The mean time that elapsed between final diet records and DLW measurement was 0·5 (SD 2·5) weeks.

Dietary assessment

For weighed records, subjects were instructed to weigh each individual food item using cumulative weighing and to provide notes on ingredients of composite dishes with approximate quantities. Weighing was by the PETRA system (Cherlyn Electronics, Cambridge, Cambs., UK). The weight and a spoken description of each food item were automatically recorded onto a cassette tape. A special console was used to recover the information from the women's data which was transcribed and then coded manually for computer analysis.

Before embarking on the diet histories, the interviewers standardized their technique by jointly interviewing colleagues and volunteers taking part in other studies. Each diet history was conducted in a single face-to-face semi-structured interview lasting between 1 and 2 h in the subject's own home. The interviewer took each meal and inter-meal period in turn and established a typical week's menu. The interviewer then took each food group in turn and asked about types of items eaten and frequency of consumption. Particular attention was given to alcoholic drinks, eating out, take-away foods and snacks. Portions were quantified by verbal descriptions in terms of familiar volumes and sizes and by reference to photographs derived from a Swedish atlas of food portions. The publication of Crawley (1988) was used a source of average portion sizes and typical weights of standard sized items such as biscuits, confectionery, buns and cakes. Nutrient intake was calculated using the fourth edition of *McCance & Widdowson's The Composition of Foods* (Paul & Southgate, 1978) and supplements (Holland *et al*. 1988, 1989) with additional recipes (Wiles *et al*. 1980) and manufacturers' information.

The alternative dietary assessments are fully described in the original publication (Bingham *et al*. 1994). The Oxford FFQ was based on the US Nurses Study (Willett *et al*. 1985) modified for the UK. The unstructured 24 h recall was a blank piece of paper with an example of how to record the previous day's intake. Published portion weights (Crawley, 1988) were used to calculate nutrient intake. The 7 d estimated record was that developed for the MRC National Survey of Health and Development (Braddon *et al*. 1988; Price *et al*. 1996). Food was recorded at the time of eating and quantified in household measures or with the help of photographs included in the booklet.

BMR

The subjects were brought to the unit for an evening meal of one third of energy requirements calculated as $1{\cdot}4 \times$ BMR estimated from equations (Schofield *et al*. 1985). After an overnight stay in the whole-body calorimeter, they were woken to pass urine at 06.30 hours, subsequently returning to sleep. They were woken for BMR to be measured between 08.00 and 09.00 hours at a temperature of 23°.

Total energy expenditure

A baseline urine specimen was obtained at 06.30 hours on the day of the BMR measurement. The subject then drank the DLW containing 0·07 g ^2H_2O and 0·174 g $H_2{}^{18}O$/kg body weight. Urine samples were obtained from the second voiding of the day on each of the following 14 d. Isotope enrichments were measured using an isotope-ratio mass spectrometer (Aqua Sira, Middlewich, Cheshire, UK) and pool sizes were calculated by extrapolation. The mean value for ^{18}O : ^2H spaces was 1·035 (SD 0·013). EE was calculated using the multipoint technique (Coward, 1988). Individual food quotients were calculated from the 16 d weighed records (Black *et al*. 1986). There was no difference in calculated DLW EE if the diet history food quotient was used. The mean difference between EE calculated using the different food quotients was −3 (SD 91) kJ. Correlation between food quotients from weighed records and from diet history was 0·99.

Urinary nitrogen

Urine collections (24 h) were made using boric acid as a preservative in the collection container (Bingham *et al*. 1995). In each season, two 24 h urine collections were obtained, one during the 4 d period of diet records and one in the 2 d immediately following. Completeness of collection was verified by PABACHEK (Bingham & Cummings, 1983). Three capsules containing 80 mg PABA were taken, one on rising and one each with the midday and evening meals. Urine collections containing less than 205 mg (85 % of the dose) were rejected as incomplete. N in urine was measured by the Kjeldahl technique.

Validation of nitrogen intake

For subjects in N balance, urine N reflects NI. Validation is by comparison of NI with urine N expressed as the ratio urine N : NI. A higher than expected urine N : NI value reflects either incomplete reporting of N (protein) intake, or a reduced (low energy) intake leading to oxidization of protein to supply energy. Bingham & Cummings (1985) found extra-renal losses to be proportional to total N turnover and the average urinary N : NI value to be 0·81 (SD 0·05). In the main study those in the top fifth, with values from 1·00 to 1·47, had significantly lower EI, higher body weight, higher BMI, and higher restrained eating scores than those in the lower four fifths (with values from 0·68 to 0·99) (Bingham *et al*. 1995). In the present study subjects with a urine N : NI value greater than 1·00 were deemed under-reporters.

The alternative approach of Isaksson (1980) assumes extra-renal losses of N to be constant at 2 g/d. 'True' protein intake is calculated as (urine $N + 2) \times 6{\cdot}25$ g and compared directly with the reported protein intake. A value of 1·00 is expected. Authors using both calculations have reported no practical difference between them (Heitmann, 1993). Both values are reported in the present paper.

Validation of reported energy intake

For people in energy balance, habitual EI must equal EE. The assumption is made that individuals are in energy balance and that a single dietary assessment provides a valid measure of intake (within the limits of the precision of the technique) and that a single 14 d measure by DLW provides a valid measure of EE. Validation is by direct comparison of EI with EE expressed as the ratio EI : EE. The expected value is 1·00 and the 95 % confidence limits in the present study were 0·79–1·21 (based on mean within-subject CV on daily EI of 20·8 % and of repeat DLW measurements of 8·9 % (Black *et al*. 1996). Subjects with an EI : EE value less than 0·79 or greater than 1·21 were deemed under- or over-reporters respectively.

Under-reporting is somewhat confusingly indicated by high values of urine N : NI but low value of EI : EE.

However, since each validation ratio has been established independently in previous publications, it was decided to maintain the original configuration for the present study.

Where direct measures of EE are not available, comparison may be made between EI : BMR and the expected EE expressed as EE : BMR or PAL (Goldberg *et al*. 1991; Black, 1996). PAL as measured in a subsample in the present study was 1·65. The Goldberg cut-off is the value below which the reported intake is unlikely to represent either habitual intake or low intakes obtained by chance, taking into account errors in the techniques involved. The equation to calculate the Goldberg cut-off is given by:

$$EI : BMR > PAL \times \exp\left[SD_{min} \times \frac{(S/100)}{\sqrt{n}}\right],$$

where PAL is the assumed average PAL for the population under study, SD_{min} is –2 for 95 % or –3 for 99 % confidence limits, and $S = \sqrt{[CV_{Iw}^2/d + CV_B^2 + CV_P^2]}$, where CV_{Iw} is the within-subject variation in EI, *d* is the number of days of diet assessment (where the assessment purports to measure habitual intake e.g. diet history, then $d = \infty$ but 28 is used in practice), CV_B is the variation in repeated BMR measurements or the precision of estimated compared with measured BMR, CV_P is the between-subject variation in PAL. The values used to calculate the Goldberg cut-off were CV_I 20·8 %, CV_{Best} 8 %, CV_{Bmeas} 4 %, CV_P 15 %, *n* 48, *d* 16 (the cut-off does not change when *d* is increased to 28 d for the diet history) (Black, 2000*a*, AE Black, unpublished results). The measured PAL was 1·65. The calculated Goldberg cut-off values for the group means for EI : BMR_{meas} and EI : BMR_{est} respectively were 1·58 and 1·57 for forty-eight subjects, 1·52 and 1·51 for sixteen subjects. For identifying individual under-reporters (*n* 1) they were 1·19 and 1·16 respectively.

Statistical techniques

Differences between observers were tested with Student's unpaired *t* test, and differences between techniques on the same subjects by paired *t* test. The confidence limits of agreement between weighed record and diet history in Fig. 2 were calculated as the combined error derived from the within-subject SD of daily intake. The SD derived from the 16 d weighed record was also used for the diet history which was assumed to be equivalent to a 28 d record. The confidence limits of agreement between intake and validator in Figs. 4 and 6 were calculated as the combined

Table 1. Energy and nitrogen intakes estimated by diet histories in two groups of women by different observers together with anthropometric characteristics and nutrient intake by 16 d weighed diet records in each group

(Mean values and standard deviations)

	Observer A. B.		Observer A. W.	
	Mean	SD	Mean	SD
n	24		24	
Age (years)	58·1	4·5	56·4	5·4
Height (m)	1·65	0·06	1·63	0·07
Weight (kg)	63·6	8·4	66·1	9·5
BMI	23·3	2·7	24·9	3·5
16 d weighed records				
Energy (MJ)	7·98	1·40	8·19	1·69
Nitrogen (g)	11·2	1·9	11·2	2·3
Diet histories				
Energy (MJ)	7·73	1·45	8·68	2·13
Nitrogen (g)	11·1	2·1	11·9	2·7

Table 2. Anthropometric characteristics, energy expenditure (EE) and 24 h urinary nitrogen excretion in forty-eight women and in a subgroup of sixteen women in whom EE was measured using doubly-labelled water (DLW)

(Mean values and standard deviations)

	Whole study (*n* 48)		DLW subjects (*n* 16)	
	Mean	SD	Mean	SD
Age (years)	57·4	5·0	57·5	4·6
Weight (kg)	64·9	9·0	68·7	9·8
Change in wt over year of study (kg)	–0·37	1·9	0·4	2·2
Height (m)	1·64	0·07	1·66	0·07
BMI (kg/m^2)	24·1	3·2	25·1	4·2
Urine N (g)	9·90	1·86	10·67	2·10
Protein equivalent* (g)	74·4	11·7	79·2	13·2
EE (MJ)			9·42	1·82
BMR_{est}†	5·59	0·37	5·69	0·43
PAL_{est} (EE : BMR_{est})			1·65	0·28
BMR_{meas}‡			5·66	0·66
PAL_{meas} (EE : BMR_{meas})			1·65	0·27

PAL, physical activity level.
* Estimated as (urine N + 2) × 6·25 g.
† BMR estimated from equations.
‡ Measured BMR.

Table 3. Daily energy intakes (EI) and nitrogen intakes (NI) measured by weighed records and diet history and validations in a group of forty-eight women, and in a subgroup of sixteen women in whom energy expenditure (EE) was measured using doubly-labelled water (DLW)

(Mean values and standard deviations)

	Weighed records		Diet history	
Measurement	Mean	SD	Mean	SD
All subjects (*n* 48)				
NI (g)	11·21	2·09	11·47	2·40
Urine N : NI	0·89	0·11	0·89	0·20
Protein intake (g)	69·8	12·7	71·4	15·3
Protein intake : protein equivalent*	0·94	0·10	0·98	0·22
EI (MJ)	8·08	1·54	8·20	1·86
EI : BMR_{est}†	1·45	0·27	1·47	0·32
DLW subjects (*n* 16)				
NI (g)	12·00	2·44	12·70	2·75
Urine N : NI	0·90	0·12	0·86	0·20
Protein intake (g)	74·6	14·6	79·9	17·7
Protein intake : protein equivalent*	0·94	0·11	1·03	0·26
EI (MJ)	8·27	1·59	8·99	1·78
EI – EE	–1·15	1·75	–0·43	2·40
EI : EE	0·89	0·17	0·98	0·27
EI : BMR_{est}†	1·46	0·31	1·59	0·34
EI : BMR_{meas}‡	1·46	0·23	1·60	0·37

* (Urine N + 2) × 6·25 g.
† BMR estimated from equations.
‡ Measured BMR.

error derived from the within-subject SD for repeat DLW EE measurements (Black, 2000*a*), daily EI, daily NI, or daily N excretion and the given number of days (16 d weighed record, 28 d diet history and eight urine collections).

All protocols were approved by the Dunn Nutrition Centre Ethics Committee. The work was funded by the Medical Research Council and the Ministry of Agriculture, Fisheries and Food.

Results

Observer differences in diet histories

The anthropometric characteristics, EI and NI of the subjects studied by each observer are shown in Table 1. There were no significant differences in either anthropometric characteristics, or EI and NI from 16 d weighed records showing that the attempt to obtain two equivalent groups was successful. From the diet history, A. B. obtained estimates

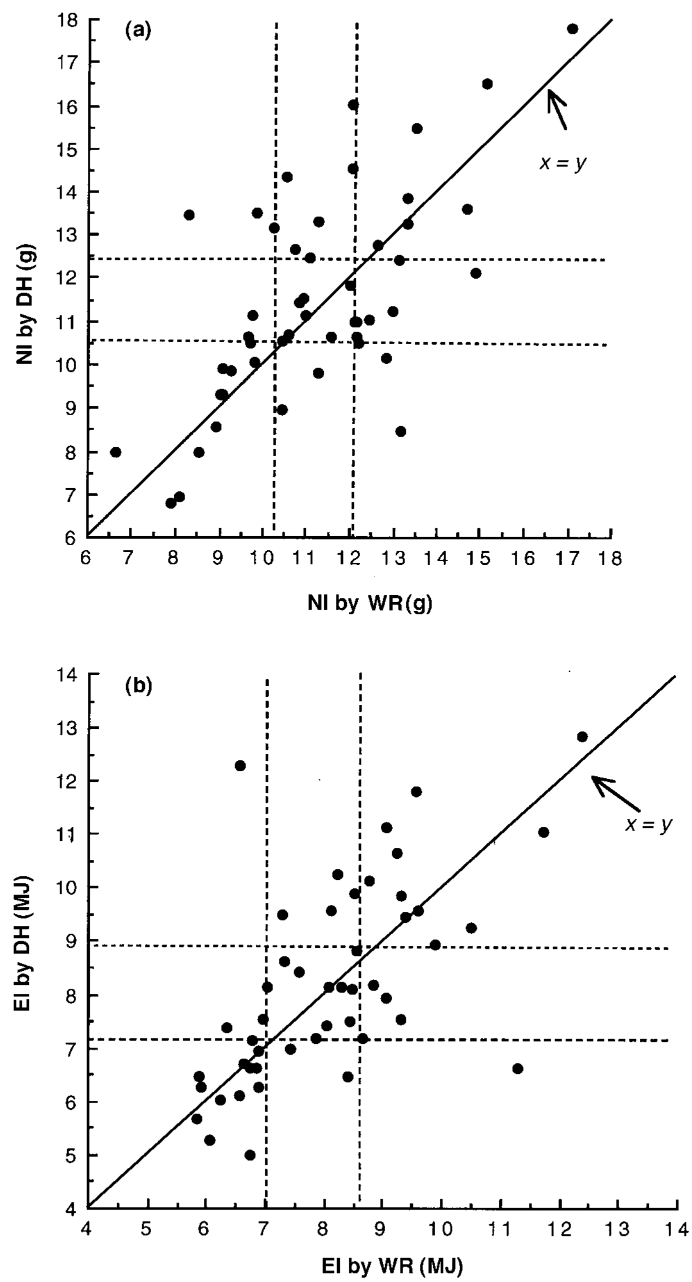

Fig. 1. Comparison of daily nutrient intakes by 16 d weighed record (WR) and by diet history (DH) in forty-eight women aged 50–65 years. The dotted lines divide intake according to thirds of the distribution. (a) Nitrogen intake (NI) (*r* 0·66, *P* < 0·001). (b) Energy intake (EI) (*r* 0·62, *P* < 0·001).

of intake that were lower and A. W. estimates of intake that were higher than those from the weighed records, but the differences between the observers were not statistically significant. The data were combined for further analysis.

Validation and comparison of methods at the group level

Table 2 shows the anthropometric characteristics, and the EE, urinary N excretion and its protein equivalent for all subjects and the subgroup with DLW measurements.

Table 3 shows the NI, protein intake and EI and the validation ratios for the 16 d weighed records and the diet history. There were no statistically significant differences between the weighed records and the diet history for any of the measurements shown. The validation ratio urine N : NI was 0·89 by both methods compared with the expected value of 0·81 indicating some bias to under-reporting of protein by both methods and no difference between them. When expressed as protein intake : protein equivalent, the values of 0·94 and 0·98, compared with an expected value of 1·00, suggest that protein was only slightly under-reported. The ratio EI : BMR_{est} was 1·45 for weighed records and 1·47 for diet history compared with an expected value of 1·65 and a Goldberg cut-off value for the group of 1·57, indicating bias to under-reporting of EI by both methods and no difference between them.

In the sixteen subjects with DLW measurements, the mean difference between EI and EE was −1·15 for weighed records and −0·43 for diet history, or under-reporting of 12·2 % and 4·6 %. The ratio EI : EE was 0·89 for weighed records and 0·98 for diet history compared with an expected value of 1·00, indicating under-reporting of 11 % and 2 % respectively. (The difference between these values and the percentages derived from EI − EE are due to the fact that in individuals similar values for EI − EE give different EI : EE depending on total EE.) EI : BMR_{meas} was 1·46 for weighed records, and 1·60 for diet history compared with a PAL_{meas} of 1·65 and a Goldberg cut-off value of 1·52. On this basis the weighed records mean intake would be regarded as biased to under-reporting and the mean intake from diet history as valid within the limits of the methodology. In this small group the diet history appeared to obtain a better mean reported intake than the weighed records, although the differences were not statistically significant. All three validation measures suggest a valid mean intake and no bias to under-reporting by the diet history.

Comparison of weighed record and diet history at the individual level

Fig. 1 shows the comparison between intakes as estimated by weighed records and by diet history for N (Fig. 1(a)) and energy (Fig. 1(b)). The figures show the $x = y$ line and the lines that divide each distribution into thirds. The correlations were 0·66 for N and 0·62 for energy ($P < 0·001$) suggesting that both methods tend to rank individuals similarly. However, twenty-two subjects out of forty-eight (46 %) for N and fourteen subjects out of forty-eight (29 %) for energy were not classified in the same third of the distribution, although only five (N) and two (energy) were placed in the opposite thirds.

Fig. 2 shows the Bland–Altman plot of differences between the methods against the mean of the two methods. Mean differences between weighed records and diet history were −0·26 (SD 1·88) g N and −0·12 (SD 1·50) MJ which were not significantly different from zero and indicated no bias between the two methods at the group level. At the individual level sixteen (33 %) (N) and eighteen (38 %) (energy) out of forty-eight subjects fell outside the 95 % confidence limits of agreement between the two methods. This could indicate either real differences in intake at

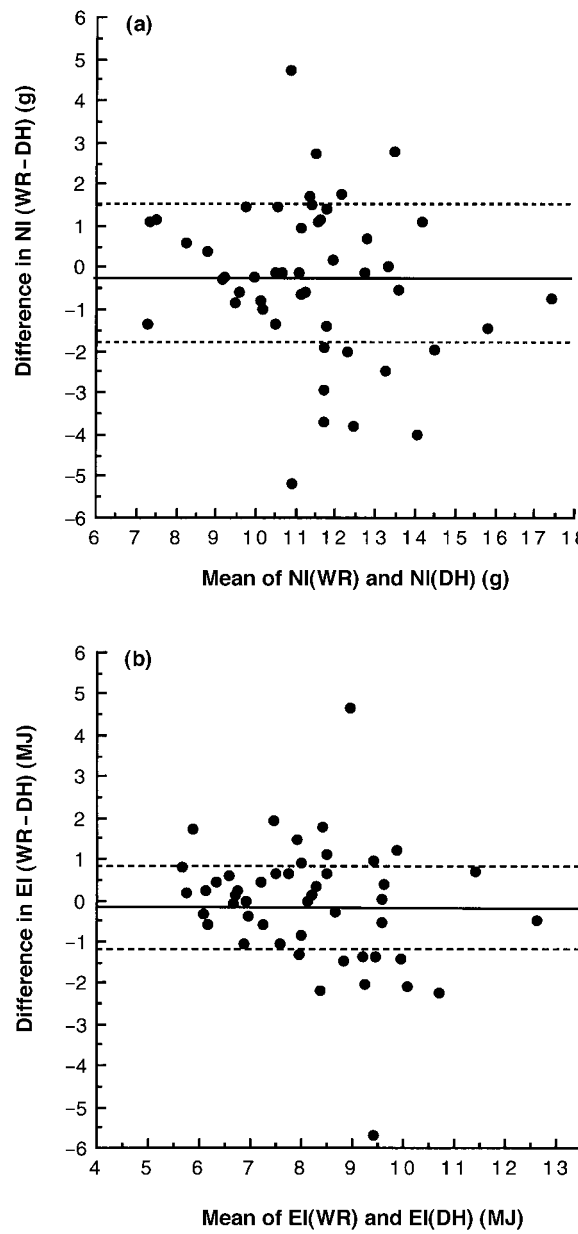

Fig. 2. Bland–Altman plot. Differences in daily intake measured by 16 d weighed records (WR) and the diet history (DH) against the mean of both measurements (*n* 48). (a) Nitrogen intake (NI), mean difference = −0·26 (SD 1·88) g N. (b) Energy intake (EI), mean difference = −0·12 (SD 1·50) MJ. (------), 95 % confidence limits of agreement between the two methods calculated from within-subject daily variation in intake, with the DH assumed to be equivalent to a 28 d record.

different time points since the assessments were not done concurrently, or poor agreement between the methods. Alternatively, it could indicate an underestimation of the errors and hence of the confidence limits of the measurements. Figs. 1 and 2 show that the two methods gave similar results at the group level as used by these observers and in this population. However, agreement at the individual level was sometimes poor.

Validation of weighed record and diet history for ranking individuals

Fig. 3 shows the relationship between urine N and NI by each technique. The correlation between urine N and NI was *r* 0·81 for weighed records and *r* 0·38 for diet history indicating that ranking of subjects was closer to true ranking (as estimated by urinary excretion) by weighed records than by diet history. The Bland–Altman plots of individual differences between the methods (Fig. 4) gave a similar picture. The mean difference between reported N intakes and measured N excretion was 1·31 (SD 1·25) g N for weighed records and 1·57 (SD 2·41) g N for diet history (NS). The mean differences were not statistically significantly different but the larger standard deviation for diet history indicated poorer ranking of individuals.

Figs. 5 and 6 shows the limited validation against DLW (*n* 16). The correlation between measured EE and reported EI was 0·48 for weighed records and 0·11 for diet history

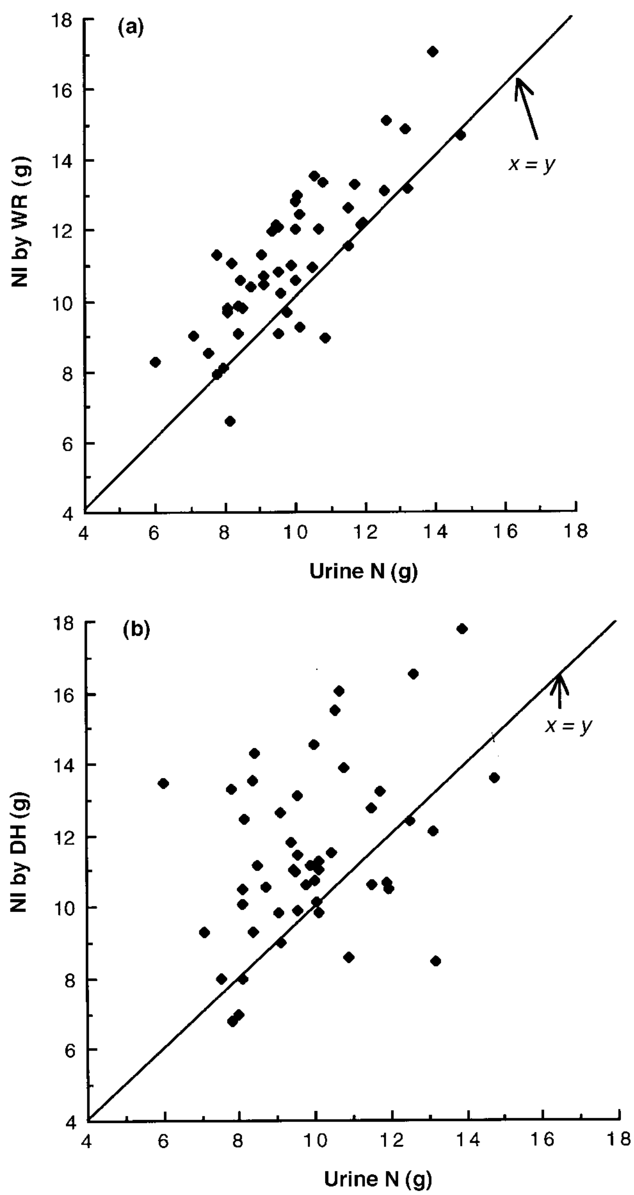

Fig. 3. Comparison of daily nitrogen intake (NI) with 24 h urinary N excretion in forty-eight women aged 50–65 years obtained from eight 24 h collections. (a) 16 d weighed records (WR), *r* 0·81. (b) Diet history (DH), *r* 0·38.

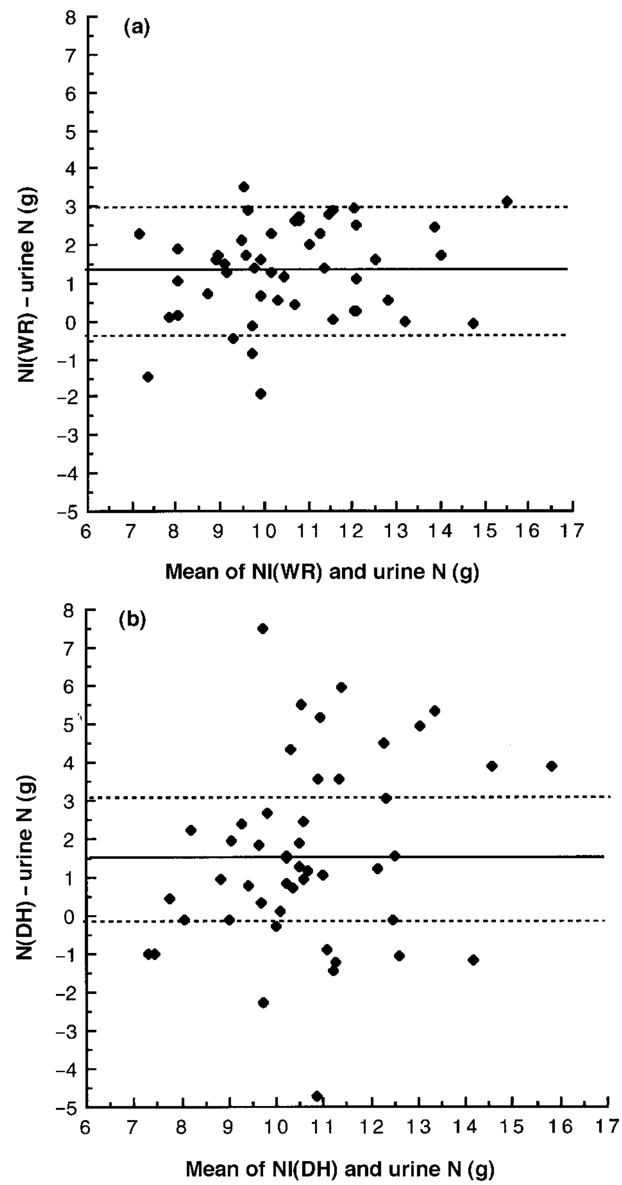

Fig. 4. Bland–Altman plot. Difference between daily nitrogen intake (NI) and 24 h nitrogen excretion plotted against the mean of the two measurements (*n* 48). (a) Weighed records (WR), mean difference = 1.31 (SD 1·25) g. (b) Diet history (DH), mean difference = 1·57 (SD 2·41) g. (------), 95 % confidence limits of agreement between the two methods calculated from within-subject daily variation in intake and excretion.

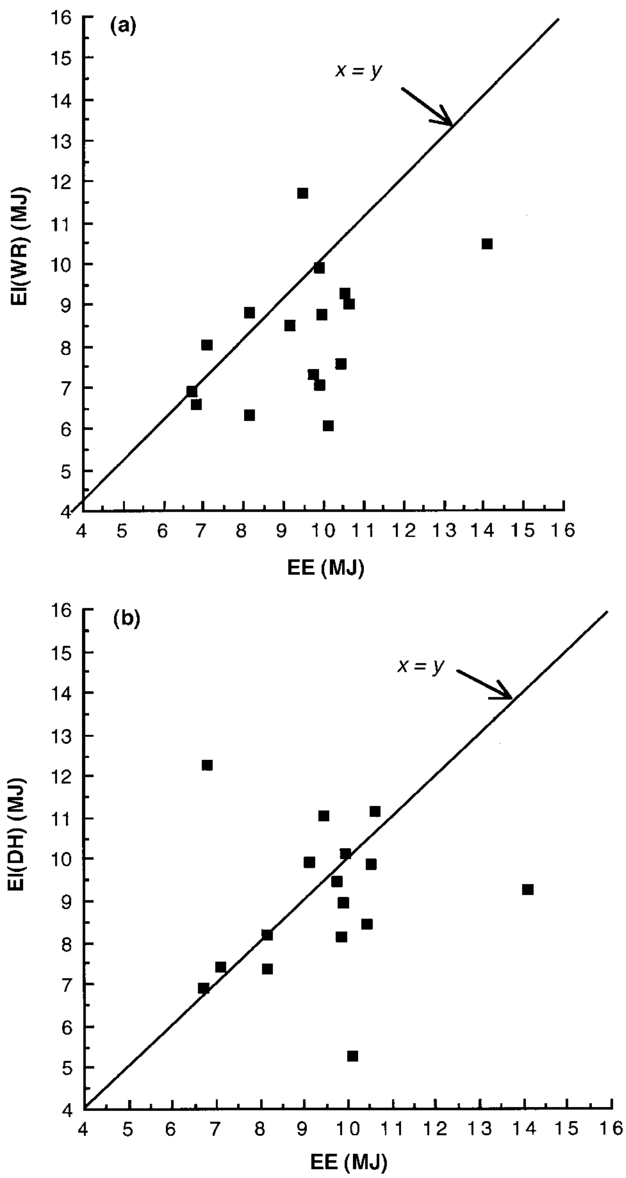

Fig. 5. Comparison of energy intake (EI) in sixteen women measured by two different methods with energy expenditure (EE) measured by doubly-labelled water. (a) 16 d weighed record (WR), *r* 0·48. (b) Diet history (DH), *r* 0·11.

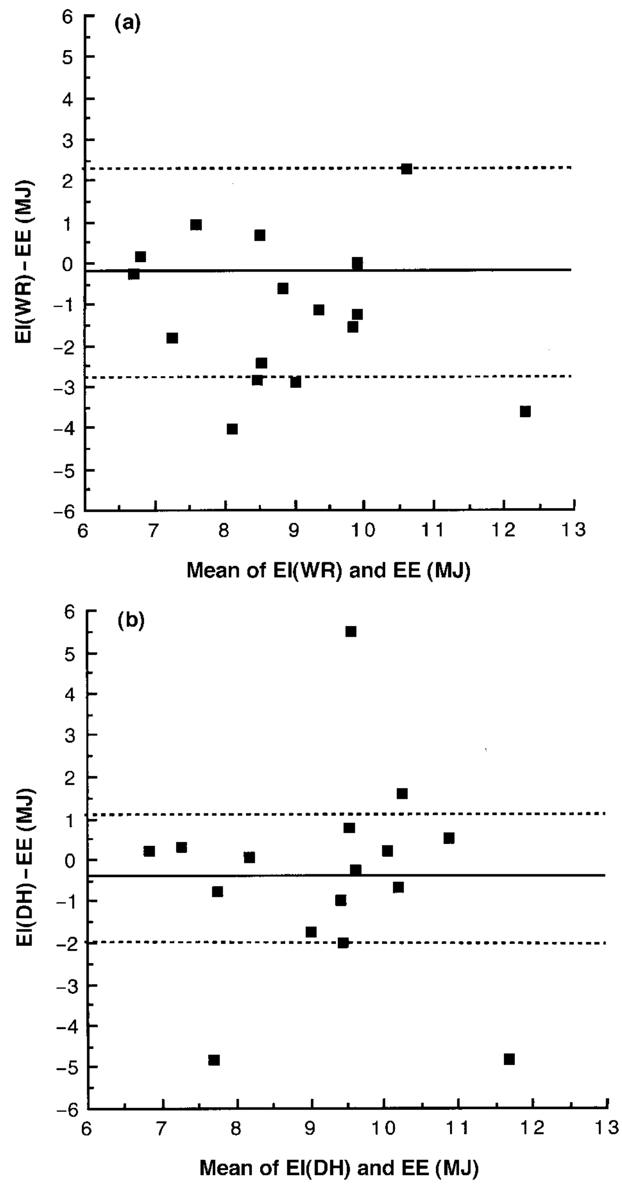

Fig. 6. Bland–Altman plot. Difference between daily energy intake (EI) measured by two different methods, and energy expenditure (EE) measured by the doubly-labelled water (DLW) method plotted against the mean of the two measurements for sixteen women aged 50–65 years. (a) Weighed records (WR), mean difference = −1·15 (SD 1·75) MJ. (b) Diet history (DH), mean difference = −0·43 (SD 2·40) MJ. (------), 95 % confidence limits of agreement between the two methods calculated from within-subject daily variation in EI and repeat measures of DLW EE.

data, indicating better ranking of individuals by the weighed record method. Mean differences were −1·147 (SD 1·75) MJ for weighed records and −0·43 (SD 2·40) MJ for diet history (NS). The mean differences were not statistically significantly different but the larger standard deviation for diet history indicated poorer ranking of individuals.

Comparison with other methods of dietary assessment

In the main study from which the participants in the present study were drawn (Bingham *et al.* 1994), the Oxford FFQ obtained an EI that was significantly higher than the weighed record, 7 d estimated record or 24 h recall. However, in the present study of forty-eight subjects in whom the diet history was completed, there were no significant differences in mean EI between any of five methods. EI (MJ) in the forty-eight subjects were 8·1 (SD 1·5) for weighed record, 8·2 (SD 1·9) for diet history, 8·6 (SD 1·6) for the Oxford FFQ, 8·4 (SD 1·6) for the 7 d estimated record and 8·1 (SD 1·9) for the unstructured 24 h recall.

Table 4 shows the Pearson correlation coefficients for urinary N *v.* NI and for DLW EE *v.* EI from each method. Values are shown for the total sample in the main study, the diet history subgroup and the DLW subgroup.

The pattern of the validation against urinary N was broadly similar for the whole study and for the two subgroups. The

Table 4. Pearson correlation coefficients between urinary nitrogen (UN) excretion and nitrogen intake (NI) measured by five different methods in women aged 50–65 years and in subgroups who provided a diet history and had their energy expenditure measured by doubly-labelled water (DLW), together with Pearson correlation coefficients between energy expenditure (EE) measured by DLW and energy intake (EI) measured by five different methods

	UN *v.* NI			
Method*	All subjects	Diet history subgroup (*n* 48)	DLW subgroup (*n* 16)	EI *v.* EE by DLW (*n* 16)
16 d weighed record	0·71 (*n* 155)	0·81	0·83	0·48
Diet history		0·38	0·43	0·11
7 d estimated diet record	0·69 (*n* 80)	0·58 (*n* 24)	0·65	0·24
Oxford FFQ	0·25 (*n* 137)	0·37 (*n* 46)	0·39 (*n* 14)	0·45
24 h recall (unstructured)	0·10 (*n* 155)	0·39	0·53	0·44

FFQ, food-frequency questionnaire.
* For details of methods, see Bingham *et al.* (1994).

correlation between intake and excretion was best for 16 d weighed records, followed by 7 d estimated diet records, and then FFQ. The correlations for 24 h recall improved as the sample became more highly selected: the DLW subjects were selected to represent the full range of urine N : NI values. The correlations for the diet history were similar to those for the Oxford FFQ in both subgroups. In the subgroup with measured EE, the pattern was reversed in that the FFQ and 24 h recall performed similarly to the 16 d weighed record, and the 7 d estimated record and the diet history performed poorly. However, with only sixteen subjects, none of the correlations was significant and it was not possible to draw conclusions about the relative ability of these methods to rank subjects correctly for energy.

Comparison of validation techniques

Fig. 7 shows the relationship between urinary N excretion and total EE. The correlation was 0·68. Fig. 8(a) shows the relationship between the validation ratios EI : EE and

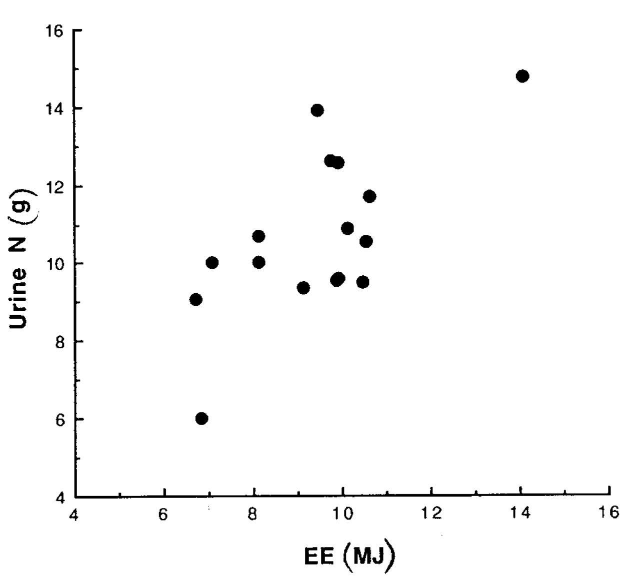

Fig. 7. Relationshp between urinary nitrogen excretion and total energy expenditure (EE) in sixteen women aged 50–65 years in whom EE was measured by the doubly-labelled water method.

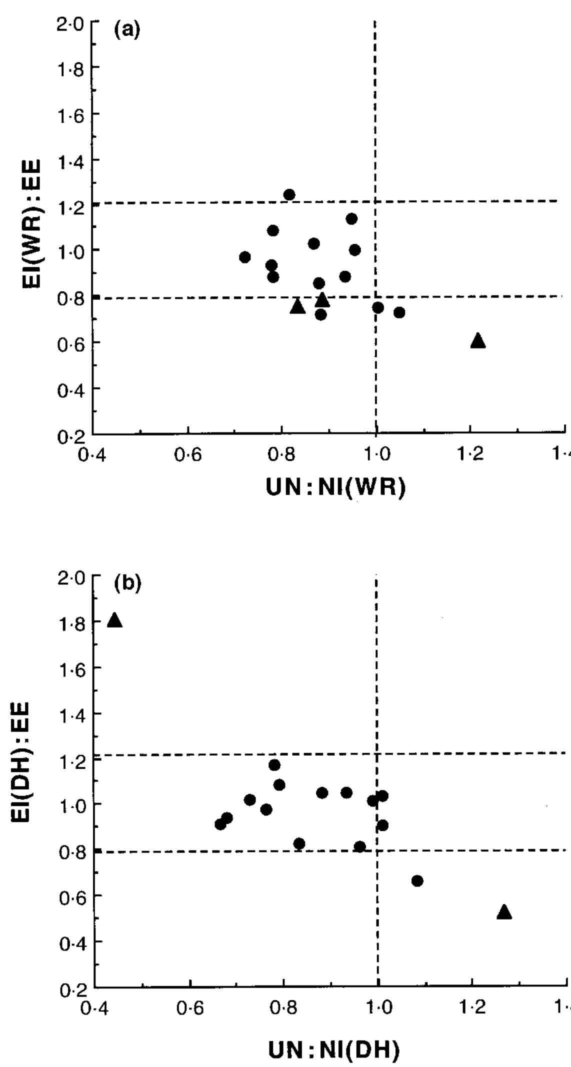

Fig. 8. Relationships between the validation ratios urine nitrogen (UN) : nitrogen intake (NI) and energy intake (EI) : energy expenditure (EE) in sixteen subjects for whom measurements of EE by the doubly-labelled water method were available. (a) Validation of the weighed record (WR) method. (b) Validation of the diet history (DH) method. Horizontal lines define the cut-off values for under-reporters (EI : EE < 0·79) and over-reporters (EI : EE > 1·21). The vertical line defines the cut-off value for under-reporters (UN : NI > 1·0).

Table 5. Externally validated studies of the diet history (DH) method from the published literature

Reference	Subjects	Age (years)	Diet method	EI : BMR	EI : EE	PrI : PrEq†	EI *v.* EE (*r*)	UN *v.* NI (*r*)	Notes
Livingstone *et al.* (1992)	20 MF	3 & 5	DH		1·13				EE measured by DLW
	24 MF	7 & 9			1·10		0·42***		
	12 MF	12			1·14		(*n* 80)		
	24 MF	15 & 18			0·99				
	24 MF	7 & 9	7 d WR		1·03				
	12 MF	12			0·89		0·77***		
	24 MF	15 & 18			0·76		(*n* 60)		
Petersen *et al.* (1992)	24 M 9 F	20–31	DH		1·15	1·07			1 × 24 h urine, PABA validated
	19 M 1 F		4 d WR		1·07	0·96			EE = EI to maintain weight
Heitmann (1993)	152 M	35–65	DH (1 month)	1·40		0·88			1 × 24 h urine, PABA validated
	171 F			1·27		0·82			
Lindroos *et al.* (1993)	22 M 23 F		DH	1·48	1·04	1·16	0·32*	0·26	1 × 24 h urine unvalidated
	obese		4 d WR	1·10	0·77	0·86	0·29	0·40**	EE estimated from PAQ
	8 M 11 F		DH	1·49	1·04	0·90	–0·01	0·56*	
	non-obese		4 d WR	1·46	1·02	0·88	0·43	0·80***	
Visser *et al.* (1995)	12 F	69–82	DH (1 month)		0·86	0·87			2 × 24 h urine unvalidated at home + 2 × 24 h urine in calorimeter. Volumes similar in both situations. EE measured in a calorimeter and by PAQ
Körtzinger *et al.* (1997)	50	26–38	DH	1·56	0·90				EE measured by PAQ
			WR	1·46	0·83				
Rothenberg *et al.* (1998)	9 F 3 M	70	DH	1·49	0·87				EE measured by DLW
									PAL = 1·71
Black *et al.*	48 F		DH	1·47		0·98		0·38**	5 to 8 × 24 h urine, PABA validated
(present study)			WR	1·45		0·94		0·81***	
Black *et al.*	16 F		DH	1·60	0·98	1·03	0·11		EE measured by DLW
(present study)			WR	1·46	0·89	0·94	0·48*		PAL = 1·67

EI, energy intake; EE, energy expenditure; PrI, protein intake; PrEq, protein equivalent; UN, urine N; NI, N intake; M, males; F, females; WR, weighed record; DLW, doubly-labelled water; PABA, *p*-amino benzoic acid; PAQ, physical activity questionnaire; PAL, physical activity level.

* $P < 0{\cdot}05$, ** $P < 0{\cdot}01$, *** $P < 0{\cdot}001$.

† PrEq = PI estimated as (urine N + 2) × 6·25 g.

urinary UN : NI for the weighed record. Three subjects were identified as under-reporters by both validators, three by energy only and one as an over-reporter by energy. The three triangles show subjects with EI : BMRo below the Goldberg cut-off. Fig. 8(b) shows the relationship between the validation ratios for diet history. Two subjects were identified as under-reporters by both techniques and one as an over-reporter. Two subjects (triangles) were respectively below and above the Goldberg cut-off values for EI : BMRo.

Discussion

The present study of forty-eight women aged 50–65 years demonstrated under-reporting by both the diet history and the weighed record techniques. If the PAL of 1·65 for the subgroup with DLW measurements also applied to the whole group, then the EI : BMR_{est} values of 1·45 and 1·47 indicate under-reporting of energy by both techniques. The values for protein intake : protein equivalent of 0·94 and 0·98 suggest that protein was more completely reported than energy. In the sixteen subjects with DLW measurements, however, there was a difference between the techniques. In this subgroup the weighed record led to under-reporting and the diet history provided a valid measurement. The selection of subjects for the diet history and DLW measurements was dependent on them agreeing to additional investigations at the end of a long study. Subjects were therefore selected for compliance or interest in the results. They had also completed repeated dietary assessments and it is possible that this ‘training’ enabled the diet history to obtain better results than might be expected from a single isolated assessment.

The external validation against urinary N and EE should determine whether the weighed record or the diet history gives better measures of individual intakes and ranking of subjects. The standard deviations of the differences between the intakes and the validators (energy and N) were greater for diet history than for weighed record indicating greater discrepancies at the individual level. Correlation between urine N and NI was 0·80 for 16 d weighed records and 0·38 for diet history suggesting that the diet history was less able to rank individuals correctly. It might be argued that the poorer agreement for the diet history was because urine N and NI were not measured concurrently, rather than due to errors of the diet history technique. However, urine N appears to be a useful validator even when separated in time from intake measurements. Bingham *et al.* (1997) found a correlation of 0·50 between the last 8 d of weighed records (seasons 3 and 4) and the first single urine specimen collected 6 or 9 months earlier in season 1. Further, in the present study, the correlation of EE with EI from weighed records (not measured concurrently) was 0·48, but that for EE and EI by diet history (measured within 2 weeks) was only 0·11.

It could also be argued that the urine N was derived from only eight 24 h collections whereas the diet history is deemed to measure ‘habitual’ intake. However, it is doubtful if the diet history does measure ‘habitual’ intake. ‘Habitual’ intake is a theoretical concept, i.e. the intake, averaged over a prolonged period, that maintains body weight. Intake is enormously variable in the short term and has also been shown to vary with time (Tarasuk & Beaton, 1991). It undoubtedly varies with season and with periods of dieting to lose weight. Reported intake from retrospective techniques is more strongly influenced by recent than by distant intake (van Staveren *et al*. 1986). Beaton (1991) has also suggested that the errors of the diet history are similar to the errors of a 7 d record.

In their comparison of several dietary techniques, Bingham *et al*. (1995) found that the weighed record outperformed all others in ranking individuals as judged by the correlation coefficient between urine N and NI. The conclusions were the same whether all 8 d of urinary N were used for the comparison or whether each method was correlated only with the urinary N measurements from the same season in which the assessment was done (Bingham *et al*. 1997). The present work (Table 4) added the diet history to the methods evaluated. The external validation against urinary N excretion showed all retrospective questionnaire techniques including the diet history to perform less well than the prospective recording techniques. Retrospective questionnaires require the subject to mentally integrate complicated eating patterns. Further, the way food is perceived by nutritionists, which influences the structure of questionnaires, may not relate well to the non-nutritionist’s perception of food. Prospective techniques require simply that food is noted as eaten. While open to omissions and distortions and certainly subject to severe under-reporting by some subjects, these may nevertheless obtain the most valid representation of food intake.

Table 5 summarizes the results of other externally validated studies of the diet history technique. Where the weighed record was also studied, the relative validity showed that the diet history obtained a more valid mean intake than the weighed record. However, there is no clear picture of the absolute validity of the diet history. The EI : BMR values ranged from 1·10 to 1·56; all indicate an element of under-reporting. EI : EE measures indicated over-reporting in two studies (Livingstone *et al.* 1992; Petersen *et al*. 1992), valid mean intakes in two studies (Lindroos *et al*. 1993; present study) and under-reporting in four (Heitmann, 1993; Visser *et al*. 1995; Körtzinger *et al*. 1997; Rothenberg *et al*. 1998). The validations by N excretion, expressed as protein intake : protein equivalent to facilitate comparisons, are in broad agreement with those of EI : EE, again showing that the absolute validity of the technique varied between studies. Correlations between intakes and validators were worse for diet history than for weighed record with the exception of the obese subjects studied by Lindroos *et al*. (1993), where the weighed record also had a poor correlation.

It is difficult to assess the absolute validity of the diet history. The studies listed in Table 5, with the exception of that from Heitmann (1993), were on small numbers and the results were not consistent. The diet history is not a standardized technique. It is the most open-ended of the retrospective interview-based survey techniques. It is strongly dependent on the communication skills and experience of the interviewer. Of all techniques, it is most open to variations in interviewer phraseology, intonation and body language. It may be conducted using a structured questionnaire, which is inevitably lengthy, probably repetitive, may include elements irrelevant to individual subjects and will differ from study to study. Alternatively the interview may be unstructured, and supposedly tailored to the individual, thus avoiding unnecessary questions and repetitions but risking the omission of important aspects of the diet. The interview usually takes a minimum of 1 h and both interviewer and subject may lose concentration before all necessary information has been obtained. Subjects are asked to undertake the very difficult task of integrating dietary patterns over a long time span. Answers depend on subjects' memories and perceptions of dietary patterns. It is probably more difficult for subjects of limited education; there is evidence, for example, of poorer validity of 24 h recalls in subjects with lower literacy scores (Johnson *et al*. 1998). It is probably more difficult for the elderly; there is evidence of poorer performance in FFQ by older women (Sawaya *et al*. 1996). It is certainly more difficult for both interviewer and subject if eating patterns are irregular. Some of these many sources of variation may account for the lack of consistency in comparative studies of diet records and diet histories. Certainly conclusions from one study are not necessarily transferable to another.

The present study attempted to evaluate observer differences in measuring dietary intake. There was a (non-significant) difference between A. B. and A. W. in reported intake, but, since the circumstances of the study did not permit repeated diet histories by both observers on the same subjects, it was uncertain whether this was due to observer differences, true difference in intakes, the errors of the technique, or differential biases in reporting. However, a subject-specific response to dietary assessment has been demonstrated (AE Black, unpublished results), making it probable that differential biases in reporting operated similarly on the diet records that were the basis for selecting subjects and on the subsequent diet histories. Further, in a previous study A. B. obtained reported intakes lower than those of another observer. This suggests that observer differences can contribute to the total errors of the diet history.

Comparison of validation techniques

Urinary N excretion and DLW EE are both established validators for reported dietary intake. Can they be used interchangeably? The present study found a correlation of 0·68 between urinary N and DLW EE, but excluding two extreme values reduced the correlation to 0·32 (NS). In a larger body of data (including the present data) (Black *et al*. 1997), the correlation between urinary N excretion and DLW EE was 0·36 (n 45) ($P < 0{\cdot}05$), but that between urinary N excretion and protein intake was 0·80 (excluding under-reporters). Thus, these methods are not completely interchangeable. Validation against urinary N excretion must be the preferred method if protein, or nutrients whose intake is highly correlated with protein, are the variables of interest. Validation for energy must be the preferred method if total energy, or the major energy contributing macronutrients, are the variables of interest. The criterion for under-reporting (EI : EE $< 0{\cdot}79$) tends to identify a higher proportion of under-reporters than the criterion urinary N : NI $> 1{\cdot}0$. It is uncertain whether this is due to inherent differences in the criteria or to genuine differences in the reporting of protein and energy (Black *et al*. 1997). As an alternative to measured EE, EI : BMR has limitations as a validator. The confidence limits of the Goldberg cut-off are wider than those for the direct validators and cannot identify as many under-reporters. Further, the sensitivity of the Goldberg cut-off for identifying under-reporters among those with high energy requirements is limited (Black, 2000*b*).

In conclusion, the present study was unable to establish the errors of the diet history or to separate methodological errors from imprecision due to normal variation in intake. However, it found that in middle-aged women mean intakes reported by the diet history did not differ from those of the weighed record, although in a subset of subjects it obtained a better mean intake. Although the diet history identified the extremes of intake equally well, the weighed record provided better ranking of individuals overall.

Acknowledgements

We thank Shirley Runswick and Malcolm Sawyer for technical help and all the women who participated in these studies for their patience and co-operation.

References

Ballard-Barbash R, Graubard I, Krebs-Smith SM, Schatzkin A & Thompson FE (1996) Contribution of dieting to the inverse association between energy intake and body mass index. *European Journal of Clinical Nutrition* **50**, 98–106.

Beaton GH (1991) Interpretation of results from diet history studies. In *The Diet History Method. Proceedings of the 2nd Berlin Meeting on Nutritional Epidemiology*, pp. 15–38 [L Kohlmeier, editor]. London: Smith-Gordon.

Bingham S (1987) The dietary assessment of individuals: methods, accuracy, new techniques and recommendations. *Nutrition Abstracts and Reviews* **57**, 705–742.

Bingham SA, Cassidy A, Cole TJ, Welch A, Runswick S, Black AE, Thurnham D, Bates CJ, Khaw KT, Key TJA & Day NE (1995) Validation of weighed records and other methods of dietary assessment using the 24 h urine nitrogen technique and other biological markers. *British Journal of Nutrition* **73**, 531–550.

Bingham S & Cummings JH (1983) The use of 4-amino benzoic acid as a marker to validate the completeness of 24-h urine collections in man. *Clinical Science* **64**, 629–635.

Bingham SA & Cummings JH (1985) Urine nitrogen as an independent validatory measure of dietary intake. *American Journal of Clinical Nutrition* **42**, 1276–1289.

Bingham SA, Gill C, Welch A, Cassidy A, Runswick S, Oakes S, Lubin R, Thurnham DI, Key TJA, Roe L, Khaw K-T & Day NE (1997) Validation of dietary assessment methods in the UK arm of EPIC using weighed records and 24-hour urinary nitrogen and potassium and serum vitamin C and carotenoids as biomarkers. *International Journal of Epidemiology* **26**, S137–S151.

Bingham SA, Gill C, Welch A, Day K, Cassiday A, Khaw KT, Sneyd MJ, Key TJA, Roe L & Day NE (1994) Comparison of dietary assessment methods in nutritional epidemiology: weighed records *v.* 24 h recalls, food frequency questionnaires and estimated diet records. *British Journal of Nutrition* **72**, 619–643.

Bingham SA, Wiggins HS, Englyst H, Seppänen R, Helms P, Strand R, Burton R, Jørgenson IM, Poulsen L, Paerregaard A, Bjerrum L & James WPT (1982) Methods and validity of dietary

assessments in four Scandinavian populations. *Nutrition and Cancer* **4**, 23–32.
Black AE (1996) Physical activity levels (PAL) from a meta-analysis of doubly labelled water studies for validating energy intake as measured by dietary assessment. *Nutrition Reviews* **54**, 170–174.
Black AE (1999) Small eaters or under-reporters? In *Progress in Obesity Research*, pp. 223–228 [G Ailhaud and B Guy-Grand, editors]. Paris: John Libbey.
Black AE (2000*a*) Within- and between-subject variation in energy expenditure measured by doubly-labelled water: implications for assessing the validity of reported dietary energy intake. *European Journal of Clinical Nutrition* (In the Press).
Black AE (2000*b*) The sensitivity and specificity of the Goldberg cut-off for EI : BMR for identifying dietary reports of poor validity. *European Journal of Clinical Nutrition* (In the Press).
Black AE, Bingham SA, Johansson G & Coward WA (1997) Validation of dietary intakes of protein and energy against 24 hour urinary N and DLW energy expenditure in middle aged women and retired men: comparison with validation against presumed energy requirements. *European Journal of Clinical Nutrition* **51**, 405–413.
Black AE, Coward WA, Cole TJ & Prentice AM (1996) Human energy expenditure in affluent societies: analysis of 574 doubly-labelled water measurements. *European Journal of Clinical Nutrition* **50**, 72–92.
Black AE, Goldberg GR, Jebb SA, Livingstone MBE & Prentice AM (1991) Critical evaluation of energy intake data using fundamental principles of energy physiology: 2. Evaluating the results of dietary surveys. *European Journal of Clinical Nutrition* **45**, 583–599.
Black AE, Prentice AM & Coward WA (1986) Use of food quotients to predict respiratory quotients for the doubly-labelled water method of measuring energy expenditure. *Human Nutrition: Clinical Nutrition* **40C**, 381–391.
Braam L, Ocke MC, Buenode Mesquita HB & Seidell JC (1998) Determinants of obesity-related underreporting of energy intake. *American Journal of Epidemiology* **147**, 1081–1086.
Braddon F, Wadsworth M, Davies J & Cripps H (1988) Social and regional differences in food and alcohol consumption and their measurement in a national birth cohort. *Journal of Epidemiology and Community Health* **42**, 341–349.
Briefel RR, Sempos CT, McDowell MA, Chien S & Alaimo K (1997) Dietary methods research in the third National Health and Nutrition Examination Survey: under-reporting of energy intake. *American Journal of Clinical Nutrition* **65**, Suppl., 1203S–1209S.
Coward WA (1988) The doubly labelled-water ($^2H_2{}^{18}O$) method: principles and practice. *Proceedings of the Nutrition Society* **47**, 209–218.
Crawley H (1988) *Food Portion Sizes*. London: H.M. Stationery Office.
Fogelholm M, Männisto S, Vartiainen E & Pietinen P (1996) Determinants of energy balance and overweight in Finland 1982 and 1992. *International Journal of Obesity* **20**, 1097–1104.
Gnardellis C, Boulou C & Trichopolou A (1998) Magnitude, determinants and impact of under-reporting of energy intake in a cohort study of Greece. *Public Health Nutrition* **1**, 131–137.
Goldberg GR, Black AE, Jebb SA, Cole TJ, Murgatroyd PR, Coward WA & Prentice AM (1991) Critical evaluation of energy intake data using fundamental principles of energy physiology. 1. Derivation of cut-off values to identify under-recording. *European Journal of Clinical Nutrition* **45**, 569–581.
Heitmann BL (1993) The influence of fatness, weight change, slimming history and other lifestyle variables on diet reporting in adult Danish men and women aged 35–65 years. *International Journal of Obesity* **17**, 329–336.
Heywood P, Harvey PJW & Marks GC (1993) An evaluation of energy intake in the 1983 Australian National Dietary Survey of Adults. *European Journal of Clinical Nutrition* **47**, 604–606.
Holland B, Unwin ID & Buss DH (1988) *Cereals and Cereal Products. Third Supplement to McCance and Widdowson's The Composition of Foods (4th edition)*. Nottingham and London: The Royal Society of Chemistry and Ministry of Agriculture, Fisheries and Food.
Holland B, Unwin ID & Buss DH (1989) *Milk Products and Eggs. Fourth Supplement to McCance and Widdowson's The Composition of Foods (4th edition)*. Cambridge and London: The Royal Society of Chemistry and Ministry of Agriculture, Fisheries and Food.
Hultén B, Bengtsson C & Isaksson B (1990) Some errors inherent in a longitudinal dietary survey revealed by the urine nitrogen test. *European Journal of Clinical Nutrition* **44**, 169–174.
Isaksson B (1980) Urinary nitrogen output as a validity test in dietary surveys. *American Journal of Clinical Nutrition* **33**, 4–5.
Johnson RK, Soultanakis RP & Matthews DE (1998) Literacy and body fatness are associated with underreporting of energy intake in US low-income women using the multiple-pass 24-hour recall; a doubly labeled water study. *Journal of the American Dietetic Association* **98**, 1136–1140.
Klesges RC, Eck LH & Ray JW (1995) Who underreports dietary intake in dietary recall? Evidence from the second National Health and Nutrition Examination Survey. *Journal of Consulting and Clinical Psychology* **63**, 438–444.
Körtzinger I, Bierwag A, Mast M & Müller MJ (1997) Dietary underreporting: validity of dietary measurements of energy intake using a 7-day dietary record and a diet history in non-obese subjects. *Annals of Nutrition and Metabolism* **41**, 37–44.
Lafay L, Basdevant A, Charles M-A, Vray M, Balkau B, Borys J-M, Eschwege E & Romon M (1997) Determinants and nature of dietary underreporting in a free-living population: the Fleurbaix Laventie Ville-Santé (FLVS) Study. *International Journal of Obesity* **27**, 567–573.
Lindroos A-K, Lissner L & Sjöström L (1993) Validity and reproducibility of a self-administered dietary questionnaire in obese and non-obese subjects. *European Journal of Clinical Nutrition* **47**, 461–481.
Livingstone MBE, Prentice AM, Coward WA, Strain JJ, Black AE, Davies PSW, Stewart CM, McKenna PG & Whitehead RG (1992) Validation of estimates of energy intake by weighed dietary record and diet history in children and adolescents. *American Journal of Clinical Nutrition* **56**, 29–35.
Paul AA & Southgate DAT (1978) *McCance and Widdowson's The Composition of Foods*. London: H.M. Stationery Office.
Petersen MA, Haraldsdóttir J, Hansen HB, Jensen H & Sandström B (1992) A new simplified dietary history method for measuring intake of energy and macronutrients. *European Journal of Clinical Nutrition* **46**, 551–559.
Price GM, Paul AA, Cole TJ & Wadsworth MEJ (1997) Characteristics of the low-energy reporters in a longitudinal national dietary survey. *British Journal of Nutrition* **77**, 833–851.
Price GM, Paul AA, Key FB, Harter AC, Cole TJ, Day KC & Wadsworth MEJ (1996) Measurement of diet in a large national survey: comparison of computerized and manual coding of records in household measures. *Journal of Human Nutrition and Dietetics* **8**, 417–428.
Pryer J, Vrijheid M, Nichols R, Kiggins M & Elliott P (1997) Who are the 'Low energy reporters' in the dietary and nutritional survey of British adults? *International Journal of Epidemiology* **26**, 146–154.
Rothenberg E, Bosaeus I, Lernfelt B, Landahl S & Steen B (1998) Energy intake and expenditure: validation of a diet history by heart rate monitoring, activity diary and doubly labelled water. *European Journal of Clinical Nutrition* **52**, 832–838.

Rothenberg E, Bosaeus I & Steen B (1997) Evaluation of energy intake estimated by a diet history in three free-living 70 year old populations in Gothenburg, Sweden. *European Journal of Clinical Nutrition* **51**, 60–66.

Sawaya AL, Tucker K, Tsay R, Willett W, Saltzman E, Dallal GE & Roberts SB (1996) Evaluation of four methods for determining energy intake in young and older women: comparison with doubly labeled water measurements of total energy expenditure. *American Journal of Clinical Nutrition* **63**, 491–499.

Schofield W, Schofield C & James WPT (1985) Basal metabolic rate – review and prediction, together with an annotated bibliography of source material. *Human Nutrition: Clinical Nutrition* **39C**, Suppl. 1.

Steen B, Isaksson B & Svanborg A (1977) Intake of energy and nutrients and meal habits in 70 year old males and females in Gothenborg, Sweden: a population study. *Acta Medica Scandinavica* **611**, Suppl., 39–86.

Tarasuk V & Beaton GH (1991) The nature and individuality of within-subject variation in energy intake. *American Journal of Clinical Nutrition* **54**, 464–470.

van Staveren WA, de Boer JO & Burema J (1985) Validity and reproducibility of a dietary history method estimating the usual food intake during one month. *American Journal of Clinical Nutrition* **42**, 554–559.

van Staveren WA, West CE, Hoffmans M, Bos P, Kardinaal A, van Poppel GAF, Schipper HJ-A, Hautvast J & Hayes R (1986) Comparison of contemporaneous and retrospective estimates of food consumption made by a dietary history method. *American Journal of Epidemiology* **123**, 884–893.

Visser M, de Groot LCPMG, Deurenburg P & van Staveren WA (1995) Validation of dietary history method in a group of elderly women using measurements of total energy expenditure. *British Journal of Nutrition* **74**, 775–785.

Voss S, Kroke A, Klipstein Grobusch K & Boeing H (1998) Is macronutrient composition of dietary intake data affected by underreporting? Results from the EPIC-Potsdam study. *European Journal of Clinical Nutrition* **52**, 119–126.

Wiles SJ, Nettleton PA, Black AE & Paul AA (1980) The nutrient composition of some cooked dishes eaten in Britain: a supplementary food composition table. *Journal of Human Nutrition* **34**, 241–248.

Willett WC, Sampson L, Stampfer MJ, Rosner B, Bain C, Witsch J, Hennekens CH & Speizer FE (1985) Reproducibility and validity of a semiquantitative food frequency questionnaire. *American Journal of Epidemiology* **122**, 51–65.