

# A comparison of selected nutrient intakes derived from three diet assessment methods used in a low-fat maintenance trial

James R Hebert<sup>1,\*</sup>, Thomas G Hurley<sup>1</sup>, David E Chiriboga<sup>2</sup> and Jeanine Barone<sup>3</sup>

<sup>1</sup> Division of Preventive and Behavioral Medicine, Department of Medicine, University of Massachusetts Medical School, 55 Lake Avenue North, Worcester, MA 01655, USA; <sup>2</sup> Department of Surgery, University of Massachusetts Medical School, 55 Lake Avenue North, Worcester, MA 01655, USA; <sup>3</sup> Currently affiliated with the University of California at Berkeley Wellness Letter, 632 Broadway, New York, NY 10012, USA

Submitted 10 March 1998; Accepted 6 July 1998

## Abstract

**Objective:** In the vast majority of surveys and research in humans, dietary data are obtained from self-reports: recalls; records; or historical methods, usually food frequency questionnaires (FFQ). This study provides a rare opportunity to compare data derived from all three methods.

**Design:** A crossover study of dietary fat in which data were collected using an average of 11.4 food records and 11.7 24-h diet recalls. Using simple subtraction and correlation, energy and nutrient intakes derived from the three methods were compared to each other and with those derived from a single FFQ. Analysis of variance was used to evaluate sources of variability in nutrient intakes estimated from the individual days of records and recalls.

**Setting:** An independent, free-standing medical research institute.

**Subjects:** 13 men who were compliant with study procedures.

**Results:** FFQ-derived estimates of energy and nutrient intake were highest (e.g. 1967 kcal versus 1858 kcal and 1936 kcal for the records and recalls, respectively). Mean differences in energy and nutrient intakes and their variances were lowest and correlation coefficients highest in comparing the records and recalls (e.g. for fat the mean difference was 5.0 g, and  $r = 0.85$ ). Analysis of variance of individual days of record- and recall-derived data ( $n = 300$ ) revealed that there was no effect due to either method (record or recall) or the sequence of administration.

**Conclusions:** Results of this study indicate that the FFQ overestimated dietary intake. Energy and nutrient results obtained from the records and recalls were interchangeable. However, based on smaller SDs around the means, it appears that the recalls may perform slightly better in estimating dietary intake in groups such as these well-educated, highly compliant men.

**Keywords**  
Diet surveys  
Dietary assessment methods  
Nutrient intake  
Variability  
Validity  
Reliability  
Calibration

The most commonly used methods of dietary assessment in surveys and studies of humans are: records; recalls; and histories, such as the food frequency questionnaire (FFQ). With some rare exceptions, records and recalls focus on a single day or a 24-h period. Therefore, many days must be assessed unless the intent is to estimate group mean intake. By contrast, a single administration of the FFQ ostensibly estimates long-term average intake in an individual. These instruments vary dramatically in terms of cost and participant involvement, with important concomitants for the scale of the study that may be feasible and for response bias. Hence, understanding the statistical properties of these tools is important to making an informed decision about their use. Because short-term recalls and records are thought to represent current diet most accurately<sup>1–4</sup>, these instruments, when adminis-

tered over a sufficient number of days, also provide a means to evaluate the validity of other assessment tools, such as the FFQ, that are more feasible for large-scale research use.

In dietary research, validity refers to the ability of an assessment tool to provide an accurate estimate of food and nutrient intake<sup>5</sup>. Because diets of free-living people cannot be assessed without error, absolute validity cannot be judged<sup>6–9</sup>. Despite the absence of a 'gold standard', in studies comparing dietary assessment methods the term *validity* is taken to mean criterion validity and not content validity (which is assumed) or construct validity (which rarely is considered)<sup>6,10,11</sup>. This relative, or inter-method, validity is determined by comparing an assessment tool to a practical standard that is assumed to assess diet more accurately but which, inevitably, has its own strengths and weaknesses.

\*Corresponding author: E-mail James.Hebert@ummed.edu

Generally, multiple food records are considered to be most indicative of true dietary intake<sup>4</sup>; however, some studies have shown that FFQ and multiple 24-h recalls are superior<sup>1,12</sup>. Because these instruments are employed by humans, their performance would be expected to vary according to a number of factors including the ability and willingness to cooperate, motivation, changes in eating habits or misrepresentation (conscious or subconscious) of the true diet<sup>6,13–16</sup>.

Optimally, research studies that assess dietary intake should utilize more than one method for determining nutrient intakes<sup>1</sup>. However, the expense of dietary assessments, especially records and recalls, normally preclude this. For this study, we were fortunate to have access to a large number of records and recalls, and an FFQ administered in a group of highly motivated men. The purpose of this study is to compare nutrient intake estimates derived from the three assessment methods, with a special emphasis on the recall–records comparison, which is this dataset's unique attribute. Methods were compared in terms of total variance or overall 'error', the sources of variance and differences in average method-specific nutrient intake.

## Materials and methods

Diet was assessed in an intervention trial conducted to test the effect of specific dietary fats on the activity of natural killer cells in humans<sup>17–19</sup>. Briefly, the study consisted of a crossover design conducted over a 10-month period from June 1986 to March 1987 (Fig. 1). Participants became familiar with the low-fat (20% of energy as fat) eating plan during the initial 2-month run-in period. Following this, all participants were assigned to the low-fat intervention for 3 months. Participants were then assigned randomly to consume supplements of either coconut oil or safflower oil for 2 months. Following a 1-month washout period, the subjects crossed over to take the other oil supplement for 2 months.

Subjects were recruited in New York City from flyers and newspaper advertisements. Eligible subjects had to be male, non-smoking, not overweight (within 10% of ideal body weight), neither currently engaging nor

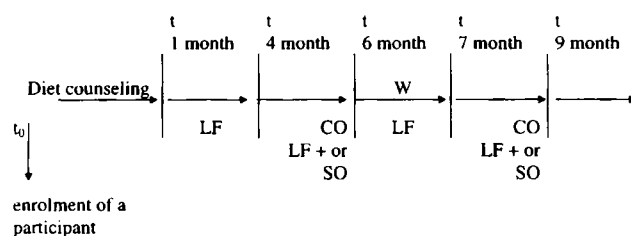
planning to engage in vigorous physical exercise, and between the ages of 21 and 39 years. In addition, no participant had allergies, and each was healthy, as determined by a medical history. The men were instructed not to attempt to lose weight or begin using nutritional supplements or medications for the entire study period. Informed consent was obtained from each participant. All enrolment and data collection procedures were approved by the Institutional Review Board of the American Health Foundation.

During the study run-in period, a trained nutritionist spent 60–90 min counselling subjects about effective steps for reducing dietary fat to ~20% of energy. At that time, examples of low-fat menus and other nutrition education materials were distributed and discussed. Subjects were taught how to keep a food record by a trained nutritionist, who also reviewed the use of standardized measuring cups and spoons. Counter scales for weighing foods were distributed to those participants who did not have them already. Four consecutive days of dietary intake were to be recorded, using the scales to establish weights of foods eaten, at the end of each intervention period (at the end of the low-fat experimental period and the end of each of the crossover periods), for a total of 12 days per subject.

During the entire study period, a trained nutritionist contacted subjects on randomly selected days by phone at home or work to conduct the recall interview. Although each subject's availability differed, an average of 1.2 recalls per subject per month were collected. Individual problems with complying with the low-fat diet were discussed at the time of the call.

The semiquantitative FFQ was pretested on 11 healthy subjects (unpublished results). One hundred and two food items which represented about 87% of total energy, 97% of total fat and 97% of the three classes of fatty acids (i.e. saturated (SFA), mono-unsaturated (MFA), and polyunsaturated (PFA)) in the American diet<sup>20</sup> were included in the FFQ. Food ingredients and additives such as cooking oils, margarines, nuts and other vegetable products such as tofu were included because they affect the intake of total fat and the ratio of PFA to SFA. The FFQ was administered by a trained nutritionist at the end of the study protocol to assess nutrient intake during the previous 10 months. Three-dimensional food models were used so that portion sizes could be visualized. Nutrient scores were calculated from the FFQ by multiplying a weight assigned to the frequency of use (e.g. 'once per day' = 1) by the nutrient composition of each food item.

The records, recalls and FFQ were used to determine intake of energy, per cent of energy from fat, total fat, SFA, PFA, MFA, cholesterol and dietary fibre. In addition, estimates of vitamins A and C were obtained



**Fig. 1** Study design: t = time; LF = low-fat diet; CO = coconut oil supplement; SO = safflower oil supplement; W = washout period.

from the records and recalls. Nutrient values were derived from the computerized diet analysis system (DIAN, American Health Foundation, New York), which obtained its nutrient data from the US Department of Agriculture's food composition tables<sup>21</sup> as well as from manufacturers' data. When reported foods were not found in the database, foods with similar nutrient content were substituted<sup>22</sup>.

Dietary data were analysed both as absolute nutrient intakes (unadjusted) and as intakes per 1000 kcal (adjusted). Summary statistics were computed, including skewness and kurtosis, to assess normality. Based on these analyses, both the energy unadjusted and the adjusted intakes of fibre, cholesterol and vitamins A and C were transformed by  $\log_e$  to achieve normality for statistical analyses. Person-specific mean energy and nutrient intakes from records and recalls were averaged to determine the overall mean intakes estimated from each method. Difference scores based on nutrient intakes estimated from each of the three methods were computed for each of the pairwise comparisons for all nutrients with the exception of vitamins A and C, for which only the record versus recall comparison was possible (because of limitations of the FFQ). The SD of this difference was computed. Pearson product-moment ( $r_p$ ) and Spearman rank ( $r_s$ ) correlations assessed linear and rank order agreement between nutrient intakes derived from the three methods.

Linear regression modelling was used to determine the predictors of variability for intakes of energy and nutrients estimated by the recalls and records, because it was from those methods that daily estimates of intake could be derived. The general linear model (GLM) was used to fit an analysis of variance (ANOVA) because it

can handle imbalances in the data resulting from missing values. Nutrient scores were fit as dependent variables in the GLM. These were fit both as raw daily intake and energy adjusted in order to remove variation due to energy. Independent variables included: 12 dummy variables to indicate the individuals from whom the dietary data were collected; a variable to indicate whether supplements of oil were being consumed at the time the data were collected; the method (record or recall) by which the data were obtained; and the sequence (i.e. order) in which the data were collected. The total  $n$  for these analyses was 300 and the degrees of freedom for the model was 15. Alternative models were fit that included a method-by-sequence interaction term. All statistical analyses were conducted using the PC version of the Statistical Analysis System<sup>23,24</sup>.

## Results

Of the original 26 sample members, 13 complied with all study procedures including satisfactory completion of all three of the dietary assessment measures and were used for these analyses. Nine of the original sample members were dropped because of extremely poor compliance with the low-fat diet, inability to be reached by telephone or a permanent move from New York City. Four additional subjects were dropped because of their unwillingness to supply complete dietary information and/or their refusal to take the dietary supplement.

An average of 11.4 records and 11.7 recalls were collected per person. Table 1 presents the mean energy and nutrient intakes as determined by each of the dietary assessment methods. Mean per cent of

**Table 1** Estimates\* of energy, per cent of energy as fat, and nutrients by three diet assessment methods

Nutrient	Record <sup>†</sup>	Recall <sup>‡</sup>	FFQ
Energy (kJ)	7766 ± 1822	8092 ± 1973	8222 ± 2261
Energy (kcal)	1858 ± 436	1936 ± 429	1967 ± 541
Fat energy (%)	24.0 ± 5.8	20.8 ± 3.8	33.2 ± 6.5
Fat (g)	50.6 ± 17.4	45.6 ± 15.9	72.8 ± 26.2
energy adjusted <sup>§</sup>	26.7 ± 6.4	23.1 ± 4.2	36.9 ± 7.3
Saturated fat (g)	13.3 ± 4.4	12.3 ± 3.6	24.7 ± 8.3
energy adjusted <sup>§</sup>	7.1 ± 2.1	6.4 ± 1.6	12.7 ± 3.0
Monounsaturated fat (g)	15.6 ± 5.1	4.3 ± 4.7	20.5 ± 8.5
energy adjusted <sup>§</sup>	8.2 ± 2.2	7.4 ± 1.7	10.3 ± 3.1
Polyunsaturated fat (g)	11.5 ± 5.4	9.6 ± 5.7	21.8 ± 11.6
energy adjusted <sup>§</sup>	6.1 ± 2.5	4.8 ± 1.8	10.8 ± 3.6
Cholesterol (mg)	125 ± 78	135 ± 97	151 ± 111
energy adjusted <sup>§</sup>	73 ± 48	78 ± 64	77 ± 50
Dietary Fibre (g)	21.8 ± 11.1	25.6 ± 8.7	28.6 ± 12.0
energy adjusted <sup>§</sup>	11.9 ± 3.8	14.1 ± 3.6	15.1 ± 5.2
Vitamin A (IU)	12,424 ± 6160	14,069 ± 10,438	N/A
energy adjusted <sup>§</sup>	7724 ± 4760	7695 ± 6115	N/A
Vitamin C (mg)	162 ± 69	176 ± 97	N/A
energy adjusted <sup>§</sup>	90 ± 37	91 ± 38	N/A

\* Tabulated values are daily intakes estimated by each of the three methods (means ± SD).

<sup>†</sup> Values are obtained from an average of 11.4 days of records.

<sup>‡</sup> Values are obtained from an average of 11.7 days of recalls.

<sup>§</sup> These nutrient values are expressed as the value per 1000 kcal day<sup>-1</sup>.

**Table 2.** Mean differences in nutrient intakes and correlations between diet assessment methods ( $n = 13$  research subjects)

Nutrient	Record vs. recall				Record vs. FFQ				Recall vs. FFQ			
	Mean difference*	SD*	$r_p^\dagger$	$r_s^\ddagger$	Mean difference <sup>§</sup>	SD <sup>§</sup>	$r_p^\dagger$	$r_s^\ddagger$	Mean difference <sup>  </sup>	SD <sup>  </sup>	$r_p^\dagger$	$r_s^\ddagger$
Energy (kJ)	-366	966	0.86	0.65	-456	2571	0.22	0.07	-130	2349	0.35	0.20
Energy (kcal)	-78	231	0.86	0.65	-109	615	0.22	0.07	-31	562	0.35	0.20
Fat (g)	5.0	9.2	0.85	0.84	-22.2	23.0	0.51	0.50	-27.2	23.1	0.49	0.53
energy-adjusted <sup>¶</sup>	3.6	4.6	0.70	0.85	-10.2	5.7	0.66	0.68	-13.8	5.6	0.64	0.71
Saturated fat (g)	1.0	3.0	0.74	0.92	-11.4	7.4	0.44	0.47	-12.4	7.1	0.55	0.47
energy adjusted <sup>¶</sup>	0.7	1.9	0.50	0.45	-5.6	3.0	0.37	0.36	-6.3	2.6	0.50	0.40
Monounsaturated fat (g)	1.3	3.7	0.72	0.86	-4.9	7.1	0.56	0.46	-6.2	7.2	0.54	0.58
energy adjusted <sup>¶</sup>	0.8	1.8	0.63	0.66	-2.1	2.5	0.61	0.75	-2.9	2.4	0.62	0.54
Polyunsaturated fat (g)	1.9	5.3	0.75	0.85	-10.3	5.1	0.57	0.55	-12.2	11.7	0.51	0.61
energy adjusted <sup>¶</sup>	1.3	2.2	0.53	0.66	-4.7	2.8	0.63	0.68	-6.0	3.3	0.43	0.46
Cholesterol (mg)	-10	40	0.79	0.75	-26	73	0.62	0.64	-16	56	0.81	0.81
energy adjusted <sup>¶</sup>	-5	32	0.80	0.80	-4	36	0.66	0.61	1	39	0.75	0.73
Dietary Fibre (g)	-3.8	8.1	0.66	0.70	-6.8	11.9	0.43	0.54	-3.0	12.6	0.36	0.47
energy adjusted <sup>¶</sup>	-2.2	4.5	0.35	0.31	-3.2	3.8	0.63	0.54	-1.0	5.4	0.51	0.45
Vitamin A (IU)	-1645	8978	0.78	0.80	-	-	-	-	-	-	-	-
energy adjusted <sup>¶</sup>	29	5644	0.68	0.58	-	-	-	-	-	-	-	-
Vitamin C (mg)	-14	104	0.53	0.58	-	-	-	-	-	-	-	-
energy adjusted <sup>¶</sup>	-1.0	39	0.50	0.53	-	-	-	-	-	-	-	-

\* Shown is the mean difference of record–recall and its SD.

† This is the Pearson product moment correlation.

‡ This is the Spearman, rank order correlation coefficient.

§ Shown is the mean difference of record–FFQ and its SD.

|| Shown is the mean difference of recall–FFQ and its SD.

¶ These nutrient values are expressed as the value per 1000 kcal day<sup>-1</sup>.

energy from fat and intakes of energy and nutrients were highest with the FFQ. Mean intakes of energy, cholesterol, dietary fibre and vitamin C were lowest as determined by the records. The mean unadjusted intake of vitamin A was lowest with the records while the mean adjusted intake of vitamin A was lowest when determined by the recalls.

Table 2 presents the mean difference, SD and correlation coefficients in energy and nutrient intakes between diet assessment methods. Consistent with data presented in Table 1, mean differences were smallest in comparing the records with the recalls. Differences between the mean intakes of nutrients as determined by the FFQ compared to either the records or recalls were generally much larger. For example, the FFQ consistently overestimated nutrient intake, even after adjusting for energy intake. For fat, the difference between either the records or the recalls and the FFQ is approximately

10–12% as a percentage of energy while the record–recall difference was only about a quarter as large.

The SDs of the mean differences between the records and the recalls were small compared to those obtained when comparing the FFQ to either the recall- or record-derived values. The SD of the record–FFQ and the recall–FFQ differences were not only much larger than the record–recall difference, but they were very similar to one another.

In general, Pearson and Spearman correlation coefficients, also shown in Table 2, were not appreciably different from each other. Consistently, correlations were highest for records versus recalls. Notably high correlations (all  $r > 0.65$ ) were obtained for energy intake, unadjusted intakes of fat, SFA, MFA, PFA and vitamin A. With the exception of the recall versus FFQ comparison of cholesterol intake (both energy adjusted and unadjusted) and the adjusted intakes of PFA and

**Table 3.** Per cent contribution to variance in nutrient intakes reported via multiple food records and 24-h diet recalls

	Energy	Fat	SFA	MFA	PFA	Fibre*	Cholesterol*	Vitamin C*	Vitamin A*
Per cent variability due to:									
subjects (inter-person)	37.7	30.8	19.9	18.6	22.7	23.6	30.9	19.2	17.6
treatment <sup>†</sup>	0	0	0	0	0	0	0	0.7	1.8
method <sup>‡</sup>	0	0	0	0	0	1.4	0	0	1.4
sequence <sup>§</sup>	0	0	0	0	0	0	0	0.5	0
error (intra-person)	62.3	69.0	80.1	81.4	77.3	75.0	69.1	79.6	82.4
Model $R^{2  }$	0.38	0.32	0.21	0.19	0.24	0.25	0.31	0.20	0.18

\* Transformed by  $\log_e$  to improve normality.

† Low-fat diet only, or supplement.

‡ Food record or 24-h diet recall.

§ The order of the assessment in a series of about 23 records and recalls per subject.

|| This is the proportion of variability explained by the variables in this model.



dietary fibre, correlations for records versus FFQ and recalls versus FFQ were considerably lower than those of records versus recalls. Most notable was the low correlation for energy intake evident in the records versus FFQ comparison.

Table 3 presents the results of the ANOVA, which examined sources of variability in the record- and recall-derived estimates of nutrient intakes. The dependent variables were the daily intake of energy and specific nutrients as estimated by either the records or the recalls. Independent variables included factors that could influence the estimated intakes of energy and nutrients, as described in the methods section. For energy and every nutrient examined, inter-person variability was significantly ( $P < 0.0001$ ) associated with estimates of intake. For fat, SFA, MFA, PFA, cholesterol and dietary fibre, there was no effect by treatment (ingestion of coconut or safflower oil supplement versus no supplement), method (record versus recall) or sequence (the placement of the specific record or recall in the sequence of assessments for each person) in the models in which an interaction term was fit. Only for vitamin A were there effects by method ( $b = 1.10, P = 0.03$ ) and treatment ( $b = 0.59, P = 0.01$ ). There was also a method-by-sequence interaction for vitamin A ( $b = -0.09, P = 0.03$ ), the only such effect observed in any of these data.

Table 4 presents the energy adjusted results based on the analogous analysis. As generally is true in such data, inter-person variability was smaller than in the non-energy adjusted data. This is due to the fact that total intake (as indicated by energy) is a major determinant of the consumption of specific nutrients. The only exceptions to this were cholesterol and vitamin C. However, it should be noted that in addition to inter-person variability, intakes of vitamin A (transformed by  $\log_e$ ) were influenced by treatment ( $b = -0.55, P = 0.02$ ) and method ( $b = 1.06, P = 0.04$ ). As for the non-energy adjusted data, there was an interaction of method-by-sequence ( $b = -0.09, P = 0.02$ ). Intake of vitamin C was affected by sequence ( $b = -2.0, P = 0.004$ ), in addition to inter-person variability.

## Discussion

In the parent study from which these data were obtained, diet was assessed in order to have period-specific estimates of intake to be used in regression analyses where the dependent variable was natural killer cell activity. Although this study was not designed primarily as a 'validation' of the dietary assessment methods, at the time we conducted the intervention trial we were not sure which method would have superior statistical properties. Therefore, we were careful to approximate the kind of design used for comparisons of dietary assessment methods.

We found that energy and nutrient results obtained from records and recalls were generally interchangeable. The multiple 24-h recalls may have performed slightly better in estimating dietary intake, as evidenced by smaller SDs around the means obtained by that method compared to those obtained by the records or, especially, the FFQ. With few exceptions, total variability was smallest for the recalls. As expected, the SD of the mean values of cholesterol and vitamin A were large, regardless of method. This is because consumption of cholesterol and vitamin A tends to vary considerably within individuals over time<sup>5,25,26</sup>. The small SD from recalls could be the result of food omissions, causing the method not to capture all of the true variability. However, this explanation is inconsistent with our results, which showed similar intakes of energy and nutrients across methods. Further, if there were confounders compensating for this effect, the resulting error variance would probably be larger.

Because nutrient intakes are determined, to a large extent, simply by the total amount of food consumed, we adjusted nutrients for total energy intake. Data were presented both as absolute and as energy adjusted intakes because the nutrient density of the diet may be more relevant than the actual gross intake. In general, adjusting for energy has an inconsistent effect on correlation coefficients. Usually, controlling for energy intake will cause the apparent range of nutrient intakes

**Table 4.** Per cent contribution to variance in energy adjusted nutrient intakes reported via multiple food records and 24-h diet recalls

	Fat	SFA	MFA	PFA	Fibre*	Cholesterol*	Vitamin C*	Vitamin A*
Per cent variability due to								
subjects (inter-person)	20.6	16.7	16.3	18.7	12.7	36.0	17.5	11.1
treatment <sup>†</sup>	0.2	0	0.4	0.9	0	0.1	2.7	0.2
method <sup>‡</sup>	1.0	0	0.8	1.5	1.0	0.1	0	0
sequence <sup>§</sup>	0.5	0	0	0	0.5	0	2.3	0
error (intra-person)	77.7	83.3	82.5	78.9	85.8	63.8	77.5	88.7
Model $R^2$ <sup>  </sup>	0.24	0.18	0.17	0.21	0.13	0.36	0.22	0.11

\* Transformed by  $\log_e$  to improve normality.

<sup>†</sup> Low-fat diet only, or supplement.

<sup>‡</sup> Food record or 24-h diet recall.

<sup>§</sup> The order of the assessment in a series of about 23 records and recalls per subject.

<sup>||</sup> This is the proportion of variability explained by the variables in this model.

to decrease, thereby decreasing the correlation coefficients<sup>27</sup>. In contrast, if subjects are biasing the data by under- or overreporting total intakes, nutrients may be better expressed as a proportion, and adjusting for total energy intake could cause the correlation coefficients to increase<sup>28</sup>.

In these data, the effect of adjusting for energy intake on the correlation coefficients was variable. For example, for the records versus recalls comparison, controlling for energy caused the correlation coefficients for all nutrients to decrease with the exception of the  $r_p$  and  $r_s$  for cholesterol and the  $r_s$  for total fat. However, for records or recalls versus FFQ, adjusting for energy produced mixed effects, indicating that the FFQ may have been subject to a bias relative to the other two methods. Block *et al.*<sup>29</sup> showed that adjusting for energy had little to no effect on correlation coefficients in their study. However, Willett *et al.*<sup>30</sup> found that adjusting for energy intake caused the correlation coefficients for macronutrients (i.e. fat, protein, carbohydrate) to decrease. In contrast, Willett *et al.*<sup>28</sup> also showed, in a previous landmark study, that correlation coefficients increased when nutrients were adjusted by energy.

Besides examining differences and level of agreement in energy and nutrient intakes as estimated by the three methods, we sought to determine what specifically accounted for the variation in daily intake as estimated by the records and the 24-h recalls. From the outset, we expected that inter-person factors, method (i.e. record or recall) and sequence (i.e. order of administration) could help in determining daily estimated intake. Because this was an intervention study, we also had to account for the timing of the assessment relative to the treatment. It is important to note that the assignment of the assessment method was not random. Although the exact number of recalls per person varied and the exact day of the 24-h recall interview was randomly selected, the pattern was a series of recalls followed by four records, then more recalls followed by four records, and finally more recalls followed by four records. Because of this pattern, the effect of sequence cannot be completely disentangled from that of method.

Using the GLM to fit the variables noted in Tables 3 and 4, we found that for total energy intake and all of the nutrients, inter-person variability was a significant predictor of intake ( $P < 0.0001$ ). These results confirm the findings of other researchers<sup>2,25</sup> and the intuitive notion that individuals differ from one another in their consumption of energy and specific nutrients. With the exception of vitamins A and C, there was neither an effect by treatment, nor by method or sequence. Similar results were found by Beaton *et al.*<sup>25</sup> who showed that sequence did not affect the estimated intake of nutrients when six 24-h recalls were obtained for each subject. The absence of an effect of sequence

indicates that there is no 'training effect' through which a quantitative difference in reported intake emerges over time. The absence of an effect of method indicates that the records and recalls were interchangeable for the men enrolled in this study.

In contrast to the macronutrients, estimates of vitamin A were affected by treatment and method. Because of the negative regression coefficient for treatment, intakes of vitamin A would be predicted to be lower when supplements of coconut or safflower oil were not taken (i.e. in the pre-treatment phase). Conversely, because of the positive regression coefficient for method, estimates of vitamin A derived from the recalls would be predicted to be high compared to those obtained from the records.

The sole nutrient which was affected by sequence was vitamin C. This could be indicative of seasonal variation in vitamin C-containing foods (no supplements were permitted), dietary compensation for treatment or a training effect. The effect of season is the most likely explanation. Because the study progressed from summer to late winter in New York City, it is likely that subjects decreased their intake of vitamin C from fresh fruits and vegetables over its course. Despite its statistical significance, the practical consequence is likely to be minimal because the magnitude of effect ( $b = -2.0$  mg per 1000 kcal) is quite small.

The  $R^2$  values for the ANOVA ranged from 0.13 for fibre and vitamin A (both energy adjusted) to 0.38 for energy, meaning that we could account for up to 38% of total variability by the variables included in the models. Although these  $R^2$  values indicate that a moderate amount of variability has been accounted for by the independent variables, they also indicate that there are other factors which can account for day-to-day variations in energy and nutrient intakes. Beaton *et al.*<sup>25</sup> showed that gender was a major contributor to the variability of the unadjusted intakes of nutrients. The results from this study omit that source of variability because the sample comprised of men. Two other sources of variability, age and educational level<sup>31</sup>, are largely factored out by design because the subjects' range of ages was small and they had similar levels of education.

The primary purpose of this study was to focus on the recall-records comparison, something for which it was uniquely positioned to do. Still, the study allowed for an interesting comparison between the FFQ-derived nutrient scores and those derived from both the recalls and records. It is interesting that, in contrast to the very high levels of agreement between the recalls and records, those obtained in comparing the FFQ with either of these other methods produced correlation coefficients that were similar to those obtained in other studies<sup>28,29</sup>. As noted by Delcourt *et al.*<sup>27</sup> the SD of the difference between diet assessment methods is useful as an indicator of agreement between methods.

Consistently, the SD of the mean differences between records and recalls were lowest among the three comparisons. In contrast, when the FFQ-derived nutrient values were compared to those from the records or recalls, the SD of the mean differences in intakes were usually considerably higher, indicating poor agreement between the FFQ and the two other methods.

The correlation coefficients for both the records and recalls versus FFQ were approximately the same ( $r \sim 0.54$ ) as those obtained by Block *et al.*<sup>29</sup> who also compared an FFQ to the mean of three 4-day food records. A mean correlation coefficient of about 0.54, however, shows that their FFQ accounted for less than 30% of the food record's variability. Similarly, Willett *et al.*<sup>28</sup> found that correlations between an FFQ and the mean of four 1-week food records yielded mean correlation coefficients ranging from 0.38 to 0.55, showing that between  $\sim 14$  and  $\sim 30\%$  of the variability of the records was accounted for by the FFQ.

In an intensive study where multiple recalls or records are feasible, it appears that either of these methods is superior to a single FFQ for assessing dietary intake. Given the tremendous amount of time and effort that must be expended conducting multiple days of recalls or records, and the fact that they are often considered as comparison methods, this should not be too surprising. Nonetheless, many larger scale studies and ones on very limited budgets will not consider recalls or records as practicable options for assessing habitual intake of study participants.

It is noteworthy that the sum of fatty acids and cholesterol for the records and recalls (Table 1) does not come within 10% of the figure for total fat. Specifically, the sum of fatty acids and cholesterol equal less than 80% of total fat for both of these methods while the sum of fatty acids and cholesterol equals almost 92% of total fat with the FFQ, even though the same nutrient data base was used across all of the methods. The better reconciliation between the fatty acid subsets and total fat from the FFQ is explained by the fact that it is a list-based method; only a limited number of foods are presented, and the foods that were averaged to obtain nutrient scores happened to have more complete data for the fatty acids. Although it was not mentioned in their results, Posner *et al.*<sup>32</sup> also showed that the sum of fatty acids and cholesterol were not within 10% of total fat. Using the Michigan State nutrient database, their results for fatty acids and cholesterol summed to 84% of total fat for women. For men, however, fatty acids and cholesterol summed to less than 80% of total fat with the recalls and to less than 70% of total fat with the FFQ.

The results from this and the other studies discussed above highlight an often under appreciated fact: dietary assessment research is handicapped not only by the limitations of the diet assessment methods but also by the flaws of incomplete data in the nutrient databases.

In this instance, the analytical techniques for quantifying total fat differs from the methodology used to quantify fatty acids; additionally, nutrient databases often have missing data for the fatty acids<sup>33</sup>. Although analytical techniques and nutrient tables have improved greatly over the past decade, even the best of the databases are impeded by missing data.

Although FFQ 'have become the primary method for measuring dietary intake in epidemiologic studies'<sup>2</sup>, the method is certainly not without limitations. By design, FFQ focus upon specific nutrients or food components. That is, they are precoded with a limited number of foods; by choosing the frequency of consumption and a serving size, rough estimates of dietary intake can be obtained. Because of their low expense and the low level of participant commitment required, at least relative to multiple records or recalls, FFQ have become increasingly popular in nutrition research. Unfortunately, many researchers fail to observe the inherent measurement error associated with FFQ and report their results as if the data derived were without error. Furthermore, some investigators have gone a step beyond by using FFQ to estimate nutrient intake under circumstances for which they were not designed or validated, for example to detect change due to a dietary intervention<sup>13,34-36</sup>. Misunderstanding the shortcomings and careless misuse of diet assessment methods can result in the publication of conflicting results, thus clouding real diet-disease relationships and confusing the public.

This study purposely had high internal validity at the expense of external validity because the intent was to test the effect of dietary fat on natural killer cell activity, a hypothesis that had never been tested previously in humans. Consequently, the sample comprised 13 fairly homogeneous men who were willing to reduce their consumption of dietary fat and provide intake data over a period of nearly 1 year. Due to a rigorous assessment schedule in a very compliant but small group of men, we were able to compare the assessment tools in an 'ideal' situation with a minimal number of confounders; however, we acknowledge that this is just one step in understanding the properties of these methods.

There is tremendous, though often misplaced, interest in external validity. External validity really is a second-order issue; there is not much point in generalizing an invalid result. This study, though small, has good internal validity. We intend to make no claim about extrapolating results to a group very different in terms of demographic characteristics or level of compliance. This study is unique in that there are a large number of dietary measurements made using both the recalls and records. Of course, to obtain such results requires a much higher level of compliance than what may be attainable in larger-scale epidemiological studies. Still, in studies where one would have to choose between using multiple recalls or records, a



relatively high level of compliance would be assumed. So, the comparability of the recalls and records could reasonably be extrapolated to such a circumstance. Clearly, future work will need to include women, different age groups and individuals with diverse educational backgrounds because these factors are known or thought to affect self-reports<sup>37,38</sup>.

## References

- Buzzard IM, Faucett CL, Jeffery RW, *et al.* Monitoring dietary change in a low-fat diet intervention study: advantages of using 24-hour dietary recalls vs food records. *J. Am. Diet. Assoc.* 1996; **96**: 574–9.
- Willett WC. *Nutritional Epidemiology*. Oxford: Oxford University Press, 1990.
- Bazzarre TL, Yuhas JA. Comparative evaluation of methods of collecting food intake data for cancer epidemiology studies. *Nutr. Cancer* 1983; **5**: 201–14.
- 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.
- Livingstone MBE. Assessment of food intakes: are we measuring what people eat? *Br. J. Biomed. Sci.* 1994; **52**: 58–67.
- Hebert JR, Miller DR. Methodologic considerations for investigating the diet–cancer link. *Am. J. Clin. Nutr.* 1988; **47**: 1068–77.
- Gibson RS. *Principles of Nutritional Assessment*. New York: Oxford University Press, 1990.
- Potosky A, Block G, Hartman A. The apparent validity of diet questionnaires is influenced by number of diet-record days used for comparison. *J. Am. Diet. Assoc.* 1990; **90**: 810–13.
- Smith AF. Cognitive psychological issues of relevance to the validity of dietary reports. *Eur. J. Clin. Nutr.* 1993; **47**, suppl 2: S6–S18.
- Last JM, ed. *A Dictionary of Epidemiology*, 2nd edn. New York: Oxford University Press, 1988.
- Hebert JR, Ockene IS, Hurley TG, Luippold R, Well AD, Harmatz MG. Development and testing of a seven-day dietary recall. *J. Clin. Epidemiol.* 1997; **50**: 925–37.
- Rothenberg E. Validation of the food frequency questionnaire with the 4-day record method and analysis of 24-h urinary nitrogen. *Eur. J. Clin. Nutr.* 1994; **48**: 725–35.
- Hebert JR, Clemow L, Pbert L, Ockene IS, Ockene JK. Social desirability and approval biases in dietary self-report may profoundly compromise the validity of diet–disease studies. *Int. J. Epidemiol.* 1995; **24**: 389–98.
- Black AE, Goldberg GR, Jebb SA, Livingstone MBE, Cole TJ, Prentice AM. Critical evaluation of energy intake data using fundamental principles of energy physiology. 2. evaluating the results of published surveys. *Eur. J. Clin. Nutr.* 1991; **45**: 583–99.
- McConaghy J. Adults' belief about the determinants of successful dietary change. *Comm. Health Studies* 1989; **13**: 492–502.
- Naslund GK, Fredrikson M, Hellenius ML, de Faire U. Determinants of compliance in men enrolled in a diet and exercise intervention trial: a randomized, controlled study. *Patient Edu. & Counsel.* 1996; **29**: 247–56.
- Barone J, Hebert JR, Reddy MM. Dietary fat and natural killer cell activity. *Am. J. Clin. Nutr.* 1989; **50**: 861–7.
- Hebert JR, Barone J, Reddy MM, Backlund JYC. Natural killer cell activity in a longitudinal dietary fat intervention trial. *Clin. Immunol. Immunopathol.* 1990; **54**: 103–16.
- Hebert JR, Miller DR, Barone J, Richie JJ, Reddy M. Erythrocyte membrane fatty acids and natural killer cell activity in a longitudinal intervention trial. *Immunol. Infect. Dis.* 1991; **1**: 341–8.
- Block G, Dresser CM, Hartman AM, Carroll MD. Nutrient sources in the American diet: quantitative data from the NHANES II survey: II. macronutrients and fats. *Am. J. Epidemiol.* 1985; **122**: 27–39.
- US Dept of Agriculture. *USDA Nutrient Data Base for Standard Reference Release 5*. Hyattsville, MD: USDA, 1985.
- Engle A, Hebert JR, Reddy BS. Relationships between food consumption and dietary intake among healthy volunteers and implications for meeting dietary goals. *J. Am. Diet. Assoc.* 1990; **90**: 526–33.
- SAS. *SAS User's Guide*. Cary, NC: SAS Institute, 1997.
- SAS. *SAS User's Guide - Statistics Version 6.11*. Cary, NC: SAS Institute, 1997.
- Beaton GH, Milner J, Corey P, *et al.* Source of variance in 24-hour dietary recall data: implications for nutrition study design and interpretation. *Am. J. Clin. Nutr.* 1979; **32**: 2546–59.
- Beaton GH, Milner J, McGuire V, Feather TE, Little JA. Source of variance in 24-hour dietary recall data: implications for nutrition study design and interpretation. Carbohydrate sources, vitamins, and minerals. *Am. J. Clin. Nutr.* 1983; **37**: 986–95.
- Delcourt C, Cubeau J, Balkau B, Papoz L. Limitations of the correlation coefficient in the validation of diet assessment methods. *Epidemiology* 1994; **5**: 518–24.
- Willett WC, Sampson L, Stampfer MJ, *et al.* Reproducibility and validity of a semiquantitative food frequency questionnaire. *Am. J. Epidemiol.* 1985; **122**: 51–65.
- Block G, Woods M, Potosky A, Clifford C. Validation of a self-administered diet history questionnaire using multiple diet records. *J. Clin. Epidemiol.* 1990; **43**: 1327–35.
- Willett WC, Reynolds RD, Cottrell-Hoehner S, Sampson L, Browne ML. Validation of a semi-quantitative food frequency questionnaire: comparison with a 1-year record. *J. Am. Diet. Assoc.* 1987; **87**: 43–7.
- Morgan KJ, Johnson SR, Rizek RL, Reese R, Stampley GL. Collection of food intake data: an evaluation of methods. *J. Am. Diet. Assoc.* 1987; **87**: 888–96.
- Posner BM, Martin-Munley SS, Smigelski C, *et al.* Comparison of techniques for estimating nutrient intake: the Framingham Study. *Epidemiology* 1992; **3**: 171–7.
- Guilland JC, Aubert R, Lhuissier M, *et al.* Computerized analysis of food records: role of coding and food composition database. *Eur. J. Clin. Nutr.* 1993; **47**: 445–53.
- Harlow BL, Cramer DW, Geller J, Willett WC, Bell DA, Welch WR. The influence of lactose consumption on the association of oral contraceptive use and ovarian cancer risk. *Am. J. Epidemiol.* 1991; **134**: 445–53.
- Hebert JR, Stoddard AM, Harris DR, *et al.* Measuring the effect of a worksite-based nutrition intervention on food consumption. *Ann. Epidemiol.* 1993; **3**: 629–35.
- Kristal AR, Shattuck AL, Williams A. *Food Frequency Questionnaires for Diet Intervention Research. 17th National Nutrient Databank Conference*, Baltimore, MD, 7–9 June, 1992. Washington, DC: International Life Sciences Institute, 1994: 110–25.
- Suitor CJW, Gardner J, Willett WC. A comparison of food frequency and diet recall methods in studies of nutrient intake of low-income pregnant women. *J. Am. Diet. Assoc.* 1989; **89**: 1786–94.
- Hebert JR, Ma Y, Clemow L, *et al.* Gender differences in social desirability and social approval bias in dietary self-report. *Am. J. Epidemiol.* 1997; **146**: 1046–55.