Comparison of dietary assessment methods in nutritional epidemiology: weighed records *v.* 24 h recalls, food-frequency questionnaires and estimated-diet records

BY S. A. BINGHAM[1], C. GILL[2], A. WELCH[1], K. DAY[1], A. CASSIDY[1], K. T. KHAW[3], M. J. SNEYD[3], T. J. A. KEY[4], L. ROE[5] AND N. E. DAY[2]

[1] *MRC Dunn Clinical Nutrition Centre*, 100 Tennis Court Road, Cambridge CB2 1QL*

[2] *MRC Biostatistics Unit, Institute of Public Health, University of Cambridge, Forvie Site, Robinson Way, Cambridge CB2 2SR*

[3] *Clinical Gerontology Unit, F & G Block, Addenbrooke's Hospital, Cambridge CB2 2QQ*

[4] *Imperial Cancer Research Fund, Cancer Epidemiology Unit, Gibson Building, The Radcliffe Infirmary, Oxford OX2 6HE*

[5] *Imperial Cancer Research Fund, General Practice Research Group, Gibson Building, The Radcliffe Infirmary, Oxford OX2 6HE*

Women (*n* 160) aged 50 to 65 years were asked to weigh their food for 4 d on four occasions over the period of 1 year, using the PETRA (Portable Electronic Tape Recorded Automatic) scales. Throughout the year, they were asked to complete seven other dietary assessment methods: a simple 24 h recall, a structured 24 h recall with portion size assessments using photographs, two food-frequency questionnaires, a 7 d estimated record or open-ended food diary, a structured food-frequency (menu) record, and a structured food-frequency (menu) record with portion sizes assessed using photographs. Comparisons between the average of the 16 d weighed records and the first presentation of each method indicated that food-frequency questionnaires were not appreciably better at placing individuals in the distribution of habitual diet than 24 h recalls, due partly to inaccuracies in the estimation of frequency of food consumption. With a 7 d estimated record or open-ended food diary, however, individual values of nutrients were most closely associated with those obtained from 16 d weighed records, and there were no significant differences in average food or nutrient intakes.

Recognition of the need for prospective investigation of future individual risk of cancer, and other diseases of middle life, in relation to present day diet is one of the factors that has prompted the search for accurate measures of habitual diet. However, the methodological aims of nutritional epidemiology are inevitably tempered by two opposing interests: that of accuracy in the dietary assessment, and that of feasibility for use in large numbers of individuals, in order to have sufficient power to detect modest relative risks and to demonstrate dose–response relationships (Bingham, 1987). As part of the preparation for a large European collaborative study of diet and cancer risk (EPIC, European Prospective Investigation of Cancer) we have therefore assessed the accuracy of various methods commonly used in epidemiological studies of diet in large numbers of subjects. We assessed the accuracy of each method by comparison with 16 d of weighed records kept over 1 year.

* Now at Hills Rd, Cambridge CB2 2DH.

Table 1. *Number of items and their ordering on food-frequency questionnaires developed in Oxford and Cambridge*

	Oxford			Cambridge	
Category*	No. of items	% of total	Category*	No. of items	% of total
Bread and cereals	11	9	Vegetables	35	27
Potatoes etc.	10	8	Fruit	9	7
Dairy products and fats	22	17	Potatoes etc.	10	8
Sweets and snacks	12	9	Bread and cereals	11	8
Soups and sauces	8	6	Meat and fish	21	16
Meat and fish	16	13	Dairy products and fats	23	18
Drinks	14	11	Snacks and drinks	21	16
Fruit	11	9			
Vegetables	23	18			
Total	127	100		130	100

* Listed by order of appearance on the questionnaires. Frequency categories in both questionnaires were as follows: never or less than once per month; 1–3 per month; 1 per week, 2–4 per week; 5–6 per week; 1 per day, 2–3 per day; 4–5 per day; and six or more times per day.

METHODS

Volunteers

All women aged 50–65 years from the lists of two general practices in Cambridge were contacted by post in October 1988 and October 1989, and asked to fill in a general health, food and lifestyle questionnaire. All those who answered 'Yes' to the question 'Would you be interested in more detailed studies of diet?' were contacted by telephone and the purpose of the study was explained. Those who remained interested were then visited at home for further discussion, and those who still wished to participate were entered into the study. No exclusions were made on the grounds of ill health and the protocol was approved by the Dunn Nutrition Unit Ethics Committee in October 1988.

Protocol

The baseline measures for this study were 4 d of weighed-diet records and two 24 h urine collections obtained on each of four occasions (seasons) over the course of one year. Season 1 was October–January; 2, February–March; 3, April–June; 4, July–September. All individuals were therefore to complete a total of 16 d weighed records and eight 24 h urine collections. The urine collections were used to assess the extent of under-reporting and results are published elsewhere (Bingham *et al.* 1991, 1994).

The weighed records were obtained using PETRA (Portable Electronic Tape Recorded Automatic) scales (Cherlyn Electronics, Cambridge). These are accurate to ±1 g and automatically record verbal descriptions and weights of food on a dual track cassette, thus avoiding the necessity for subjects to keep written records. The PETRA scales do not disclose the weights of foods eaten to the subject (Bingham, 1987).

Subjects were visited in their homes the day before they were due to begin to weigh their food (day 0) when they were given a demonstration of the PETRA scales and asked to try them out themselves. The following day (day 1) they were revisited and the verbal descriptions recorded on the tapes were checked for completeness using a personal cassette player. Subjects were left with written instructions and with a notebook for recording recipes and food eaten out of the home which had not been recorded on the PETRA scales. The scales, notebook and completed tapes were collected from the subjects 1 or 2 d after

PLEASE PUT A TICK (✓) ON EVERY LINE

FOODS AND AMOUNTS	AVERAGE USE LAST YEAR								
SWEETS AND SNACKS (medium serving)	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day
Sweet biscuits, chocolate, eg. digestive (one)									
Sweet biscuits, plain, eg. Nice, ginger (one)									
Cakes eg. fruit, sponge, home baked									
Cakes eg. fruit, sponge, ready made									
Buns, pastries eg. scones, flapjacks, home baked									
Buns, pastries eg. croissants, doughnuts, ready made									
Fruit pies, tarts, crumbles, home baked									
Fruit pies, tarts, crumbles, ready made									
Sponge puddings, home baked									
Sponge puddings, ready made									
Milk puddings, eg. rice, custard, trifle									
Ice cream, choc ices									
Chocolates, single or squares									
Chocolate snack bars eg. Mars, Crunchie									
Sweets, toffees, mints									
Sugar added to tea, coffee, cereal (teaspoon)									
Crisps or other packet snacks, eg. Wotsits									
Peanuts or other nuts									
SOUPS, SAUCES, AND SPREADS									
Vegetable soups (bowl)									
Meat soups (bowl)									
Sauces, eg. white sauce, cheese sauce, gravy (tablespoon)									
Tomato ketchup (tablespoon)									
Pickles, chutney (tablespoon)									
Marmite, Bovril (teaspoon)									
Jam, marmalade, honey (teaspoon)									
Peanut butter (teaspoon)									
	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day

Please check that you have a tick (✓) on EVERY line

PLEASE PUT A TICK (✓) ON EVERY LINE

FOODS AND AMOUNTS	AVERAGE USE LAST YEAR								
DRINKS	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day
Tea (cup)									
Coffee, instant or ground (cup)									
Coffee, decaffeinated (cup)									
Coffee whitener, eg. Coffee-mate (teaspoon)									
Cocoa, hot chocolate (cup)									
Horlicks, Ovaltine (cup)									
Wine (glass)									
Beer, lager or cider (half pint)									
Port, sherry, vermouth, liqueurs (glass)									
Spirits, eg. gin, brandy, whisky, vodka (single)									
Low calorie or diet fizzy soft drinks (glass)									
Fizzy soft drinks, eg. Coca cola, lemonade (glass)									
Pure fruit juice (100%) eg. orange, apple juice (glass)									
Fruit squash or cordial (glass)									
FRUIT (1 fruit or medium serving) For very seasonal fruits such as strawberries, please estimate your average use when the fruit is in season									
Apples									
Pears									
Oranges, satsumas, mandarins									
Grapefruit									
Bananas									
Grapes									
Melon									
Peaches, plums, apricots									
Strawberries, raspberries, kiwi fruit									
Tinned fruit									
Dried fruit, eg. raisins, prunes									
	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day

Please check that you have a tick (✓) on EVERY line

Fig. 1. Sample pages of the final version of the Oxford food-frequency questionnaire, to be used in the European Prospective Investigation of Cancer (EPIC). Information concerning copies of and related software can be obtained from Ms S. Oakes, EPIC, Institute of Public Health, Cambridge.

completion of the 4 d weighing period, on days 5 or 6. Each 4 d period included different days chosen to ensure that all days of the week were studied, and that there was an appropriate ratio of weekend to weekdays included during the year. Subjects who submitted unsatisfactory records for all or part of the 4 d for technical and other reasons were asked to weigh their food for extra days the following season if possible.

The weights and verbal descriptions of food consumed were transposed by hand from the tape cassettes using the PETRA master console, and hand-coded for computer calculation of nutrient intake using food tables (Paul & Southgate, 1978; Wiles *et al.* 1980; Bingham *et al.* 1990). Subjects from practice 1 (group 1) were studied from October 1988–1989, and those from practice 2 (group 2) from October 1989–1990.

Other methods

A number of other dietary methods were investigated for comparison with the 16 d weighed record results, and comparisons of means obtained on the first presentation of each method are reported here. Most methods were repeated on two or more occasions and the results of error variance analysis of separate measures are to be reported elsewhere (S. Bingham, M. Plummer, D. Clayton, K. Khaw and N. Day, unpublished results).

Food-frequency questionnaires. Two food-frequency questionnaires (Oxford and Cambridge) were tested, each with a similar number of total food items (Table 1) and each asking subjects to recall 'average use last year'. The frequency categories of use in both were the same (Table 1). The Cambridge questionnaire devoted more attention to vegetable consumption than the Oxford questionnaire, with this category appearing first and last respectively, and 27 and 18 % of total items included respectively. The list of foods included in each category of the Cambridge questionnaire was chosen from foods most commonly eaten in a previous study of a Cambridge population (Bingham *et al.* 1981). The Oxford questionnaire was based on that used in the US Nurses Study (Willett *et al.* 1985). The frequency categories were not changed, but the lists of foods were modified by changing American food names to their British equivalent and by using the National Food Survey to identify additional foods which are important sources of nutrients in average British diets (National Food Survey Committee, 1982).

Nutrient intake was calculated from the questionnaires by multiplying the frequency of food consumption by portion weights to obtain weight of each food consumed (g/d); these were then converted to nutrient intake using an appropriate food-table code. In the Oxford questionnaire, 'medium serving' or units were specified (e.g. pints, slices, teaspoons). The weights of 'medium serving' portions were derived from experience with other dietary surveys and from published values (Crawley, 1988). In the Cambridge questionnaire no units or portion sizes were specified and average portion weights were obtained from the first season of weighed food records obtained from group 1. Hence in the Oxford questionnaire the portion size assigned to milk for example was 567 g (one pint) but in the Cambridge questionnaire, the portion size for milk was 59 g. Portion sizes of bread were approximately 30 g in the Oxford questionnaire and 50 g in the Cambridge questionnaire. Questionnaires were excluded from the Oxford data if ten or more lines had not been filled in, or because frequency choices of four or more per day designated items (for example, cabbage) had been made. In Cambridge, no exclusions were made and an average frequency choice for that group was assigned to missing lines.

The Cambridge questionnaire was sent to subjects by post as part of recruitment (see above) before they participated in the dietary studies, and the Oxford questionnaire was given to subjects to complete immediately before they started to weigh their food on day 0 in season 3. The final version of the Oxford questionnaire is shown in Fig. 1.

DAILY FOOD CHECKLIST

PLEASE FILL IN THE DATE AND PUT A NUMBER IN THE BOX FOR THE FOODS YOU HAVE EATEN TODAY.

Date __/__/__

VEGETABLES, FRESH, FROZEN, CANNED, DRIED

Item	Code
Avocado pears	692
Beans, green	563
Beans, baked	569
Beans, lentils, peas (dried)	604
Beansprouts	573
Beetroot	575
Broccoli/calabrese	577
Brussel sprouts	579
Cabbage or spring greens	586
Carrots	588
Cauliflower	592
Celery	594
Coleslaw	1105
Cucumber	597
Garlic	1218
Leeks	602
Lettuce	606
Marrow or courgettes	608
Mushrooms	609
Mustard & cress, watercress	611
Onions, cooking	613
Onions, spring	616
Parsnips	619
Peas	623
Peas, tinned	624
Peppers (red/green)	634
Radishes	654
Spinach	657
Swede	660
Sweetcorn	663
Tomatoes, fresh	666
Tomatoes, tinned	668
Other fresh herbs	617
Mixed vegetables	1119

SOUPS

Item	Code
Meat based	938
Vegetable based	955
Pulse (lentil) based	943

FRUIT

Item	Code
Apples	676
Pears	786
Oranges, satsumas	774
Grapefruit	740
Bananas	693
Strawberries, raspberries	817
Rhubarb	816
Dried fruit, e.g. raisins	809
Fruit salad, fresh	1123
Fruit salad, canned	731
Stewed fruit with sugar	681
Pineapple	792

POTATOES, RICE AND PASTRY

Item	Code
Boiled potato	640
Mashed potato	641
Chips	645
Roast potato	644
Potato salad	1111
Jacket potato	122
Yorkshire pudding	577
Stuffing	1162
Rice, white, boiled	020
Rice, brown, boiled	1008
White or green pasta or noodles, eg. spaghetti	027
Wholemeal pasta, eg. spaghetti, lasagne	1006
Savoury pies, eg. chicken, steak & kidney	425
Pizza	178
Quiche, savoury flan	179
Sweet pies, crumbles, steamed puddings etc.	106

BREAD AND BUTTER

Item	Units	Code
Rolls	Numbers	042
Cream crackers	Numbers	059
Crispbread, Ryvita	Numbers	060
White bread	Slices	033
Brown, granary bread	Slices	031
Wholemeal bread	Slices	030
Rye bread	Slices	1016
Hovis bread	Slices	032
Low fat spread, eg. Gold		186
Butter		140
Soft margarine		190
Hard margarine		188
Polyunsaturated margarine eg. Flora		192
Butter/margarine on other foods, eg. vegetables	Teaspoons	

EGGS AND CHEESE

Item	Units	Code
Cheese, Cheddar etc.		152
Cottage cheese		157
Eggs, boiled, fried, scrambled etc.	Numbers	169

SNACKS, CAKES, DRINKS, SUGAR

Item	Units	Code
Sweet biscuits, chocolate	Numbers	058
digestive	Numbers	062
Cakes eg. fruit	Numbers	069
sponge		077
pastry type		081
scones, currant buns	Numbers	087
other		095
Ice-cream		108
Nuts and Peanut butter		837
Crisps	Bags	652

Item	Units	Code
Marmite, Bovril		961
Chocolates	Numbers	859
Chocolate bars	Numbers	857
Sweets, toffees etc.	Numbers	865
Sugar in drinks or on cereal	Teaspoons	843
Jam, marmalade, honey, syrup		849
Jelly		110
Tea	Cups	877
Coffee, instant	Cups	642
ground	Cups	1143
decaffeinated	Cups	871
Other drinks eg.		1384
Horlicks	Cups	874
Spirits eg. gin	Measures	919
Wine, red, white, rose	Glasses	904
Beer or cider	Pints	896
Port, sherry, vermouth	Glasses	912
Fizzy soft drinks eg. Coca Cola, lemonade	Glasses	881
Real fruit juice (100%) eg. orange, apple	Glasses	885
Fruit drinks, eg. squash	Glasses	1151

MILK

Item	Units	Code
Milk, silver top or red (homogenized) top	Pints	125
Milk, semi-skimmed (red and silver top)	Pints	1185
Milk, skimmed (blue)	Pints	131
Milk, Jersey (gold)	Pints	128
Evaporated milk	Cups	134
Dried milk	Teaspoons	136
Soya milk	Pints	12042

SAUCES

Item	Code
Gravy	1161
Sweet pickle, eg. Branston, Chutney	929
Ketchup, brown sauce	931
Savoury white sauce eg. cheese, parsley	934

MEAT

Do you usually eat the fat on meat? Yes ___ No ___

Item	Code
Roast beef or steak	262/263
Roast lamb or chops	285/286
Roast pork or chops	309/310
Roast chicken or chicken joints	322/321
Stewed mince or beef stew or casserole	1077

Item	Code
Chicken casserole	1094
Other stewed or roast meat	1397
Soya meat, eg. Vegeburger	025
'Quorn' eg. Sainsbury's Savoury Pie	231/230
Bacon	394
Ham	393
Corned beef or spam etc.	413
Sausages	416
Beefburgers	376
Liver	404
Liver sausage, pate	

FISH

Item	Code
White fish, fried	442
White fish, not fried	444
Fatty fish, eg. herrings, salmon, tuna	508
Shellfish, eg. crab, prawns, mussels, scampi	523

FATS, CREAM, YOGHURT, MILK PUDDING

Item	Units	Code
Olive oil	Teaspoons	199
Other vegetable oil	Teaspoons	135
Salad cream, mayonnaise	Teaspoons	926
Cream, single, or sour		143
Cream, double, or clotted		146
Yoghurt, low fat, natural or 'diet'	Cartons	161
Yoghurt, fruit	Cartons	163
Custard, milk puddings eg. rice, semolina		102

BREAKFAST CEREALS

Item	Code
Cornflakes, Special K	048
Rice Krispies, Sugar Pops	053
All Bran and bran	047
Shredded Wheat, Weetabix	057
Wheat Flakes, Bran Flakes	1019
Puffed Wheat, Sugar Puffs	051
Country Store, muesli	050
Porridge, Ready Brek	018
Other breakfast cereals	

Other foods eaten but not on the list:

Fig. 2. Sample pages of the final version of the structured food-frequency (menu) record. Information concerning copies can be obtained from Dr S. Bingham.

24 h recalls. Two types of 24 h recall were tested. The first was unstructured and consisted of a blank sheet of paper and written example attached to the back of the Cambridge food-frequency questionnaire, which was given to the subjects during recruitment before the start of season 1 (see above). Published portion weights were used to calculate nutrient intakes (Crawley, 1988).

The second was a structured 24 h recall given to subjects to fill in on day 0 of season 2 and collected on day 1. This 24 h recall contained ten pages and was subdivided into seven time periods (e.g. before breakfast, breakfast, between breakfast and lunch) with detailed enquiries as to the amount and type of food eaten. Subjects were asked to give estimates of portion size either by amount or by referring to an accompanying nineteen-page booklet of thirty sets of black and white photographs and two-dimensional shapes of food. Three to four portion choices of each food were shown in each set and most were from those prepared by the Karolinska Institute, Sweden for a similar validation study (Callmer *et al.* 1993).

Data for both types of 24 h recall were hand-coded for computer calculation of nutrient intakes, using the same food tables as were used for the weighed records.

Estimated-diet records. Three 7 d estimated-diet records were tested. All were left with the subjects on day 5 or 6 of the protocol to fill in within the next 14 d, with a stamped addressed envelope for postal return.

In season 1, subjects received a 7 d structured food-frequency (menu) record (Bingham, 1987). To avoid confusion, this method is called the 'checklist'. This booklet comprised one page of instructions, one of an example, and seven pages (one for each day) of the checklist. The checklist was a printed list of foods on which subjects were asked to check off which foods they had eaten, counting half for a small portion, and two for a large portion at the end of each day. A space was left for subjects to record foods not present on the printed list, but otherwise the list was precoded for nutrient analysis. The list of foods was largely that of the Cambridge food-frequency questionnaire, with some additions, e.g. soups. Where possible, 'units' (e.g. slices, cups) were specified, or the average portion weights obtained from the chosen food code from the first set of 4 d weighed records (season 1) in group 1 were used to calculate nutrient intake from food tables as above. Season 1 group 1 data were also used to check that the appropriate food code for each item had been chosen. The final list of 160 foods arranged in fourteen sections is shown in Fig. 2.

In season 4 an attempt was made to gain more accurate information by asking subjects to give individual estimates of portion size using photographs. To simplify the method a condensed set of the fourteen most used sets of photographs and shapes were selected from the thirty sets used to assess portion size in the structured 24 h recall and given to the subjects as a seven page booklet. The number of the appropriate picture set was also pre-printed onto the form, and a space was left for subjects to fill in the appropriate portion size.

In group 2, season 3 only, an open-ended (unstructured) estimated-diet record or 7 d food diary was used. This had been developed for the MRC National Survey of Health and Development (Braddon *et al.* 1988; F. Key, A. Paul, A. Harter, G. Price, T. Cole and M. Wadsworth, unpublished results). Subjects kept a written record of food consumed at the time of eating and the booklet contained fifteen sets of black and white photographs, taken mostly from the study of Edington *et al.* (1989). Each set of photographs had three portion choices. Subjects were allowed to state portion sizes in other measures if they so wished. No part of this diary was precoded but a computer program, DIDO (Data In, Diet Out) was developed specifically for the coding of these booklets for nutrient analysis (F. Key, A. Paul, A. Harter, G. Price, T. Cole and M. Wadsworth, unpublished results). For this diary only, the nutrient composition database was different from that used in the weighed records

and all other methods. It included values from the food table supplements for cereals and for milk and milk products (Holland *et al.* 1988, 1989).

Statistics

Means and standard deviations of untransformed data are presented unless otherwise stated, using the averages of all available days of weighed records and 7 d of food diaries and checklists for each individual. The significance of difference was assessed using paired *t* tests, except where otherwise stated, and coefficients of variation are calculated as the standard deviation divided by the average of the 16 d weighed records. Using the Statistical Analysis Systems program (version 6.07; SAS Institute Inc., Cary, NC, USA), distributions were ranked and then divided into four equal parts and results were tabulated for the weighed records and other methods, in order to see if individuals were ranked similarly by different methods. Only those subjects who had completed the weighed record exactly, that is who submitted four complete dietary records at each of the four seasons, were used for this and correlation analysis.

Results are given for energy, macronutrients, non-starch polysaccharides (NSP), K, Ca, Fe, carotene, retinol, and vitamin C, all of which had less than 1 % 'missing values' in the database. Results are also shown for 'dietary fibre', although the database contained 9 % missing values.

RESULTS

Participants

Eighty-six subjects in group 1 and ninety-six in group 2 returned completed questionnaires and undertook to take part in the study in the first season. Of these, seventy-nine in group 1 and eighty-one in group 2 completed most or all of the protocol over the year, about 15 % of the original sample. Table 2 shows that there were no significant differences between the two groups in average age, weight, height or social class (assigning 1 to social class I and 5 to social class V). There was also no significant difference in mean energy, macronutrient, NSP or micronutrient intakes calculated from the weighed records between the two groups. Except where otherwise stated, the data were therefore combined for the two groups, giving a total of 160 subjects who completed the study.

Table 2 also shows some anthropometric and dietary data from a representative sample of 50–64-year-old British women (Gregory *et al.* 1990). There was no significant difference in body weight between the group of 160 women studied here and the representative population sample of 283 women, but the Cambridge women were significantly taller, and reported significantly greater intakes of energy, macronutrients, vitamin C, Fe and Ca.

Seasonal differences in consumption reported from the PETRA weighed records, and correlations and variances in nutritional intake

One hundred and sixty subjects completed seasons 1–3, but four subjects had moved out of the area by season 4, so that 156 completed all four seasons. The results of the four subjects who completed only three seasons of data were included in these results.

Table 3 shows average nutrient consumption by season. There was a trend towards lower consumption in seasons 3 and 4, and there were significant ($P < 0.01$) differences in carbohydrate (as starch), Ca, Fe and retinol between seasons 4 and 1. Examination of food consumption in each season (results not tabulated) showed that the only significant difference was in soups and sauces; consumption was 390 (SD 286) g in season 1 and 320 (SD 279) g in season 4, $P < 0.01$. The difference in Ca consumption was largely accounted for by a (non significant) 30 g reduction in milk consumption in season 4 compared with season 1, and a 4 g reduction in cheddar cheese consumption. There was no obvious change

Table 2. *Response rate, diet and anthropometric data from female volunteers aged 50–65 years from two general practices in Cambridge (groups 1 and 2), and some comparisons with data from a representative sample of British women (Gregory* et al. *1990)†*

(Mean values and standard deviations)

	Group 1		Group 2		Representative sample of women aged 50–64 years‡	
No. of initial questionnaires sent out	482		542		—	
Those answering 'Yes' to the volunteering question (%)	20		28		—	
Subjects who started the study (%)	18		18		—	
No. that completed the study	79 (16%)		81 (15%)		70%	
	Mean	SD	Mean	SD	Mean	SD
Age (years)	56	5	57	4	—	—
Weight (kg)	66	14	64	10	66	14
Height (m)	1·64	0·07	1·64	0·07	1·60**	0·06
Social class average	2·3	1·0	2·5	0·8	—	—
Energy intake (MJ/d)	8·0	1·5	7·7	1·5	6·7**	1·6
Protein (g/d)	69	12	68	13	63**	15
Fat (g/d)	78	18	75	19	71**	20
Carbohydrate (g/d)	228	51	218	52	182**	50
NSP (g/d)	15·2	5·0	16·4	5·2	—	—
Vitamin C(mg/d)	98	35	100	47	76**	60
Calcium (mg/d)	970	249	935	241	739**	217
Iron (mg/d)	12·8	4·0	12·9	3·7	10·6**	3·9

NSP, non-starch polysaccharides.
Mean value was significantly different from those for groups 1 and 2: ** $P < 0.01$.
† For details of procedures, see p. 620.
‡ Data from Gregory *et al.* (1990).

in food consumption to account for the differences in starch. Fe and retinol between seasons 1 and 4. The total weight of food and drink consumed was approximately 2630 g in seasons 1 and 3, and 2560 g in seasons 2 and 4.

Table 3 also shows estimates of within- and between-person variance in nutrient intake. Pooled within-person variation was in excess of 100% for alcohol, carotene and retinol, and ranged from 23% for energy to 56% for vitamin C for the other nutrients. The 16 d of weighed records therefore allowed means of energy, protein, carbohydrate and K to be within 6% standard error for each individual mean on average, and within 10% for all others except vitamin C (14%) and the highly variable nutrients. In general the between-person range was small and the within-person variation exceeded the between-person variation.

Correlations between nutrients

Table 4 is a correlation matrix between dietary nutrients from the 16 d weighed records. Highest correlations were between NSP and 'fibre' (r 0·98) and between fat and carbohydrates and total energy, r 0·86 and 0·82 respectively. Sugars and starch separately were less well correlated with energy (r 0·67 and 0·57), and fat as a percentage of energy was poorly correlated with total energy intake (r 0·18, data not shown). Protein and carbohydrate as percentages of energy were not correlated with total energy (r −0·45 and 0·00 respectively, data not shown). Alcohol was poorly correlated with total energy intake,

Table 3. *Intakes of nutrients by season from weighed records kept by 160 women aged 50–65 years, together with estimates of within- and between-person variation*†

	Season 1 October–January		Season 2 February–March		Season 3 April–June		Season 4 July–September		Within-person coefficient of variation	Between-person coefficient of variation	Within:between ratio	Mean standard error (%)
	Mean	SD	Mean	SD	Mean	SD	Mean	SD				
Energy (MJ/d)	7·93	1·73	7·98	1·87	7·86	1·78	7·58	1·57	23	19	1·21	6
Fat (g/d)	77	23	78	23	78	24	73	21	34	24	1·42	9
Protein (g/d)	69	16	70	15	68	15	66	13	25	19	1·32	6
Carbohydrates (g/d)	229	57	226	62	220	58	218**	54	23	23	1·00	6
Sugars (g/d)	110	33	108	41	110	40	107	38	32	32	1·00	8
Starch (g/d)	117	42	116	39	105	31	106**	32	38	27	1·41	10
NSP (g/d)	16	6	16	6	15	6	16	6	32	32	1·00	8
'Fibre' (g/d)	22	7	22	8	22	7	22	8	31	32	0·97	8
Potassium (g/d)	3·21	0·76	3·18	0·79	3·17	0·82	3·14	0·77	24	22	1·09	6
Calcium (mg/d)	997	319	964	273	941	290	913**	272	32	26	1·23	8
Iron (mg/d)	13·1	4·5	13·4	5·0	12·7	4·2	12·4	4·4	38	30	1·27	10
Carotene (mg/d)	3·63	3·06	3·91	2·80	3·08	2·19	3·38	2·49	113	54	2·09	28
Retinol (µg/d)	897	1436	849	1630	839	1350	627**	933	311	99	3·14	78
Vitamin C (mg/d)	92	46	99	50	112	53	94	48	56	41	1·37	14
Alcohol (g/d)	7·7	9·6	9·8	13·4	10·7	14·0	7·8	13·4	121	36	3·36	30

NSP, non-starch polysaccharides.

Mean value was significantly different from that for season 1: ** $P < 0.01$.

† For details of subjects and procedures, see p. 620.

Table 4. *Correlation matrix between nutrients recorded by 160 women aged 50–65 years completing 16 d weighed records over 1 year*†‡

	Energy	Fat	Protein	Carbohydrates	Sugars	Starch	NSP	'Fibre'	Potassium	Calcium	Iron	Carotene	Retinol	Vitamin C	Alcohol
Energy	1·0														
Fat	0·86	1·0													
Protein	0·68	0·48	1·0												
Carbohydrates	0·82	0·53	0·51	1·0											
Sugars	0·67	0·41	0·42	0·86	1·0										
Starch	0·57	0·38	0·35	0·66	0·28	1·0									
NSP	0·28	0·01	0·44	0·38	0·28	0·31	1·0								
'Fibre'	0·30	0·04	0·42	0·42	0·44	0·35	0·98	1·0							
Potassium	0·54	0·29	0·74	0·54	0·52	0·27	0·76	0·74	1·0						
Calcium	0·64	0·44	0·73	0·65	0·62	0·41	0·45	0·44	0·70	1·0					
Iron	0·42	0·21	0·51	0·38	0·26	0·36	0·63	0·61	0·59	0·17	1·0				
Carotene	-0·02	-0·12	0·24	0·02	0·14	-0·09	0·48	0·47	0·42	0·19	0·20	1·0			
Retinol	0·21	0·26	0·19	0·11	0·09	-0·07	0·03	0·01	0·15	0·17	0·11	-0·08	1·0		
Vitamin C	0·17	0·02	0·32	0·18	0·31	-0·01	0·60	0·58	0·68	0·29	0·32	0·45	0·09	1·0	
Alcohol	0·19	0·09	0·14	-0·22	-0·25	-0·11	0·00	-0·03	0·09	-0·13	0·15	-0·04	0·00	0·07	1·0

NSP, non-starch polysaccharides.
† For details of subjects and procedures, see p. 620.
‡ Coefficients > 0·159 are significant at the 5% level, and those > 0·208 significant at the 1% level.

and with all other nutrients. K, Ca and protein intakes were interrelated (r 0·70–0·74), as were NSP, K and vitamin C (r 0·60–0·76).

Differences between methods

Food-frequency questionnaires v. *PETRA-weighed records.* Thirty-three questionnaires were excluded from the Oxford data due to missing information. Table 5 shows that both questionnaires significantly overestimated means of intakes of NSP, fibre, K and carotene when compared with the average of 16 d weighed PETRA records. In addition, the Cambridge questionnaire significantly overestimated vitamin C and significantly under-estimated energy, fat and Ca. The Oxford questionnaire significantly overestimated for most other nutrients.

Table 6 shows that the differences in carotene, vitamin C and NSP were partly accounted for by the greater reported consumption of vegetables (approximately 120 g/d more) in both questionnaires than in the weighed records. Some of the portion weights used in the Oxford questionnaire were greater than those recorded on average in the weighed records (for example, carrots: 80 g questionnaire, 65 g weighed records) but others were the same or lower (for example, cabbage: 80 g questionnaire, 93 g weighed records). The portion weights used in the Cambridge questionnaire were very similar to those obtained on average over 16 d data from the weighed records, for example, carrots: 59 g questionnaire, 65 g weighed records. Hence the discrepancy in average total vegetable consumption was due to a greater reported frequency of consumption in the questionnaire methods than actually measured by the weighed records.

Consumption of another major contributor to NSP intake, wholemeal bread, was also overestimated by 16 and 14 g in the Oxford and Cambridge questionnaires respectively. Given the smaller portion sizes used in the Oxford questionnaire compared with the weighed records and the Cambridge version, this overestimation most likely arose from overestimation of frequency of consumption. However, the main proportional contributors to NSP were the same in all three methods, namely wholemeal bread (11–16%), apples (4–5%), and muesli or breakfast cereal containing bran (4–9%). Carrots contributed 51 to 52% carotene in all three methods, and tomatoes, broccoli, spinach and lettuce between 4 and 6% each. Much of the difference in vitamin C consumption found between the weighed records and the Cambridge questionnaire was due to overestimation of the frequency of consumption of orange juice and oranges, leading to a difference of 53 g in average reported daily consumption of orange juice, and a 30-g difference in consumption of oranges. Portion weights for these items were 143 g (weighed) and 139 g (questionnaire) for orange juice, and 116 g (weighed) and 118 g (questionnaire) for oranges. The two items made up 12 and 8% vitamin C in the weighed records and 23 and 17% in the questionnaires respectively.

The major contributors to Ca consumption in the PETRA-weighed records were milk (225 g/d, 29% of total Ca intake), and cheddar cheese (10 g/d, 8% of total Ca intake). However, daily milk consumption was overestimated by 150 g in the Oxford questionnaire and underestimated by 135 g in the Cambridge questionnaire (Table 6). As stated above, average portion sizes used or assigned were 567 and 59 g for the Oxford and Cambridge questionnaires respectively. Cheddar cheese was also overestimated by 15 g in the Oxford questionnaire, and the portion size of cheddar cheese used (60 g) was approximately double that recorded in the weighed records (31 g) and used in the Cambridge questionnaire (33 g). These differences in milk and cheese consumption largely accounted also for the significant differences in energy, fat and protein found between the Oxford questionnaire and weighed records, and the discrepancies in milk consumption contributed to the differences in K, Ca and sugars. Coffee, which supplied 4 to 7% of dietary K, was overestimated by 160 and

Table 5. *Daily intakes of nutrients obtained using seven different dietary assessment methods completed by 160 women aged 50–65 years‡*

(Mean values and standard deviations)

	16 d weighed All four seasons		Oxford food-frequency questionnaire Season 3		Cambridge food-frequency questionnaire Season 1		24 h recall (unstructured) Season 1		24 h recall (structured) Season 2		7 d checklist Season 1		7 d checklist and portions Season 4	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Energy (MJ/d)	7·83	1·50	8·81†	2·40	7·08†	2·1	7·57	2·2	8·64†	2·73	7·56	1·73	8·43***	2·03
Fat (g/d)	76	19	87†	31	58†	21	77	34	89†	40	71***	22	81	29
Protein (g/d)	69	12	82†	22	66	20	70	24	76***	30	67	13	76†	18
Carbohydrates (g/d)	223	52	245***	77	221	83	206**	68	239**	75	219	59	243†	64
Sugars (g/d)	109	35	135†	55	103	44	102	42	117**	51	108	39	114	43
Starch (g/d)	108	27	93†	33	112	53	98**	41	115	44	103	27	123†	33
NSP (g/d)	16	5	19†	7	19†	8	15	6	16	7	15**	4	19†	7
'Fibre' (g/d)	22	7	27†	10	27†	11	21	10	21	11	20†	6	25†	9
Potassium (g/d)	3·2	0·6	4·0†	2·2	3·4**	0·9	3·1	0·9	3·6	1·7	3·2	0·6	3·7†	8·7
Calcium (mg/d)	952	245	1308†	453	763†	282	879	360	1025	391	981	322	1033**	336
Iron (mg/d)	12·9	3·8	13·6	7·0	12·7	4·2	12·3	5·9	13·6	6·6	10·9†	2·4	13·4	4·5
Carotene (mg/d)	3·4	1·9	5·1†	3·2	4·7†	3·0	3·5	3·7	3·0	3·3	3·5	1·9	6·2†	4·7
Retinol (µg/d)	797	773	1010	952	1000	1048	1070	4178	702	1606	867	1430	609**	833
Vitamin C (mg/d)	99	41	122	53	134†	68	87***	56	110	65	99	45	110	64
Alcohol (g/d)	9	13	8	11	7	13	7***	14	6†	14	9	15	8	12

NSP, non-starch polysaccharides.

Mean values were significantly different from those obtained by 16 d weighed record: ** $P < 0.01$; *** $P < 0.001$; † $P < 0.0001$.

‡ For details of subjects and procedures, see pp. 620–628.

Table 6. *Daily food consumption values (g/d) obtained using seven different dietary assessment methods completed by 160 women aged 50–65 years‡*

(Mean values and standard deviations)

	16 d weighed All four seasons		Oxford food-frequency questionnaire Season 3		Cambridge food-frequency questionnaire Season 3		24 h recall (unstructured) Season 1		24 h recall (structured) Season 2		7 d checklist Season 1		7 d checklist and portions Season 4	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Total cereals	247	91	199†	91	265	166	216	131	256	141	266	110	277**	101
Bread	82	34	74	54	91	77	74	49	88	59	75	37	93	48
Breakfast cereals	21	21	21	20	26	33	21	29	18	23	18	18	28	33
Milk and milk products	296	148	491†	381	175†	131	269	173	307	196	363	210	335	208
Eggs	23	18	23	17	20	14	23	47	22	49	20	23	18	17
Fats	8	7	12**	14	15†	14	8	23	9	12	10	10	9	9
Meat	96	42	91	49	99	57	115	108	93	94	96	46	90	45
Fish	31	22	35	24	27	21	36	69	47**	83	25	23	36	32
Vegetables	272	85	406†	153	386†	161	273	143	294	157	246**	74	397†	164
Potatoes	88	42	99	44	105**	55	78	71	107	99	86	38	114†	52
Fruit	206	130	219	145	231	167	224	194	218	211	198	125	232	157
Nuts	3	5	4	8	3	7	3	11	4	17	3	6	10***	22
Sugars	26	22	36†	37	27	28	25	30	28	33	32	34	21	21
Beverages	940	342	1174	415	1108†	485	1090**	466	1282†	522	1127†	400	1163†	387
Alcoholic beverages	97	133	96	155	77	138	95	187	93	170	91	123	96	149
Sauces	342	219	63†	52	2	4	167	236	273**	259	40	42	58	48

Mean values were significantly different from those obtained by 16 d weighed record: ** $P < 0.01$; *** $P < 0.001$; † $P < 0.0001$.

‡ For details of subjects and procedures, see pp. 620–628.

50 g/d in the Oxford and Cambridge questionnaires respectively, and also contributed to the significant differences in reported K consumption.

24 h recalls v. *weighed records.* Table 5 shows that apart from vitamin C, starch, and alcohol there were no other differences between the weighed records and the unstructured 24 h recall in reported average intakes of nutrients. There were however significant overestimations in energy, fat, protein and sugars when the structured 24 h recall was compared with the weighed records. Significantly greater consumptions of beverages, mainly tea, coffee and water, in both types of 24 h recall, and of fish and potatoes in the structured 24 h recall were reported (Table 6).

Seasonal variations contributed to the significantly lower estimate of vitamin C consumption by the unstructured 24 h recall method, since there was no significant difference between vitamin C assessed by the 24 h recall and by the 4 d of weighed records collected at season 1, the season closest to the collection of the 24 h recall (Table 3). The significant difference in starch consumption was accounted for by a 10 g lower reported intake of bread in the 24 h recall and a 13 g lower reported intake of potatoes. In both methods, muesli contributed 4% to starch intake, potatoes 16% and bread 30%.

No clear reason for the significant overestimation of protein, fat and sugar in the structured 24 h recall was apparent, inspection of portion weight data revealing no obvious differences. Butter, cheese, polyunsaturated margarine and milk supplied 7, 4, 4, and 5% fat respectively to the total in both methods, and cheese, bread, milk, chicken and muesli 4, 5, 8, 2 and 1% protein respectively in both methods. Sugar, apples, marmalade, milk and orange juice supplied 7, 4, 3, 10 and 3% to the total sugars in the weighed records; apart from a 4% contribution from sugar, the pattern was otherwise similar in the structured 24 h recall.

Structured 7 d checklist (food-frequency menu record). Table 5 shows that the checklist underestimated mean Fe intake by 16% compared with the result obtained from the PETRA-weighed records. This was largely due to aggregation of the codes for breakfast cereals; of all possible Fe-enriched cereals, only those for muesli and All-Bran were used, and these supplied 0·7 mg/d according to the food-frequency record and 0·6 mg/d according to the weighed records. Other breakfast cereals were consumed in much smaller amounts and questions and codes for them were therefore not included in the checklist. However, in the weighed records the average daily consumption of only 2 g bran flakes and 1 g Sultana Bran contributed 0·8 mg and 0·3 mg Fe/d respectively.

Intake of fat was also underestimated but milk and fats were calculated to be consumed in significantly higher amounts in the checklist than measured on the PETRA weighed records (Table 5). Butter, cheddar cheese, polyunsaturated margarine and whole milk were the main contributors to fat intake in both methods (21% weighed, 26% checklist). NSP consumption according to the checklist was also lower than according to the weighed records. Wholemeal bread, apples and muesli supplied the major proportion of NSP, 19% in both cases, and differences in their consumption did not account for any of the difference in NSP between the two methods. There was however a significant difference in calculated vegetable consumption, particularly 'other' (not leafy or root) vegetables (other vegetables, weighed records 98 (SD 47) g, checklist 68 (SD 35) g, $P < 0.001$). Since portion weights for the checklist were similar to those of the weighed record, this difference must have been due to underestimation of the frequency of consumption in the checklist compared with weighed records.

Structured 7 d checklist with individual portion weights (food frequency menu record with portions). Table 5 shows that nearly all nutrients were assessed with a significantly higher mean intake by this method than by the weighed records, and Table 6 shows that intakes of cereal, milk, vegetables, potatoes and beverages were all significantly greater. Most of

Table 7. *Daily food and nutrient consumption values obtained using an unstructured 7 d estimated-diet record compared with weighed records completed by eighty-one women aged 50–65 years‡*

(Mean values and standard deviations)

Nutrients	16 d weighed records All four seasons		7 d estimated record Season 3		4 d weighed records Season 3	
	Mean	SD	Mean	SD	Mean	SD
Energy (MJ/d)	7·68	1·46	7·89	1·49	7·69	1·64
Fat (g/d)	75	18	80**	23	76	24
Protein (g/d)	68	13	69	12	68	15
Carbohydrates (g/d)	217	50	215	47	212	53
Sugars (g/d)	105	35	106	33	105	38
Starch (g/d)	109	30	104	25	103	32
NSP (g/d)	16	5	15†	4	16	5
'Fibre' (g/d)	22	7	22	6	22	7
Potassium (g/d)	3·2	0·7	3·2	0·6	3·2	0·8
Calcium (mg/d)	935	238	864***	248	909	273
Iron (mg/d)	12·9	3·7	12·4	2·9	12·8	4·0
Carotene (mg/d)	3·8	2·2	3·2	1·8	3·1	2·4
Retinol (µg/d)	617	710	740	821	638	988
Vitamin C (mg/d)	102	47	106	48	114	57
Alcohol (g/d)	10	10	9	10	12	13

Food (g/d)	16 d weighed records All four seasons		7 d estimated record Season 3	
	Mean	SD	Mean	SD
Total cereals	235	80	203	68
Bread	84	36	90	37
Milk and milk products	284	143	329	152
Eggs	24	19	18	18
Fats	14	8	17	12
Meat	88	42	93	46
Fish	34	23	38	27
Vegetables	277	90	267	83
Potatoes	84	42	95	51
Fruit	224	145	216	137
Nuts	3	5	3	10
Sugars	25	21	25	20
Beverages	953	347	1084	367
Alcoholic beverages	99	108	121	163
Sauces	344	215	244	229

NSP, non-starch polysaccharides.

Mean values were significantly different from those obtained by 16 d weighed record: ** $P < 0.01$; *** $P < 0.001$; † $P < 0.0001$.

‡ For details of subjects and procedures, see pp. 620–628.

Table 8. *Comparison of Spearman correlation coefficients for values obtained from the average of 16 d weighed records and those obtained from seven different dietary assessment methods completed by 146* women aged 50–64 years†*

	Oxford food-frequency questionnaire Season 3	Cambridge food-frequency questionnaire Season 1	24 h recall (unstructured) Season 1	24 h recall (structured) Season 2	7 d structured checklist Season 1	7 d structured checklist and portions Season 4	7 d estimated-diet record Season 3
Energy	0·52	0·32	0·42	0·46	0·60	0·53	0·59
Fat	0·55	0·35	0·40	0·40	0·62	0·52	0·63
Protein	0·43	0·13	0·21	0·34	0·51	0·46	0·66
Carbohydrates	0·55	0·41	0·60	0·60	0·63	0·67	0·71
Sugars	0·51	0·44	0·63	0·65	0·68	0·70	0·77
Starch	0·53	0·39	0·56	0·38	0·55	0·52	0·70
NSP	0·57	0·36	0·61	0·49	0·70	0·66	0·74
'Fibre'	0·55	0·33	0·58	0·45	0·73	0·67	0·73
Potassium	0·39	0·24	0·56	0·47	0·54	0·49	0·76
Calcium	0·50	0·32	0·28	0·57	0·60	0·59	0·67
Iron	0·43	0·26	0·53	0·36	0·49	0·68	0·83
Carotene	0·45	0·37	0·28	0·21	0·48	0·47	0·66
Retinol	0·55	0·32	0·54	0·33	0·50	0·29	0·35
Vitamin C	0·54	0·41	0·54	0·35	0·66	0·65	0·70
Alcohol	0·90	0·89	0·60	0·54	0·90	0·87	0·88

NSP, non-starch polysaccharides.

* n 73 in the comparison of 7 d estimated-diet record with 16 d weighed records; n 127 in the comparison of Oxford food-frequency questionnaires with 16 d records.

† For details of subjects and procedures, see pp. 620–628.
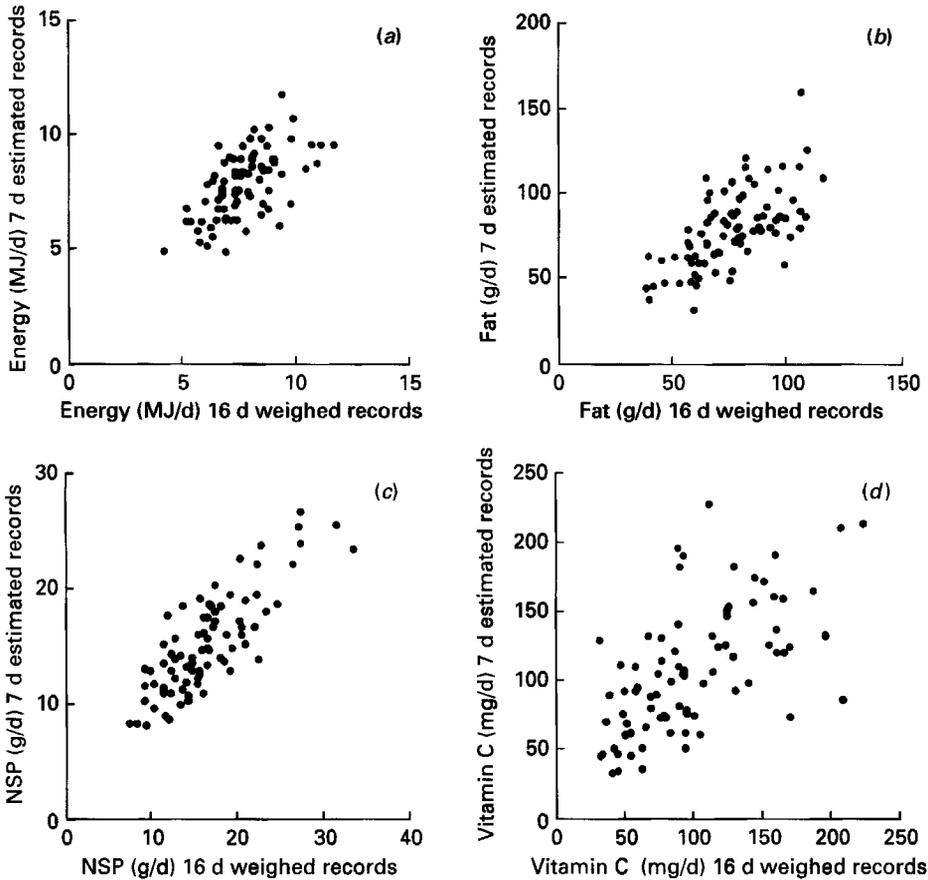
Fig. 3. Scatter plots of the results obtained from 7 d estimated-diet records in relation to those obtained from 16 d records for (a) energy (r 0·59), (b) fat (r 0·63), (c) non-starch polysaccharides (NSP) (r 0·74) and (d) vitamin C (r 0·70). Information concerning copies of the black and white version can be obtained from Dr M. Wadsworth, Department of Community Medicine, University College, London, and concerning copies of the coloured European Prospective Investigation of Cancer (EPIC) version and related software can be obtained from Ms S. Oakes, EPIC, Institute of Public Health, Cambridge.

these differences were attributable to larger portion sizes estimated in this method where photographs were used. Examples of differences in average portion size include muesli 80 g v. 50 g weighed, carrots 130 g v. 65 g weighed, potatoes 160 g v. 108 g weighed, bread 84 g v. 50 g weighed.

Unstructured 7 d food diary (estimated 7 d record). Table 7 shows that there were few differences in mean intakes of nutrients measured by this method compared with those measured by the 16 d of weighed records and that these differences were no longer significant when compared with the averages of nutrients assessed at the same season the diary was administered, season 3. There were no significant differences in average food consumption between the two methods.

Agreement in individual results

Correlation coefficients. One hundred and forty-six individuals completed 4 d of weighed dietary records at each season. Table 8 shows Spearman rank correlation coefficients between the individual values obtained from each method and the averages of all 16 d

Table 9. *Comparison of the percentage classification of results from seven different dietary assessment methods into the same quartile of the distribution of nutrient intake as that obtained from the mean of 16 d weighed records obtained from 146* women aged 50–65 years†*

	Oxford food-frequency questionnaire Season 3	Cambridge food-frequency questionnaire Season 1	24 h recall (unstructured) Season 1	24 h recall (structured) Season 2	7 d checklist Season 1	7 d checklist and portions Season 4	7 d estimated-diet record Season 3
Energy	46	30	37	41	53	39	40
Fat	38	37	38	39	47	40	44
Protein	39	30	34	39	45	38	49
Carbohydrates	39	37	41	43	52	46	38
Sugars	44	34	56	47	55	46	58
Starch	43	37	46	39	40	40	47
NSP	43	39	44	38	46	50	51
'Fibre'	47	34	49	44	49	48	47
Potassium	35	33	42	41	44	42	58
Calcium	46	27	35	42	46	40	51
Iron	39	33	41	37	48	35	56
Carotene	40	38	30	30	43	37	42
Retinol	47	34	45	32	41	36	37
Vitamin C	37	38	39	36	48	44	48
Alcohol	75	68	39	33	76	65	70

NSP, non-starch polysaccharides.

* *n* 73 in the comparison of 7 d estimated-diet record with 16 d weighed records; *n* 127 in the comparison of Oxford food-frequency questionnaires with 16 d records.

† For details of subjects and procedures, see pp. 620–628.

Table 10. *Comparison of percentage misclassification of results from seven different dietary assessment methods into extreme quartiles of the distribution of nutrient intake, compared with that obtained from the mean of 16 d records obtained from 146* women aged 50–65 years†*

	Oxford food-frequency questionnaire Season 3	Cambridge food-frequency questionnaire Season 1	24 h recall (unstructured) Season 1	24 h recall (structured) Season 2	7 d checklist Season 1	7 d structured checklist and portions Season 4	7 d estimated-diet record Season 3
Energy	2	8	6	4	1	3	4
Fat	2	6	6	4	1	4	3
Protein	4	9	9	8	4	3	0
Carbohydrates	5	4	1	3	2	1	0
Sugars	5	3	1	3	2	1	1
Starch	5	4	3	6	1	3	0
NSP	2	6	3	3	1	2	1
'Fibre'	2	4	3	7	1	1	0
Potassium	6	8	1	5	2	6	1
Calcium	6	4	7	2	2	1	1
Iron	4	8	1	5	4	2	0
Carotene	6	6	7	10	4	3	0
Retinol	3	6	1	7	3	10	10
Vitamin C	2	3	3	5	0	1	0
Alcohol	0	0	9	7	0	0	0

NSP, non-starch polysaccharides.

* *n* 73 in the comparison of 7 d estimated-diet record with 16 d weighed records; *n* 127 in the comparison of Oxford food-frequency questionnaires with 16 d records.

† For details of subjects and procedures, see pp. 620–628.

weighed records. Correlation coefficients were highest for alcohol, and lowest for protein, Fe and carotene. The Cambridge food-frequency questionnaire performed least well on this analysis, with coefficients ranging from 0·13 to 0·41 excluding alcohol, compared with 0·39 to 0·57 in the Oxford questionnaire. Correlation coefficients between the 16 d weighed records and the Cambridge food-frequency questionnaire examined in groups 1 and 2 separately were, respectively, energy 0·38, 0·27; fat 0·35, 0·35; protein 0·16, 0·09; starch 0·34, 0·45; sugars 0·53, 0·38; vitamin C 0·50, 0·32.

Correlation coefficients were of the same order for both 24 h recall methods, and ranged from 0·21 to 0·60 in the simple method and from 0·14 to 0·65 in the structured method with pictures to assess portion size. Correlation coefficients were highest for all three record methods, and tended to be of a greater magnitude between the 16 d of weighed records and the 7 d open-ended food diary. Fig. 3 shows plots of the individual values for energy (r 0·59), fat (r 0·63), NSP (r 0·74) and vitamin C (r 0·70). Correlation coefficients between the 16 d weighed records and the 7 d checklist were also examined for groups 1 and 2 separately. Coefficients for groups 1 and 2 respectively were energy 0·55, 0·65; fat 0·53, 0·69; protein 0·45, 0·57; starch 0·67, 0·60; sugars 0·67, 0·69; vitamin C 0·68, 0·69.

Classification into quartiles. Tables 9 and 10 show the extent to which each method was able to classify individuals into the same quartile of intake and to misclassify into opposite quartiles. Alcohol was poorly classified by the 24 h recall but for other methods it was correctly classified in over 60% of individuals and less than 10% of individuals were misclassified. The food records tended to classify a higher proportion of individuals into the correct quartile, but otherwise little difference was evident either between methods or between nutrients. In general, methods correctly classified 30–50% of individuals into their correct quartile of intake for most nutrients, and misclassified in extreme quartiles 0 to 10%. Overall the 7 d estimated record classified fewest individuals into the incorrect fourth of the distribution for most nutrients, although it was less accurate than other methods for retinol.

DISCUSSION

This comparison of various methods of dietary assessment was carried out on a group of individuals who had come through a number of stages of recruitment, so that only 15% of the initial population approached finally completed the study. Hence the average data reported here cannot be said to be representative of the true mean intake of women of this age living in the Cambridge area. Table 2 shows that, compared with the results gained from a representative sample of the UK population, the average results obtained from the weighed records in this group of women were generally greater. However, the aim of the present survey was to assess the ability of different survey methods to classify individuals into the distribution of values, rather than to survey the population mean.

Over the year, up to eighteen home visits were made to each individual by the study team to instruct in the methods being used, and to collect biological samples. Hence, the results obtained from each method must represent the 'best estimate' possible from a highly cooperative group of women. Further studies are under way in men and in a more representative population sample to determine whether similar results would be obtained from these groups. There was no evidence of a socio-economic bias in terms of ability to undertake dietary measurements in the present study, however, since there was no association between social class index and under-reporting (Bingham *et al.* 1994).

The baseline dietary measurement was 16 d of weighed records using the PETRA system, divided into 4 d periods or seasons across 1 year. The validity of these measurements was in itself assessed from 24 h urine collections, as described elsewhere (Bingham *et al.* 1991, 1994). This protocol had previously been devised in order to accumulate sufficient numbers

of observations to reduce the error associated with random (daily) within-person variation in nutrient intake to $\pm 10\%$ standard error on average for energy and energy-yielding nutrients, based on existing literature estimates (Bingham, 1987). Table 3 shows that the estimates of within-person variation found in this survey are very similar to those reported previously from a variety of different population groups and that these limits of precision were achieved for energy and macronutrient intakes. However, the variability of micronutrients is greater and 16 d of weighed intake was associated with 30% standard error in the individual mean for vitamin C.

A major problem in nutritional epidemiology is that of collinearity, which interferes with the ability to distinguish the specific nutrient related to an end-point when nutrients are highly correlated. However, the majority of nutrients were not correlated in this study (Table 4). Fat intake was highly correlated with total energy, but not when considered as a percentage of total energy. The magnitude of correlations between sugar and starch was much less than those with total carbohydrates and energy, and would be considered separately in studies of diet and cancer.

The aim of the present investigation was to examine to what extent other methods, less troublesome for participants, could replace weighed records in the assessment of habitual diet for a large prospective study of diet and cancer. Despite the known problems of a single day's 24 h recall method, the simple unstructured version used in this survey compared surprisingly well with weighed records. Apart from comparatively small differences in starch intake, probably arising from differences in assigned portion weights for bread and potatoes, there were no major biases in average macronutrients or food group consumption compared with results obtained from weighed records. The difference in vitamin C consumption was no longer significant when the results were compared with those obtained from the 4 d weighed records kept at the same time of year at which the 24 h recall was filled in by the subjects, season 1. Although not intended to assess habitual diet, the simple 24 h recall was able to classify individuals to almost the same extent within the distribution as those methods which asked about food use over the year (the food-frequency questionnaires) as judged by comparisons based on correlation coefficients, and classification into appropriate quartiles. Since only a single 24 h recall was used, even closer agreement would be expected if repeat measures were to be available.

There is considerable discussion in the literature concerning the most appropriate statistical method for comparison of results in individual dietary surveys (Woolf, 1954; Bland & Altman, 1986; Willett, 1990). Most published reports use the correlation coefficient to compare individual results within the overall mean, and these are reported here for comparison with these previous reports (Block, 1982; Willett, 1990; Nelson, 1991). Correlation coefficients between the questionnaires and weighed records were generally in the order 0·4 to 0·6 for the Oxford version, and 0·3 to 0·4 in the Cambridge version (excluding alcohol). These are similar to values comparing a 131-item questionnaire with means of 14 d records (not weighed) used in the recently published US male health professionals prospective study (Rimm *et al.* 1992), and lower than those of 0·5 to 0·7 obtained between comparisons of a combined photograph booklet and 276-item questionnaire and means of 24 d of unweighed food records in men in Finland (Pietinen *et al.* 1988). Comparisons with a single 24 h record and a sixty-five-item questionnaire in 433 UK men and women gave lower (0·2 to 0·4) values (Margetts *et al.* 1989), as did a fifty-four-item questionnaire and 7 d weighed record comparison in 119 men in South Wales, where *r* values ranged from 0·27 for carbohydrate to 0·37 for total fibre (Yarnell *et al.* 1983). Results from a microcomputer 120-item questionnaire gave similarly low correlation coefficients (0·16 to 0·37) when compared with those obtained from a 7 d weighed dietary record (Engle *et al.* 1990).

The use of correlation coefficients in this type of analysis has been criticized however, because the magnitude of coefficients depends partly on the range of observations (Woolf, 1954). Without exception, high correlations between measures of alcohol consumption in questionnaires and other methods are obtained, due to the fact that a substantial proportion of the individual values obtained are zero. This is also true of the findings of the present study, where correlation coefficients for alcohol consumption were in the order of 0·9 for methods other than the 24 h recalls (Table 8). To avoid this criticism, an assessment of the ability of the simpler techniques to classify individuals into the same quartile of the population distribution as defined by the baseline method was also considered. Misclassification into extreme quartiles occurred to a greater extent with the food-frequency questionnaires and 24 h recalls than with the food record methods. Similar levels of misclassification using a 140-item food-frequency questionnaire which included estimates of portion size were achieved in a comparison against 8 d records (Mares-Perlman et al. 1993).

Both statistical comparisons ranked the accuracy of simpler methods of dietary assessment in a similar way when compared with weighed records. Using both criteria, diet records gave the best agreement with weighed records. Generally there was little to choose between the 7 d estimated record (7 d diary) and the structured checklist (food-frequency menu record) version. The latter is precoded for computer nutrient analysis, and hence is capable of further development to allow machine readability in surveys of very large numbers of subjects, as is presently the case with food-frequency questionnaires. One advantage of this method is that there is no evidence of bias from overestimation of the frequency of food consumption, for example in vegetable consumption which occurred in both Oxford and Cambridge food-frequency questionnaires.

The inclusion of opportunities for subjects to assess their own portion size did not improve the accuracy of the checklist, rather the reverse. Overestimation from this method was explained by overestimation of portion size and it was not investigated further. However, inclusion of photographs in the 7 d estimated record and in the structured 24 h recall was not associated with greater error. Average portion weights obtained from the first 4 d of weighed records kept by subjects in group 1 were assigned to the checklist and Cambridge food-frequency questionnaire, rather than average portion weights published elsewhere (Crawley, 1988), because these were more appropriate for an all-female population. Correlation coefficients between the checklist results and those obtained from the 16 d weighed records in group 1 were rather lower than those found in group 2, in which the portion size data would have been independent of the reference method. The relatively high level of agreement between the reference method and the checklist cannot therefore be attributed to common portion-size data. Correlation coefficients between the Cambridge food-frequency results and the weighed records were higher in group 2 than in group 1, but correlations from this method were generally low, and lower than those obtained from the Oxford questionnaire.

A key problem that cannot be resolved with 'list' or structured precoded methods of dietary analysis, both retrospective and prospective, is their inflexibility and inability to cope with a variety of dietary hypotheses likely to emerge over the course of a prospective study lasting several years. Although it is possible to develop questionnaires generating results that can be shown to be highly correlated with standard methods such as weighed records, this is only the case if single or related nutrients are of interest. The problem is much more difficult if the aim is to measure the total diet, as in prospective studies of diet and cancer where most nutrients and food groups have been associated either with protection or causation of cancer at different sites. Given a finite number of items that can be included in a questionnaire, the choice depends largely on the nutrients or food of interest at the time the questionnaire was devised and the accuracy of estimates of intake

of items not originally assessed will remain in doubt (Sempos, 1992). In the present study for example, estimates of starch and NSP using the checklist were relatively accurate, but the omission of two choices of Fe-fortified breakfast cereal led to significant bias in the overall mean Fe estimates.

Due to the concentration of interest on particular nutrients, comparative studies in the dietary survey literature do not usually investigate how well the primary information on food consumption relates to that obtained in the standard method. Investigation of this however emphasizes the importance of gaining accurate measures of both portion size and frequency of consumption. Both food-frequency questionnaires yielded evidence of overestimation of consumption of a number of categories of foods, especially vegetables. The problem of overestimation of frequency is well known in questionnaires and in the present survey, for example, vegetable consumption assessed by questionnaire was almost double that assessed from weighed records. This led to a significant overestimation of carotene, vitamin C and NSP in the food-frequency questionnaire results compared with the weighed records. Overestimation did not seem to be related to the length of the list of choices since the list differed in the two questionnaires used. When subjects were presented with the same list used in the Cambridge food-frequency questionnaire, but asked in the checklist to make their choices on a daily basis, rather than choose a category of frequency of consumption, overestimation did not occur.

A further problem arose in the choice of portion sizes, particularly for milk. The importance of milk as a source of energy, protein, fat, Ca and sugars, and inaccurate estimation of consumption, led to significant under and overestimation respectively in the Cambridge and Oxford versions of the food-frequency questionnaires. A revised version of the Oxford food-frequency questionnaire, in which the format of questions concerning milk has been altered, is presently undergoing trials to see if this problem can be resolved. Despite these problems, reported correlation coefficients appeared to be no worse than those reported elsewhere in studies using this method.

The disadvantage of the structured 7 d diary method (estimated-food record) is the time needed to code it for nutrient analysis, although this has been shortened substantially with the computer program DIDO developed especially for use with this diary (F. Key, A. Paul, A. Harter, G. Price, T. Cole and M. Wadsworth, unpublished results). Furthermore, a nested-case-control approach is to be used in the UK arm of the EPIC study, so that it will be necessary to code only the estimated records for a few thousand of the 50 000 subjects to be recruited. Preliminary studies have shown that compliance with this method is similar, and probably preferred by manual workers, to that obtained with the checklist. A significantly greater number of items were recorded by manual-graded civil servants in the open-ended 7 d diary than in the checklist, although there was no such difference amongst non-manual grades (M. Beksinska, personal communication).

Overall, the unstructured 7 d diary (estimated record) had the highest correlation coefficients and was able to classify a greater proportion of individual values into the correct quartile of the distribution. No bias in mean intakes of either food or nutrients was obtained, and Fig. 3 shows that results obtained from it were almost as good as those obtained from 16 d weighed records. For this reason, and because of the flexibility of a food diary in investigating any further hypotheses, it is to be used as the main method in the EPIC study, with repeat investigations as the cohort progresses over time.

to assess portion size, and the sponsorship of Beach Villas and J. Sainsbury PLC are gratefully acknowledged. Dr M. Wadsworth is thanked for the 7 d food diaries. The general practitioners who assisted in the recruitment of volunteers, are also thanked. Additional support for this study was obtained from the Cancer Research Campaign, the Ministry of Agriculture, Fisheries and Food and the Vegetarian Society.

REFERENCES

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

Bingham, S. A., McNeill, N. I. & Cummings, J. H. (1981). The diet of individuals: a study of a randomly chosen cross section of British adults in a Cambridgeshire village. *British Journal of Nutrition* **45**, 23–35.

Bingham, S. A., Pett, S. & Day, K. C. (1990). Non-starch polysaccharide intake of a representative sample of British adults. *Journal of Human Nutrition and Dietetics* **3**, 333–337.

Bingham, S. A., Cassidy, A., Cole, T., Welch, A., Runswick, S. A., Black, A. E., Thurnham, D., Bates, C., Khaw, K. T. & Day, N. E. (1994). 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* (In the Press).

Bingham, S. A., Welch, A., Cassidy, A., Runswick, S., Gill, C. & Khaw, K. T. (1991). The use of 24 h urine nitrogen to detect bias in the reported habitual food intake of individuals assessed from weighed dietary records. *Proceedings of the Nutrition Society* **50**, 32A.

Bland, J. M. & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* **i**, 307–310.

Block, G. (1982). A review of validations of dietary assessment methods. *American Journal of Epidemiology* **115**, 492–505.

Braddon, F. E. M., Wadsworth, M. E. J., Davies, J. M. C. & Cripps, H. A. (1988). Social and regional differences in food and alcohol consumption in Britain. *Journal of Epidemiology and Community Health* **42**, 341–349.

Callmer, E., Riboli, E., Saracci, R., Akesson, B. & Lindegarde, F. (1993). Dietary assessment methods evaluated in the Malmo study. *Journal of Internal Medicine* **233**, 53–57.

Crawley, H. (1988). *Food Portion Sizes.* London: H.M. Stationery Office.

Edington, J., Thorogood, M., Geekie, M., Ball, M. & Mann, J. (1989). Assessment of nutritional intake using dietary records with estimated weights. *Journal of Human Nutrition and Dietetics* **2**, 407–414.

Engle, A., Lynn, L. L., Koury, K. & Boyar, A. P. (1990). Reproducibility and comparability of a computerised, self administered food frequency questionnaire. *Nutrition and Cancer* **13**, 281–292.

Gregory, J., Foster, K., Tyler, H. & Wiseman, M. (1990). *The Dietary and Nutritional Survey of British Adults.* London: H.M. Stationery Office.

Holland, B., Unwin, I. D. & Buss, D. H. (1988). *Cereals and Cereal Products, Third Supplement to McCance and Widdowson's Composition of Foods.* London: H.M. Stationery Office.

Holland, B., Unwin, I. & Buss, D. H. (1989). *Milk, Milk Products and Eggs. Fourth Supplement to McCance and Widdowson's The Composition of Foods.* London: H.M. Stationery Office.

Mares-Perlman, J., Klein, B. E., Klein, R., Ritter, L., Fisher, M. R. & Freudenheim, J. L. (1993). A diet history questionnaire ranks nutrient intakes in middle aged and older men and women similarly to multiple food records. *Journal of Nutrition* **123**, 489–501.

Margetts, B. M., Cade, J. E. & Osmond, C. (1989). Comparison of food frequency questionnaire with a diet record. *International Journal of Epidemiology* **18**, 868–873.

National Food Survey Committee (1982). *Annual Report of Household Food Consumption and Expenditure: 1980.* London: H.M. Stationery Office.

Nelson, M. (1991). Assessment of food consumption and nutrient intake: past intake. In *Design Concepts in Nutritional Epidemiology*, pp. 167–191 [B. M. Margetts and M. Nelson, editors]. New York: Oxford University Press.

Paul, A. A. & Southgate, D. A. T. (1978). *McCance and Widdowson's The Composition of Foods*, 4th ed. London: H.M. Stationery Office.

Pietinen, P., Hartman, A. M., Haapa, E., Rasenen, L., Haapakovski, J., Plamgren, J., Albanes, D., Virtamo, J. & Huttunen, J. K. (1988). Reproducibility and validity of dietary assessment instruments. 1. A self administered food use questionnaire with a portion size booklet. *American Journal of Epidemiology* **128**, 655–666.

Rimm, E. B., Giovanucci, E. L., Stampfer, M. J., Colditz, G. A., Litin, L. B. & Willett, W. C. (1992). Reproducibility and validity of an expanded self administered semiquantitative food frequency questionnaire among male health professionals. *American Journal of Epidemiology* **135**, 1114–1126.

Sempos, C. T. (1992). Some limitations of semiquantitative food frequency questionnaires. *American Journal of Epidemiology* **135**, 1127–1133.

Wiles, S. J., Nettleton, P. A., Black, A. E. & Paul, A. A. (1980). The nutrient consumption of some cooked dishes eaten in Britain: a supplementary consumption table. *Journal of Human Nutrition* **34**, 189–223.

Willett, W. C. (1990). Reproducibility and validity of food frequency questionnaires. In *Nutritional Epidemiology*, pp. 92–126 [W. C. Willett, editor]. New York: Oxford University Press.

Willett, W. C., Sampson, L., Stampfer, M., Rosner, B., Bain, C., Witschi, J., Hennekens, C. H. & Speizer, F. E. (1985). Reproducibility and validity of a semi-quantitative food frequency questionnaire. *American Journal of Epidemiology* **122**, 51–65.

Woolf, B. C. (1954). Statistical aspects of dietary surveys. *Proceedings of the Nutrition Society* **12**, 82–94.

Yarnell, J. W. G., Fehily, A. M., Milbank, J. E., Sweetnam, P. M. & Walker, C. L. (1983). A short dietary questionnaire for use in an epidemiological survey: comparison with weighed dietary records. *Human Nutrition: Applied Nutrition* **37A**, 103–112.