Attainment of precision in implementation of 24 h dietary recalls: INTERMAP UK

Claire Robertson[1]*, Rana Conway[2], Barbara Dennis[3], John Yarnell[4], Jeremiah Stamler[5] and Paul Elliott[1] for the INTERMAP Research Group

[1]*Department of Epidemiology and Public Health, Imperial College London, St Mary's Campus, Norfolk Place, London W2 1PG, UK*
[2]*Department of Nutrition and Dietetics, King's College London, 150 Stamford Street, London SE1 8WA, UK*
[3]*Department of Biostatistics, Collaborative Studies Coordinating Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA*
[4]*Department of Epidemiology and Public Health, The Queen's University of Belfast, Mulhouse Building, Grosvenor Road, Belfast BT12 6BA, UK*
[5]*Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA*

Collection of complete and accurate dietary intake data is necessary to investigate the association of nutrient intakes with disease outcomes. A standardised multiple-pass 24 h dietary recall method was used in the International Collaborative Study of Macro- and Micronutrients and Blood Pressure (INTERMAP) to obtain maximally objective data. Dietary interviewers were intensively trained and recalls taped, with consent, for randomly selected evaluations by the local site nutritionist (SN) and/or country nutritionists (CN) using a twelve-criterion checklist marked on a four-point scale (1, retrain, to 4, excellent). In the Belfast centre, seven dietary interviewers collected 932 24 h recalls from 40–59-year-old men and women. Total scores from the 134 evaluated recalls ranged from thirty-four to the maximum forty-eight points. All twelve aspects of the interviews were completed satisfactorily on average whether scored by the SN (*n* 53, range: probing 3·25 to privacy of interview 3·98) or CN (*n* 19, range: probing 3·26 to pace of interview and general manner of interviewer 3·95); the CN gave significantly lower scores than the SN for recalls evaluated by both nutritionists (*n* 31, Wilcoxon signed rank test, $P=0.001$). Five evaluations of three recalls identified areas requiring retraining or work to improve performance. Reporting accuracy was estimated using BMR; energy intake estimates less than 1·2 × BMR identifying under-reporting. Mean ratios in all age, sex and body-mass groups were above this cut-off point; overall, 26·1 % were below. Experiences from the INTERMAP Belfast centre indicate that difficulties in collection of dietary information can be anticipated and contained by the systematic use of methods to prevent, detect and correct errors.

Methodology of 24 h recalls: Implementation: INTERMAP: Precision

The International Collaborative Study of Macro- and Micronutrients and Blood Pressure (INTERMAP) aimed to investigate associations between nutrient intakes of individuals and their blood pressures (Stamler *et al.* 2003). Altogether, 4680 men and women aged 40–59 years were recruited from seventeen diverse population samples in Japan, The People's Republic of China, UK and USA. Dietary intake was assessed on four occasions with a multiple-pass 24 h recall method used extensively in the USA (Gorder *et al.* 1986; Tillotson *et al.* 1986; Phillips *et al.* 2000; Dennis *et al.* 2003; Dwyer *et al.* 2003*b*; Hajjar & Kotchen, 2003) and less frequently in the UK. The Nutrition Data System for Research from the Nutrition Coordinating Center, University of Minnesota has facilitated the use of this method (Schakel, 2001) through the incorporation of an interview structure and food and nutrient database in one computerised software system, reducing the burden of data collection, processing and

quality control (QC) usually associated with 24 h recalls (Sievert *et al.* 1989). Adaptation of existing methodologies, for example, creating an interview system specific to INTERMAP for completion either on-line (USA) or on paper and then transcribed to computer (Japan, China and UK), was completed in all four INTERMAP countries. This helped take advantage of the in-house training, built-in QC checks and procedures contained in the Nutrition Data System, and the considerable knowledge by senior INTERMAP researchers of their use.

Accurate assessment of how dietary factors impact disease outcomes is determined by the precision of all involved measurements (Hilner *et al.* 1992). In INTERMAP, multifaceted approaches were used to reduce imprecision associated with dietary assessment, including several QC procedures. Reports of such systems and their results are essential to help explain disparities between studies, enable assessment of the reliability of presented

data, and enhance confidence in the accuracy and merit of nutrition epidemiology studies (van Horn *et al.* 1990; Hilner *et al.* 1992; Dwyer *et al.* 2003*a*).

Collection of dietary intake data requires integration of procedures to ensure accurate and reliable estimation of nutrient intakes (Dwyer *et al.* 2003*a*). No central computerised data collection and coding system is as yet available in the UK. Local research groups develop their own methods to collect and process dietary information. This requires considerable effort, especially in the design of protocols to reduce potential reporting inaccuracies associated with dietary assessment (Karvetti & Knuts, 1985; Bingham, 1987, 1991; Bingham *et al.* 1994; Dennis *et al.* 2003). Given these acknowledged difficulties, the aim of the present paper is to share experiences on types of problems identified in one of the INTERMAP UK centres (Belfast) and results with methods developed before fieldwork to help prevent and resolve them.

Methods

Participants recruited for INTERMAP were seen on four separate occasions at study centres. Detailed information on what was required throughout the four-visit protocol was given, and informed consent obtained before commencement of fieldwork. Appointments were scheduled as two pairs of visits on consecutive days, with a 3–6-week period between the second and third appointments. Participants seen in the Belfast centre were largely (96·3 %) recruited from the Northern Ireland Civil Service, based at Stormont Estate. Numbers were supplemented to reach required sample size (*n* 260) by recruiting participants from a general practice in West Belfast.

Methods, including a detailed account of the recruitment rationale and process, are described in detail elsewhere (Dennis *et al.* 2003; Stamler *et al.* 2003). In brief, dietary interviewers – trained and certified by senior INTERMAP investigators – collected information regarding the previous day's intake using standardised, semi-structured, three-phase 24 h recalls at each clinic visit (Fig. 1). Procedures were kept as informal as possible, within set protocols and dialogues to enhance standardisation and completeness, i.e. ensure no aspect of questioning was missed. During the recall, participants were first asked to report freely their consumption over the 24 h period, midnight to midnight, for the day before interview. While no particular order was asked of INTERMAP participants, for consistency across all centres and countries, and all interviews commenced with the question 'What was the first thing you ate...?'.

The initial food consumption list included dining location and eating times to mark each eating event and identify long periods with no reported intake, for possible prompting of overlooked consumption. Once complete, open questioning techniques were used to gain information on food brands and portion sizes (estimated using food models, standard plates and containers with calibrated coloured lines, food photographs, and standardised portions such as a can of cola – see Fig. 2) and additions made to items eaten. The interviewer next asked the participant to look at a 'foods commonly forgotten' sheet, to prompt recall of forgotten items. At this time, a 'documentation checklist' was used to ensure coverage of all areas requiring probing for information; for example, 'Was the skin on chicken eaten or not?' or 'Were salt, vinegar and sauces added to chips or not?'. Interviewers were trained to document all information, for example, if no
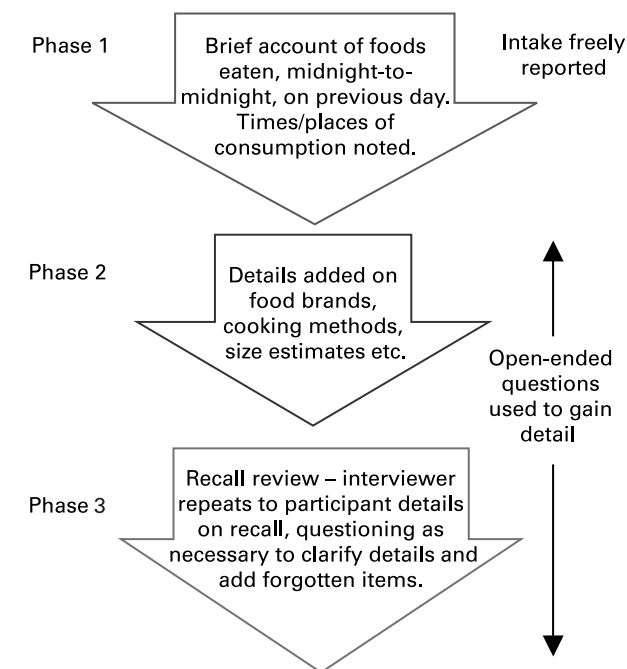

Fig. 1. Three-phase 24 h dietary recall methodology used in the International Collaborative Study of Macro- and Micronutrients and Blood Pressure (INTERMAP) UK.

salt was added to potatoes while boiling them, or no milk or sugar added to coffee, this was set down, item by item, on the recall form.

The third phase of the interview involved a review. The interviewer repeated to the participant all items set down on the recall form, providing an opportunity to correct any details recorded inaccurately, to ask any further questions on reported intake, and to cover any areas missed during the interview. When participants could not provide sufficient information on the foods they consumed, for example, when they had not prepared or cooked their own meals and therefore could not answer all questions fully, food inquiry to cooks (FIC) forms were initiated for clarification. Participants provided information including a description of the food, constituent ingredients and portion sizes eaten lest forms were not returned; however, if the individual who prepared the meal returned details of its composition, this was assumed

Fig. 2. Food models and amount estimation tools used in the International Collaborative Study of Macro- and Micronutrients and Blood Pressure (INTERMAP) UK.

more accurate, and was used instead. On completion of all recalls, the local site nutritionist (SN), responsible for ongoing local QC checks, completed a visual documentation check. This helped identify and correct unrecorded details (for example, 'no salt added'), ensuring default coding rules (for example, 'if addition of salt unknown, code as salted') did not introduce avoidable errors. Unasked questions were also identified, enabling interviewers to gain necessary additional details at subsequent visits or by telephone for completeness.

In the Belfast centre, seven interviewers collected dietary recalls for INTERMAP. Two were employed as a SN and one nutritionist from the London INTERMAP centre completed 2 weeks of dietary assessment and nutrient coding work in the Belfast centre. Work completed by this interviewer was excluded from final analyses due to small numbers.

All dietary recalls were taped, with interviewee consent, for on-going assessment of interviewer skills. One random recall of every ten collected was evaluated based on the twelve criteria listed below with a brief rationale for their inclusion:

i) Privacy of interview: to avoid disturbances hindering participant recall;
ii) General manner of interviewer: if the interviewer is not interested in what the participant is reporting, this may discourage the participant from reporting accurately;
iii) Introduction by interviewer: making the participant aware of the importance of his/her involvement to encourage complete and accurate data;
iv) Use of non-directed questions: to avoid introduction of interviewer bias;
v) Pace: rushing participants could result in guesswork being used to speed up the recall process, while going too slow could induce boredom;
vi) Manner of questioning: to avoid introduction of bias and inference of boredom;
vii) Objectivity: to avoid introduction of interviewer bias;
viii) Probing: to ensure information collected is complete and nothing is missed;
ix) Use of food models and amount estimation tools: to gain accurate amount information for subsequent nutrient coding;
x) Documentation: to ensure an accurate base for nutrient coding;
xi) Use of memory aids: to prompt reports of forgotten items;
xii) Review of recall: to correct misreported items and poor documentation of details, and collect information on forgotten items.

For each criterion, a four-point scale was used: 1 – retrain; 2 – needs work; 3 – acceptable; 4 – excellent; thus, evaluation scores could range from twelve to forty-eight points. A hand-written report was completed for each tape evaluation and was shared with the interviewer to prevent recurrence of identified problems. This ensured that all interviewers were given feedback on their interview technique even if they did well in all areas, with even small problems highlighted to prevent their repetition. This system was designed by senior INTERMAP investigators and provided an on-going QC check on data-collection methods.

Tape evaluations were completed by both local SN and/or country nutritionists (CN) (responsible for ensuring collection of accurate dietary data from centres within their country). Findings are given here for the INTERMAP Belfast centre. They include recorded type and frequency of errors in dietary interviews during fieldwork, methods to prevent their reoccurrence and assessments of the potential benefit of these QC checks with regard to achieved data accuracy and reliability.

BMR was calculated to assess 'completeness' of dietary recalls (Black *et al.* 1991; Pryer *et al.* 1997; Cook *et al.* 2000). The BMR (MJ) of men aged 30–60 years was calculated as 0·048 weight (kg) + 3·653, and for women aged 30–60 years, 0·034 weight (kg) + 3·538 (Thomas, 1994). Energy intake calculated from the 24 h recalls was then expressed as a ratio, i.e. divided by participant BMR, to estimate mis-, especially under-, reporting (Thomas, 1994); a cut-off level of 1·2 × BMR was used to define under-reporting (Goldberg *et al.* 1991).

The sample comprised white men and women aged 40–59 years, employed in both white- and blue-collar occupations. Over the 12-month period of data collection (April 1998–April 1999), 222 individuals (125 men and ninety-seven women) satisfactorily completed all phases of the study, providing data from 888 24 h recalls; altogether, 932 recalls were collected in Belfast; data from nineteen participants (forty-four recalls) were excluded from the final data file based on criteria stipulated in the INTERMAP protocol: seventeen (7·1 %) completed only two recalls, one (0·4 %) collected less than 250 ml urine over the 24 h collection period, one (0·4 %) was a volunteer and one (0·4 %) had an incomplete urine collection. No more than one participant was recruited from any household to help capture maximum variation in dietary habits. Protocols stipulated that data should still be collected from such participants, but classified as collected from a volunteer. These data therefore remained subject to the same INTERMAP protocols and due to this, QC results collected from even excluded participants is discussed here. No recalls were excluded from analyses due to excessively high or low energy intakes, namely energy intakes on any 1 d less than 2·09 MJ (500 kcal) for men or women, or exceeding 20·92 MJ (5000 kcal) for women or 33·47 MJ (8000 kcal) for men (Stamler *et al.* 2003). Since any collected recall could have been tape evaluated, data from participants excluded from the final INTERMAP data file are included in the results presented here. Of the 932 collected recalls, ten were excluded (one tape evaluation) as they were collected by the interviewer providing short-term support. That left 922 recalls, of which 103 (11·2 %) were randomly selected for evaluation by either a SN only, the CN only, or both the SN and CN. Of 140 evaluations, 134 were completed fully; there were problems with tape quality in six recalls.

Various basic descriptive statistics including frequencies (*n*, %) and means (and standard deviations) are presented. Non-parametric subgroup analyses have been used for ordinal data, assessing: (i) differences in applications of each tape evaluation criterion by the interviewers (Kruskall–Wallis test); (ii) comparability of total evaluation scores and individual interview criterion scores in recalls assessed by the SN and CN (Wilcoxon signed ranks test); (iii) number of FIC forms initiated by each interviewer and returned (ANOVA). All analyses were completed using SPSS, version 11.5 for Windows (SPSS Inc., Chicago, IL, USA).

Results

Of the 103 randomly selected recalls, fifty-three (51·5 %) were evaluated by the SN only, nineteen (18·4 %) by the CN only and thirty-one (30·1 %) by both the CN and SN. Scores from these evaluations ranged from thirty-four to the maximum of

forty-eight (mean 43·88 (SD 2·59)), indicating generally good interview techniques by all six Belfast interviewers (Table 1). Significant differences among local interviewers were evident for their use of the recall review procedure (Kruskall–Wallis test: χ^2 11·634; $P=0.040$), and pace of interview (Kruskall–Wallis test: χ^2 19·468; $P=0.002$).

Between thirteen and twenty-two evaluations of collected recalls were completed for each of the six Belfast interviewers; these constituted 7·7–23·4 % of collected recalls. Tape evaluations completed only by the SN ranged from five to twelve recalls or 3·7–10·6 % of collected recalls per interviewer and evaluations completed only by the CN from zero to seven recalls or 0–7·7 % collected recalls. Only five evaluations (of three recalls) identified areas that required interviewer retraining or work to ensure a satisfactory completion level (scores of one or two) (Table 1). Of the twelve criteria, mean scores were lower for probing (mean 3·29 (SD 0·472)), documentation (mean 3·41 (SD 0·523)), use of non-directed questions (mean 3·44 (SD 0·498)) and manner of questioning (mean 3·46 (SD 0·500)). All areas were completed satisfactorily on average whether scored by the SN (average scores ranged from 3·25 for probing to 3·98 for privacy of interview) or CN (average scores ranged from 3·26 for probing to 3·95 for pace of interview and general manner of interviewer). Based on amalgamated scores from both nutritionists, three criteria were met especially well: privacy of the interviews (3·92 (SD 0·275)), general manner of the interviewer (3·89 (SD 0·319)), and pace of interview (3·74 (SD 0·441)).

Comparison of total scores for dietary interviews reviewed by both a SN and the UK CN (n 31) showed significantly higher scores in twenty-one cases from the SN compared with the CN; in five, the reverse; in five, scores the same (Wilcoxon signed rank test; $P=0.001$). By criterion, disagreements between CN and SN scores were significant for pace of interview (twenty-four tied scores, six SN > CN, zero CN > SN; $P=0.014$); objectivity (twelve tied scores, sixteen SN > CN, three CN > SN; $P=0.003$); use of food models and amount estimation tools (twenty-one tied scores, nine SN > CN, one CN > SN; $P=0.011$) and use of memory aids (twenty tied scores, ten SN > CN, one CN > SN; $P=0.007$).

Interviewers initiated FIC forms in between 34·9 and 62·1 % of collected recalls to clarify information on food items consumed (see Table 1), a significant difference among local interviewers (ANOVA test; $P=0.001$). The proportion of these returned was not significantly different across interviewers (ANOVA test; $P=0.083$).

Of all Belfast participants who provided information for the final INTERMAP data files (n 222), under-reporting – defined as a calculated energy intake:BMR ratio of less than 1·2 × BMR (Poppitt *et al.* 2002) – was evident in fifty-eight (26·1 %). Women were more likely to under-report than men – 30·9 % women and 22·4 % men respectively, a non-significant difference ($P=0.158$). Under-reporting based on this ratio was directly related to BMI; energy intake:BMR ratio was lower with higher BMI (r −0·239; $P<0.001$), indicating a greater tendency for under-reporting with greater BMI (Table 2). Mean ratios for all age, sex and BMI strata were above the 1·2 cut-off point.

Discussion

The principal aim of the INTERMAP is to clarify the relationship of multiple dietary factors to blood pressure (Stamler *et al.* 2003). The taped INTERMAP 24 h recalls provided a unique opportunity qualitatively (and to some extent, quantitatively) to assess the objectivity of methods used to collect dietary data. Achieving accuracy and precision in assessing dietary intakes is a challenge (Todd *et al.* 1983; Morgan *et al.* 1987; Bingham, 1991; Bingham *et al.* 1994, 1997; Grandits *et al.* 1997; Macdiarmid & Blundell, 1998; Black, 2000; Dennis *et al.* 2003). Results from analyses here indicated: (i) consistently good overall application of interviewer skills; (ii) specific areas of interview techniques more of a challenge than others; (iii) value of local and national QC checks to maintain high-level, comparable estimates of accuracy; (iv) good estimated completeness of dietary data in the INTERMAP Belfast centre.

The generally good application of interview techniques observed for all interviewers reflects positively on the intensive training procedure and practice completed before fieldwork. Local SN in each INTERMAP centre were generally more

Table 1. Descriptive summary of taped recalls evaluated for each interviewer

Variable		3006	3007	3008	3009	3010	3025	Total
				Interviewer identity number				
Recalls collected	n	94	130	134	224	194	146	922
	n	36	62	52	139	63	51	403
Food inquiry to cooks forms initiated	%	38·3	47·7	38·8	62·1	32·5	34·9	43·7
	% returned	72·2	77·4	55·8	68·3	92·1	80·4	73·7
Recalls evaluated only by the SN	n	10	7	5	10	12	9	53
	%	10·6	5·4	3·7	4·5	6·2	6·2	5·7
Recalls evaluated only by the CN	n	7	1	0	3	3	5	19
	%	7·4	0·8	0	1·3	1·5	3·4	2·1
Recalls evaluated by both the SN and CN	n	5	6	8	7	0	5	31
	%	2·6	4·6	6·0	3·1	0	3·4	3·4
All* evaluations (SN or CN)	n	22	14	13	20	15	19	103
	%	23·4	10·8	9·7	8·9	7·7	13·0	11·2
Recalls with unsatisfactory scores	n	0	1	0	1	0	1	3
	%	0	0·77	0	0·45	0	0·68	0·33

CN, country nutritionist; SN, site nutritionist.
* Includes recalls evaluated twice; once by the SN, once by the CN.

Table 2. Summary statistics for estimated energy intake, estimated basal metabolic rate (MJ) and energy intake:basal metabolic rate ratio

(Mean values and standard deviations)

Age/sex group	n	Energy intake (MJ/d)*		Calculated BMR (MJ)		Energy:BMR		% <1·2 × BMR
		Mean	SD	Mean	SD	Mean	SD	
All participants	222	9·43	2·54	6·90	1·04	1·37	0·30	26·1
All men	125	10·78	2·43	7·68	0·60	1·41	0·32	22·4
All women	97	7·70	1·37	5·88	0·44	1·32	0·25	30·9
All aged 40−49 years	123	9·32	2·53	6·71	1·04	1·39	0·29	24·4
All aged 50−59 years	99	9·57	2·57	7·13	1·00	1·34	0·30	28·3
Men aged 40−49 years	59	11·01	2·41	7·65	0·61	1·45	0·33	22·0
Men aged 50−59 years	66	10·57	2·45	7·71	0·60	1·37	0·31	22·7
Women aged 40−49 years	64	7·76	1·37	5·84	0·41	1·33	0·24	26·6
Women aged 50−59 years	33	7·57	1·38	5·97	0·49	1·28	0·27	39·4
BMI <25 kg/m^2	76	8·98	2·21	6·18	0·83	1·45	0·27	19·0
BMI <25−30 kg/m^2	107	9·81	2·59	7·16	0·84	1·36	0·30	48·3
BMI >30 kg/m^2	39	9·28	2·93	7·57	1·15	1·21	0·29	32·8

* Energy intake from average of four 24 h dietary recalls per individual.

qualified and experienced in dietary assessment than local interviewers. However, all interviewers completed a 4 d intensive training session with senior investigators, followed by 1 week or more of practice in order to gain certification (Dennis *et al.* 2003). This was the first experience of these methods for all interviewers in the Belfast centre; therefore, it can be assumed that their application of the 24 h recalls is fully comparable. Interviewers were aware that the tape from any of the interviews they conducted might be evaluated. The importance of each of the twelve criteria used in evaluation was outlined in training. Published reports also attest to the effectiveness of training and QC methods to enhance accuracy (Block, 1982; Bingham & Cummings, 1985; Bingham, 1991); the INTERMAP QC procedure has been used in only a small number of other studies (van Horn *et al.* 1990; Hilner *et al.* 1992; Buzzard *et al.* 1998; Tapsell *et al.* 2000).

The CARDIA study (Friedman *et al.* 1988) utilised a tape-evaluation QC procedure, recording errors broadly as two types; deviations from protocol, and recording or documentation errors. Randomly selected recalls were evaluated also for completeness post-interview using visual checks. Of the 5111 diet histories collected, 153 (3·0%) were chosen at random to be audiotaped and evaluated, and 564 (11·0%) forms were reviewed. Deviations from protocol were more common (mean frequency 2·7 (SD 2·3) per tape form) than recording errors (mean frequency 1·3 (SD 1·4) per audiotape) in the tape evaluations. However, the mean form completion error rate for the 153 forms was only 0·31%, significantly more than the mean error rate in the 564 forms evaluated (0·19%; P=0·0351). This could have been the result of nervousness in interviewers knowing their work would be reviewed, or more errors may have been identified using the in-depth audiotape evaluation method. Thus, as in INTERMAP, this time-consuming QC method did identify errors; its usefulness was highlighted for large-scale multi-centre studies including several nutritionists (Hilner *et al.* 1992). In INTERMAP, a much larger mean percentage of recalls was tape evaluated (14·4% in the Belfast centre), and all collected recalls were visually checked for completeness of documentation shortly after the interviews. No formal assessment was made of the latter QC check; however, the mean score for documentation (of the twelve) tape evaluation

criteria was 3·41 (SD 0·52), and considered aside the optimal score in all tape evaluation criteria (4·00), indicates that the visual check is not 100% effective for identification of limitations in interviewers' data documentation.

A recent study using sixty-two audio-taped diet histories involved conversational analysis, and assessed terms like 'it depends,' as indicative of possible prevarication; 'probably,' as conjecture. It noted involved discussions around foodstuffs such as alcohol and chocolate, highlighting food items considered sensitive by the participants (Tapsell *et al.* 2000). Low-fat and high-carbohydrate foods were frequently described without prompting; high-fat foods (for example, chocolate, take-away foods and red meat and pork) encouraged 'elaborations'; alcohol was rarely reported unless prompted (Tapsell *et al.* 2000). Similar situations – particularly the 'elaboration' scenario – were evident in INTERMAP Belfast. In interviews with men particularly, recall of food eaten the previous day was uncertain, and reports including 'I think we had chicken. Yes, it was definitely chicken' were frequently amended following the return of FIC forms, where their partners indicated that they had consumed fish or beef. Extended descriptions were also provided, and advice on intake frequently asked (it was not given). One participant, for example, reported consumption of standard Coca Cola, noting immediately afterwards that 'I would usually have the diet stuff, but yesterday I had the full-sugar variety', a conversation habit repeated on several occasions by this participant when describing other foodstuffs.

In INTERMAP, an interview structure was used to help minimise differences among and within interviewers (Dolecek *et al.* 1997; Willett, 1998; Dennis *et al.* 2003; Conway *et al.* 2004). Since questions were by necessity detailed, it was at times tedious to collect complete and reliable accounts of food intake data. Asking participants multiple questions on what they ate and how much had the potential to produce questioning errors. 'Did you have margarine on your toast?', for example, could readily lead participants to report simply that they did, when they may actually have consumed butter and jam. An undirected – hence proper – question was 'Did you add anything to your toast?' It is not possible to assess in detail possible bias introduced by inadequate questioning techniques, since for free-living individuals, actual intake is not attainable with 100% validity, i.e.

there is no 'gold standard' method (Dennis *et al.* 2003). Given the in-depth nature of the tape-evaluation system, we were quickly able to identify problem areas related to questioning, and – as necessary – retrain interviewers to correct them and minimise frequency of subsequent errors. Consequently, interviewers were repeatedly advised on methods to clarify information, for example, to ask 'How much of that did you eat?' not 'Did you eat all of your toast?'.

Generally, as shown by Tapsell *et al.* (2000), 'core' foods such as cereals, meats and breads are identified unprompted by an interviewer, while additions or components of mixed dishes require prompting. Measures that can enhance accuracy in collecting dietary data also include identifying linguistic choices possibly associated with inaccurate reporting (for example, use of 'probably', or 'it depends'); dealing carefully with sensitive topics (for example, alcohol, fats, confectionery) and responding appropriately to cues related to accuracy of reporting (for example, location of dining, presence of other individuals when eating). Evidence suggests that conducting recalls in reverse (from immediately before the recall time, over the preceding 24 h period) may yield data more accurately reflecting true intake (Smith, 1991). In INTERMAP Belfast (and other studies, for example, Tapsell *et al.* 2000), all participants began their recalls with a description of breakfast. This may not be the case with individuals of other backgrounds.

FIC forms were used in INTERMAP to help overcome inaccurate reporting; an average of 73·7 % were returned completed. Several unreturned forms were sent to catering establishments; repeated telephone contact did not always produce the necessary information. Overall, however, this procedure did correct many misreported facts in 24 h recalls; its merit was therefore confirmed.

In the Belfast centre, of the 134 interview evaluations, only three recalls (five evaluations) contained examples of potentially serious problem areas. No recalls were excluded due to excessively high or low energy intakes. Data on overall participant ratio of reported energy intakes to BMR – used particularly to assess possible under-reporting bias in other studies (Black *et al.* 1991; Pryer *et al.* 1997; Cook *et al.* 2000) – also indicated relatively complete recall of dietary data. The numerous difficulties associated with using BMR to estimate under-reporting are acknowledged. Within-individual variability in energy intake and expenditure renders BMR estimates unstable. However, the Harris–Benedict equation for sedentary lifestyles (1·2 × BMR) was considered the best available option here, with participants in the Belfast sample largely recruited from desk jobs. Results in the Belfast centre indicated that women were more likely to under-report than men (22·4 % men and 30·9 % women); this tendency was also greater with increasing age in both men and women and with body mass. This is consistent with findings from other studies (Livingstone *et al.* 2003). A recent Food Standards Agency study using modelling techniques concluded that misreporting errors cannot be predicted by individual characteristics (for example, BMI, age and sex) (Rowett Research Institute, 2004). However, in the absence of any more accurate method, BMR estimates continue to be used as a check on completeness of estimated energy intakes. The accuracy apparent in the tape evaluations may be questioned by the under-reporting identified here. However, it is suggested here that this would be considerably more pronounced without application of this method.

Conclusion

The data in the present paper and in the available literature attests to the effectiveness of staff training and QC methods to enhance accuracy of dietary intake data (Block, 1982; Bingham & Cummings, 1985; Bingham, 1991). Foreseeing difficulties that may arise during a recall helps interviewers both to avoid them and to deal effectively with them when they occur, maintain and enhance accuracy. This produces greater ability to draw correct inferences concerning relationships of dietary intakes to health outcomes. The 24 h recall method used in INTERMAP drew on the knowledge and experience of senior investigators. The experiences illustrated here are probably familiar to research groups who have used a 24 h recall procedure. For groups with no experience of this method, application of the approaches discussed here is deemed useful to minimise common pitfalls and maximise accuracy in data collection. Given that dietary assessment data are only as good as the information obtained by the interviewer, the food composition database, and the quality of coding (Slimani *et al.* 2000; Dennis *et al.* 2003), the aim must be to maximise accuracy of reporting and minimise bias in methodology.

References

Bingham SA (1987) The dietary assessment of individuals; methods, accuracy, new techniques and recommendations. *Nutr Abst Rev* **57**A, 705–742.

Bingham SA (1991) Limitations of the various methods for collecting dietary intake data. *Ann Nutr Metab* **35**, 117–127.

Bingham SA & Cummings JH (1985) Urine nitrogen as an independent validatory measure of dietary intake: a study of nitrogen balance in individuals consuming their normal diet. *Am J Clin Nutr* **42**, 1276–1289.

Bingham SA, Gill C, Welch A, *et al.* (1997) Validation of dietary assessment methods in the UK arm of EPIC using weighed records, and 24-hour urinary nitrogen and potassium and serum vitamin C and carotenoids as biomarkers. *Int J Epidemiol* **26**, Suppl. 1, S137–S151.

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

Black A (2000) Critical evaluation of energy intake using the Goldberg cut-off for energy intake: basal metabolic rate. A practical guide to its calculation, use and limitations. *Int J Obes* **24**, 1119–1130.

Black A, Goldberg G, Jebb SA, Livingstone MBE, Cole T & Prentice AM (1991) Critical evaluation of energy intake data using fundamental principles of energy phisiology: 2. Evaluating the results of published surveys. *Eur J Clin Nutr* **45**, 583–599.

Block G (1982) A review of validations of dietary assessment methods. *Am J Epidemiol* **115**, 492–505.

Buzzard IM, Dennis B, Moag-Stahlberg A & Stamler J (1998) Quality control of 24 hour dietary recall interview techniques: use of audiotapes for process evaluation. *Eur J Clin Invest* **52**, Suppl. 2, S14.

Conway R, Robertson C, Dennis B, Stamler J & Elliott P (2004) Standardised coding of dietary records: experiences from INTERMAP UK. *Br J Nutr* **91**, 765–771.

Cook A, Pryer J & Shetty P (2000) The problem of accuracy in dietary surveys. Analysis of the over 65 UK national diet and nutrition survey. *J Epidemiol Community Health* **54**, 611–616.

Dennis B, Stamler J, Buzzard M, *et al.* (2003) INTERMAP: the dietary data – process and quality control. *J Hum Hypertens* **17**, 609–622.

Dolecek T, Stamler J, Caggiula A, Tillotson JL & Buzzard I (1997) Methods of dietary and nutritional assessment and intervention and other methods in the Multiple Risk Factor Intervention Trial. *Am J Clin Nutr* **65**, 196S–201S.

Dwyer J, Picciano MF & Raiten DJ (2003*a*) Collection of food and dietary supplement intake data: What We Eat in America-NHANES. *J Nutr* **133**, 590S–600S.

Dwyer J, Picciano MF & Raiten DJ (2003*b*) Food and dietary supplement databases for What We Eat in America-NHANES. *J Nutr* **133**, 624S–634S.

Friedman GD, Cutter GR, Donahue RP, Hughes GH, Hulley SB, Jacobs JR, Liu K & Savage PJ (1988) Cardia: Study design, recruitment, and some characteristics of the examined subjects. *J Clin Epidemiol* **41**, 1105–1116.

Goldberg G, Black A, Jebb SA, Cole T, Murgatroyd PR, Coward W & Prentice AM (1991) Critical evaluation of energy intake data using fundamental principles of energy physiology: 1. Derivation of cut-off limits to identify under-recording. *Eur J Clin Nutr* **45**, 569–581.

Gorder D, Dolecek T, Coleman G, Tillotson JL, Brown H, Lenz-Litzow K, Bartsch G & Grandits G (1986) Dietary intake in the Multiple Risk Factor Intervention Trial (MRFIT): nutrient and food group changes over 6 years. *J Am Diet Assoc* **86**, 744–751.

Grandits G, Bartsch G & Stamler J (1997) Chapter 4. Method issues in dietary data analyses in the Multiple Risk Factor Intervention Trial. *Am J Clin Nutr* **65**, 211S–227S.

Hajjar I & Kotchen T (2003) Regional variations of blood pressure in the United States are associated with regional variations in dietary intakes: the NHANES-III data. *J Nutr* **133**, 211–214.

Hilner J, McDonald A, Van Horn L, Bragg C, Caan B, Slattery M, Birch R, Smoak C & Wittes J (1992) Quality control of dietary data collection in the CARDIA study. *Control Clin Trials* **13**, 156–169.

Karvetti RL & Knuts LR (1985) Validity of the 24-hour dietary recall. *J Am Diet Assoc* **85**, 1437–1442.

Livingstone MBE, Robson PJ, Black A, Coward W, Wallace JMW, McKinley MC, Strain JJ & McKenna PG (2003) An evaluation of the senstivity and sensitivity of energy expenditure measured by heart rate and the Goldberg cut-off for energy intake: basal metabolic rate for identifying mis-reporting of energy intake by adults and children: a retrospective analysis. *Eur J Clin Nutr* **57**, 455–463.

Macdiarmid J & Blundell J (1998) Assessing dietary intake: who, what and why of under-reporting. *NutrRes Rev* **11**, 231–253.

Morgan KJ, Johnson SR, Riezek RL, Reese R & Stampley GL (1987) Collection of food intake data: an evaluation of methods. *J Am Diet Assoc* **87**, 888–898.

Phillips ELR, Arnett D, Himes JH, McGovern PG, Blackburn H & Luepker RV (2000) Diffrences and trends in antioxidant dietary intake in smokers and non-smokers, 1980–1992: the Minnesota Heart Survey. *Ann Epidemiol* **10**, 417–423.

Poppitt SD, Keogh GF, Prentice AM, Williams DEM, Sonnemans HMW, Valk EEJ, Robinson E & Wareham NJ (2002) Long-term effects of ad libitum low-fat, high-carbohydrate diets on body weight and serum lipids in overweight subjects with metabolic syndrome. *Am J Clin Nutr* **75**, 11–20.

Pryer JA, Vrijheid M, Nichols R, Kiggins M & Elliott P (1997) Who are the 'low energy reporters' in the Dietary and Nutritional Survey of British adults? *Int J Epidemiol* **26**, 146–153.

Schakel SF (2001) Maintaining a nutrient database in a changing marketplace: keeping pace with changing food products – a research perspective. *J Food Comp Anal* **14**, 315–322.

Sievert YA, Schakel SF & Buzzard IM (1989) Maintenance of a nutrient database for clinical trials. *Control Clin Trials* **10**, 416–425.

Slimani N, Ferrari P, Ocké M, *et al.* (2000) Standardisation of the 24-hour diet recall calibration method used in the European Prospective Investigation into Cancer and Nutrition (EPIC): general concepts and preliminary results. *Eur J Clin Nutr* **54**, 900–917.

Smith AF (1991) *Vital and Health Statistics: Cognitive Processes in Long-term Dietary Recall*. Rockville, MD: US Department of Health and Human Services.

Stamler J, Elliott P, Dennis B, Dyer A, Kesteloot H, Liu K, Ueshima H & Zhou B (2003) INTERMAP: background, aims, design, methods, and descriptive statistics (non dietary). *J Hum Hypertens* **17**, 591–608.

Stubbs RJ, O'Reilly L, Fuller Z, Horgan G, Mehar C, Deary I, Austin E, Ritz P, Milne E & James WPT (2004) *N08001: Detecting and Modelling Mis-reporting of Food Intake with Special Reference to Underreporting in the Obese*. London: Food Standards Agency.

Tapsell L, Brenninger V & Barnard J (2000) Applying conversational analysis to foster accurate reporting in the diet history interview. *J Am Diet Assoc* **100**, 818–824.

Thomas B (1994) *Manual of Dietetic Practice*. London: Blackwell Science.

Tillotson JL, Gorder D, DuChene A, Grambsch P & Wenz J (1986) Quality control in the multiple risk factor intervention trial nutrition modality. *Controlled Clini Trials* **7**, 66S–90S.

Todd KS, Hudes M & Howes Calloway D (1983) Food intake measurement: problems and approaches. *Am J Clin Nutr* **37**, 139–146.

van Horn LV, Gernhofer N, Moag-Stahlberg A, Farris R, Hartmuller G, Lasser VI, Stumbo P, Craddick S & Ballew C (1990) Dietary assessment in children using electronic methods: telephones and tape recorders. *J Am Diet Assoc* **90**, 412–416.

Willett W (1998) *Nutritional Epidemiology*. Oxford, UK: Oxford University Press.