

Standardised coding of diet records: experiences from INTERMAP UK†

Rana Conway¹, Claire Robertson^{2*}, Barbara Dennis³, Jeremiah Stamler⁴ and Paul Elliott² for the INTERMAP Research Group

¹Department of Nutrition and Dietetics, Kings College London, 150 Stamford Street, London SE1 8WA, UK

²Department of Epidemiology and Public Health, Imperial College London, St Mary's Campus, Norfolk Place, London W2 1PG, UK

³Department of Biostatistics, Collaborative Studies Coordinating Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

⁴Department of Preventive Medicine, The Feinberg School of Medicine, Northwestern University, Chicago, IL, USA

(Received 16 May 2003 – Revised 18 December 2003 – Accepted 7 January 2004)

Coding diet records is a basic element of most dietary surveys, yet it often receives little attention even though errors in coding can lead to flawed study results. In the INTERnational study of MAcro- and micronutrients and blood Pressure (INTERMAP study), efforts were made to minimise errors in coding the 18 720 diet records. Staff were centrally trained and certified before being able to process study data and ongoing quality control checks were performed. This involved the senior (site) nutritionist re-coding randomly selected diet records. To facilitate standardisation of coding in the UK, a code book was designed; it included information about coding brand items, density and portion size information, and default codes to be assigned when limited information was available for food items. It was found that trainees, despite previous experience in coding elsewhere, made coding errors that resulted in errors in estimates of daily energy and nutrient intakes. As training proceeded, the number of errors decreased. Compilation of the code book was labour-intensive, as information from food manufacturers and retailers had to be collected. Strategies are required to avoid repetition of this effort by other research groups. While the methods used in INTERMAP to reduce coding errors were time consuming, the experiences suggest that such errors are important and that they can be reduced.

Nutrient coding: 24 h Recall: Standardisation: INTERMAP

When collecting dietary intake data, errors arise from several sources (Bingham, 1991). Under- and mis-reporting by certain individuals, e.g. restrained eaters and the overweight, are increasingly recognised and efforts are made to minimise these problems (Black *et al.* 1991; Johansson *et al.* 1998). How diet records are coded to produce estimates of nutrient intake receives far less attention and in most research papers it goes unreported. This could reflect inadequate attention to the coding process that if associated with coding errors could result in flawed results.

Coding is generally carried out using an electronic database, which the coder searches to find a food code to match to each item reported in a diet record. Inaccuracies arise during coding from several sources, e.g. difficulties in interpretation of written detail, mistakes (e.g. number displacement) or poorly chosen food codes (Bingham, 1987). Coding can also be compromised if users of food compositional databases do not refer to the documented

limitations and pitfalls of using such tables. Users of the 6th edition of *McCance and Widdowson's Food Composition Tables* (Food Standards Agency, 2002), for example, are advised to refer to food labels as well as the tables to ensure that recent additions (e.g. fortification) or changes in recipes (e.g. replacement of sunflower oil with olive oil) are accounted for in resultant nutrient calculations. Nutrient labelling is similarly flawed, however. A recent Food Standards Agency report considering the accuracy of nutrient labelling on sixty-five samples of sausages (Food Standards Agency, 2003) found discrepancies of up to 30% when analysed meat contents were compared with labelling declarations. Divergence was also evident with several nutrient estimates: for example, Co-op Butchers 8 Select Pork Sausages contained 1350 mg more Na, 26 g more sugar, 29 g more carbohydrate and 34 g more protein/kg in product analysis compared with nutrient label estimates.

Abbreviations: INTERMAP, INTERnational study of MAcro- and micronutrients and blood Pressure; LSNFR, local site new food request; QC, quality control.

† For copies of the INTERMAP study manuals and INTERMAP UK code book please contact the corresponding author.

* **Corresponding author:** Dr Claire Robertson, fax + 44 20 7402 2150, email c.robertson@imperial.ac.uk

Inaccuracies in nutrient estimations linked to use of static food composition tables are also variable between individual users, and reflect differences in coder interpretation of diet records and nutrient databases used, rather than real differences between individual's nutrient intakes (Guilland *et al.* 1993). As problems during coding can greatly affect estimates of nutrient intake, enhanced efforts are needed to minimise them.

The present paper addresses efforts made in the INTERMAP (INTERnational study of MACro- and micronutrients and blood Pressure) study to maximise accuracy in coding. This study included 4680 participants, from seventeen population samples across China, Japan, the UK and the USA (Stamler *et al.* 2003). Four 24 h dietary recalls were collected from each participant with use of standardised methods (Dennis *et al.* 2003). Dietary data collected from the 2195 participants interviewed in the USA were entered directly onto the computer while the participant was recalling intake to enable automated nutrient coding; the 9940 dietary recalls collected in China, Japan and the UK were coded manually. Standardised procedures were applied to ensure this was done with the highest possible uniformity and accuracy.

Methods

INTERMAP staff in the UK were trained centrally to collect and code the 24 h recalls. Ongoing quality control (QC) checks were included with the aim of achieving a high standard of collection and processing of data throughout. Training and QC systems were adapted from the National Heart, Lung and Blood Institute procedures used in the Lipid Research Clinics study, the Multiple Risk Factor Intervention Trial and the USA–PRC Collaboration (Dennis *et al.* 1980, 2003; National Institutes of Health, 1992). As far as we are aware, attempts to standardise coding in this way have not been previously made in the UK.

Before involvement in INTERMAP fieldwork, trainees were first required to gain certification in dietary collection and/or processing procedures. For dietary coders, this included nutrient coding of five standard 24 h dietary recalls and five additional self-collected recalls. Coding was first completed on the recall forms, then this information was transferred onto the FoodBase computer program (version 1.3, 1993; The Institute of Brain Chemistry and Human Nutrition, University of North London, London, UK). The basic structure and some of the nutrient composition data held within this program was used by the UK INTERMAP research team; however, it was extensively updated to include considerably more information on individual nutrients (for example, amino acid and fatty acid data not contained here) and data on new foods that had recently become available in the UK, and whose nutrient compositions could not be matched to that from any existing food code (e.g. Yakult) (Schakel *et al.* 2003).

Throughout training and fieldwork, a code book was used, designed to facilitate standardisation of coding. For each food code, the code book specified a description, a list of food items represented by this code and portion size information. It was compiled as a rule book, structured

to ensure coders did not make subjective decisions. Each recall coded by a trainee was marked against the original recall designed and coded by a senior nutritionist to assess accuracy. Any difference between food code number and/or weight assigned and original coding was considered a line error. The number of errors in each recall was expressed as a percentage of the total number of lines of coding. If this value exceeded 6%, the diet record failed, and additional diet records were coded until the required ten passed. Only then could coders process data for the study. Since several staff members coded the same diet records in the training effort, this provided an opportunity to look at the impact of coding by different individuals on estimates of an individual's nutrient intake.

QC checks on coding during fieldwork began at the local site (Fig. 1). When coders had collated a batch of ten diet records, a senior local nutritionist (site nutritionist) randomly selected one recall (10%) to re-code blind. If it contained >6% line errors, then all ten in the batch were recoded by the initial coder. This was continued until a randomly selected recall passed with $\leq 6\%$ errors.

Many foods recorded in the diet records were not in the initial INTERMAP code book. During fieldwork, local site new food request (LSNFR) forms were completed to ask how such items should be coded. These forms were resolved whenever possible by the site nutritionist by matching the nutrient composition of the 'new' food, obtained from food tables or manufacturer's labels, to that of an existing food code. The permitted differences between the new item and the proposed match varied according to nutrient (e.g. ± 355 kJ (85 kcal), ± 5 g protein, ± 2.5 g fat, ± 50 mg Ca/100 g food). If a single suitable substitute could not be found, an attempt was made to solve the

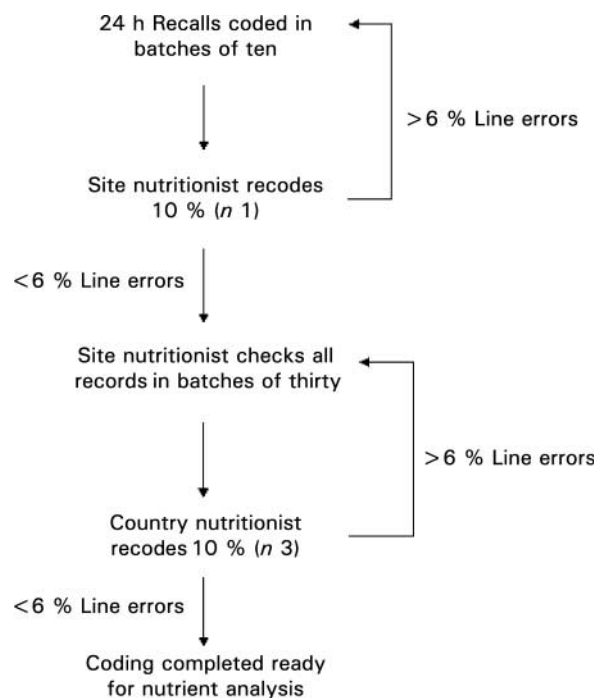


Fig. 1. Flow diagram of quality control for coding INTERMAP 24 h diet records.

problem using a recipe, comprising several codes. When it was not possible to find such a solution, a new food code was created by the Nutrition Coordinating Centre (Minneapolis, MN, USA) and added to the database to enable coding (Schakel *et al.* 1997, 2003). Related coding rules were added to the code book for future use. When all LSNFR had been resolved and entered into the recalls, and batches had passed local QC checks, completed recalls were collected into batches of thirty (see Fig. 1). The site nutritionist reviewed these recalls before passing each batch to the country nutritionist, who initiated a further 10% check (three randomly selected recalls from each batch). These were recoded blind, and passed if 6% errors were found. If this value was exceeded, the local site nutritionist was responsible for recoding the complete batch of thirty recalls.

Results

Training

In the UK, eighteen staff completed training and were certified to code INTERMAP diet records. We explored the variation in estimates of nutrient intake associated with line errors noted in recalls recoded for the first four trainees (all dietary coders in the Belfast centre) during the initial stages of training. The four trainees included a dietitian, a nutritionist and two undergraduate students. Although they had all been taught previously how to code UK diet records, they made several errors. Comparisons of the nutrient estimates calculated for the correctly coded recall and trainee-coded recalls found considerable absolute variation in estimates of nutrient intake (Table 1). Feedback was given following completion of each recall and each error was highlighted to help minimise its future occurrence. As consecutive recalls were coded the number of line errors decreased, as did variation in estimated nutrient intake.

Ongoing quality control checks

As described earlier, during the period of data collection, when a coder had completed coding a batch of ten diet records, one (10%) was re-coded by a site nutritionist. During fieldwork it was found that 25% of batches failed to meet the QC criteria the first time and 31% of the failed batches failed again after return to the initial coder. Some of the eighteen coders were better than others. One coder completed sixteen batches (160 diet records) with no batches failing the QC check. Among the first five batches coded by another person, three batches

had to be coded three times and the other two had to be coded four times. This coder improved greatly as more diet records were coded and the last six batches all passed QC checks first time.

Coding errors

Throughout the study, whenever a senior nutritionist recoded a diet record for QC, she gave written feedback to the coder about line errors. From these reports we were able to categorise different types of errors that arose:

Keying: typing in the wrong code number or weight.

Choice of code: e.g. choosing the code for standard mayonnaise when a low-fat variety was recorded.

Subjective decisions: these were probably the most common mistakes initially, e.g. coding an unknown type of milk as whole milk when the code book stated that if the type is unknown it is to be coded as semi-skimmed.

Calculations: sometimes these errors were simple, e.g. coding for weight of one slice of bread when two were eaten, and sometimes these were more complex (e.g. when the participant reported intake in g, the coder multiplied by density, which is required only when intake is estimated by volume) or making an error in calculating weight (e.g. of a sausage from its estimated length, diameter and density).

Interpretation: e.g. when it was recorded that half a cup of coffee was not drunk, halving only the weight of coffee consumed not the weight of milk and sugar.

Decisions about brand items: e.g. coding what a volunteer described as 'Co-op Weetabix' as 'Weetabix' when Weetabix is a brand; the Co-op cereal should have been treated as a possible new food. In this case it is likely that it would be coded in the same way, but the nutrient label would have to be checked first.

Code book

As we had not used this system before, we started with a blank code book; as foods arose in the diet records, decisions were made about how to code them and these were entered in the code book. The number of LSNFR per diet record varied greatly and often reached as many as sixteen. In the Belfast centre for example, 3501 LSNFR forms were generated from the 240 Belfast participants (data collected from eighteen participants were incomplete, and therefore excluded from the final database). Creating coding rules in response to these requests was very time consuming, as it involved multiple contacts with food

Table 1. Line errors (%) and absolute variation in estimation of daily energy and nutrient intakes resulting from coding errors* (Ranges for four trainees)

24 h Dietary recall number†	Line errors (%)	Energy (kj)	Protein (g)	Na (mg)
1	13–20	– 824–+1681	– 8.1–+4.6	– 153–+424
2	25–44	– 904–+142	– 31.8–+0.6	– 1275–+17
3	4–14	– 96–+38	– 2.2–+0.5	– 511–+62
4	0–7	– 188–0	0–+0.1	0–+15.5

* For details at procedures, see p. 766.

† The four dietary recalls were coded consecutively and feedback regarding errors was given after each one.

Table 2. Extract from the code book showing how new foods were matched to existing food codes

Food description and code number	Foods this code can be used for	Portion information	Portion size (g)	Any comments or coding rules
Biscuit, semi-sweet (FB 07310)	Any brand of Rich tea, Osbourne, Arrowroot	Round Rich tea	7	Iced plain biscuits: 1.2 g (FB 07912), 7 g (FB 07310) per biscuit
		Rich tea finger (If unknown code as round Rich tea)	5	
		Arrowroot	8	
	Marks and Spencer's reduced-fat sultana/sultana and cinnamon cookies	Marks and Spencer's sultana (+/- cinnamon) cookies	16	Café noir: 7 g (FB 07310) 2 g (FB 07912)
		Arrow biscuit	7	
	Tesco's honey nut and raisin biscuit	Marie	8	Chocolate-coated Rich tea: 10 g (FB 07310), 3 g (FB 08104) per biscuit (not low fat) (if plain chocolate FB 8105)
		Tesco's honey nut and raisin	16	
	McVites/Tesco's light, digestive	Light digestive	15	
	Sainsbury's reduced-fat digestives	Sainsbury's reduced fat digestive	10	
	Morning coffee biscuit	Morning coffee	5	

FB, FoodBase code number.

manufacturers and retailers, and frequent visits to shops and supermarkets. The extracts from the code book show how new foods, e.g. Tesco's light digestive biscuits, were added to existing codes (Table 2) and how simple combinations of codes were used for composite products, e.g. beef pie (Table 3). Table 4 shows one of the more complicated recipes in the code book, calculated by the Nutrition Coordinating Centre. Of the large number of UK requests to the Nutrition Coordinating Centre, most resolutions took the form of recipes such as that in Table 4. Some 9735 of these coding rules were added to the UK code book throughout data collection, and 213 new food codes were also added to the food composition database.

Default codes

An important aspect in the development of the code book was writing rules for default codes. These were needed when inadequate information was available to code an item, e.g. as a result of the interviewer not asking for sufficient detail, the participant not being able to remember or not knowing in detail or it not being possible to find missing information. For example, the default code for an unknown type of boiled potatoes was for old potatoes boiled in salted water. Even with the most carefully collected diet records, details are missing, and therefore coders have to make what are at best educated guesses. Wherever possible when deciding on a default code, information was used about common consumption patterns (Ministry of Agriculture, Fisheries and Food, 1996); when this was not available, data on national food production or food imports were used.

Portion sizes

The code book also contained information on portion sizes or their calculation. INTERMAP participants reported

portion sizes in various ways during their 24 h recall interviews. Many items were estimated by volume, with measuring spoons, cups and bowls (with coloured lines for calibration), geometric shapes, food models and food photographs (Nelson *et al.* 1997) as visual aids. If none of the apparatus was felt to be appropriate by the participant, he/she could draw the size of an item on graph paper. To convert volumetric data to weights, we collated information about food densities in the code book (Table 3). For many food items it was possible to use information from the US Department of Agriculture (Database for nationwide surveys, release 7; 1991) as the foods were not country-specific, e.g. steak or honey. For other foods, e.g. certain breads and breakfast cereals, it was necessary to do our own density measurements. This was done by purchasing at least three different items matching

Table 3. Extract from the code book showing use of simple combination of codes.

Beef pie
Density 0.65 g/cm ³ *
Tesco mince and onion pie (shortcrust pastry): 135 g
Tesco mince and onion pie (flaky pastry), from a pack of three: 115 g
Marks and Spencer's Scottish steak pie (shortcrust pastry): 170 g
Sainsbury's steak, mushroom and red wine pie (flaky pastry): 230 g
Iceland mince and onion pie (flaky pasy): 120 g
per 100 g of pie code: 82 g minced beef pie filling, canned (FB 30140)
18 g pastry, flaky, cooked homemade (FB 07644), (made with margarine, average FB 01902)
or 24 g pastry, shortcrust, plain, cooked (FB 07654)
if pastry just top or just bottom, code for 10 g pastry and 90 g beef)

FB, FoodBase code number.

*Density included to enable weight to be estimated from participants description of dimensions.

Table 4. Recipe from code book

Weight (g)	FB	Ingredient
Vegetarian chop suey (Morrisons): one portion		
205	6016	Beansprouts
205	6091	Baby sweetcorn
82.5	6012	Bamboo shoots
30	6106	Water chestnuts
50	5706	Broccoli
60	6072	Red pepper
64	6065	Onion
3	2932	Garlic
3	2933	Ginger
15.6	2768	Oyster sauce
16.4	6309	Tomato puree

FB, FoodBase code number.

a food description, then weighing them and taking measurements. For manufactured items, participants reported standard measures such as a can of cola or a packet of crisps, with a full description of the brand. Participants were also asked about the size of pack, any 'extra free' additions, and proportion of product consumed. We then used the book of *Food Portion Sizes* (Crawley, 1993), or more commonly we used manufacturers' information or purchased the products and weighed them.

Discussion

Coding errors were substantial within INTERMAP UK centres, with about 25% of records failing their first QC checks, yet with increasing knowledge of the code book, accuracy increased. Errors throughout data collection were varied; nonetheless, the use of written personal feedback for each coder following each 10% batch check helped ensure accuracy in nutrient estimates and minimise frequency of coding errors. During fieldwork 9735 new rules were incorporated into the INTERMAP UK coding manual, to match the nutrient contents of a consumed food with an existing code or recipe of codes contained in the FoodBase nutrient analysis package (FoodBase version 1.3, 1993). Some 213 new codes were also added by the Nutrition Coordinating Centre to code those food items that could not be matched to any combination of existing codes.

Staff training and certification was an important component of the INTERMAP study. It was considered essential that coders in each country be trained centrally, as it has been found that coding errors are minimised when coders have similar training (Eagles *et al.* 1966). The results in Table 1 suggest that although staff had coded diet records previously, the way in which they did this varied considerably. The focus of the INTERMAP study was consideration of the relationship of nutrients consumed by individuals to their blood pressure; therefore, it was important that errors in estimated individual intake of this magnitude did not occur during fieldwork. Training methods were found to lead to greater uniformity and reduced individual bias. The variation we found in estimates of individual intakes of energy and nutrients resulting from coding differences

were similar to results found when eleven nutritionists in an US study coded three different diet records (Eagles *et al.* 1966). Whilst nutrient differences were not statistically significant, differences of up to 2397 kJ (573 kcal), 33 g protein and 75 g fat were found. As the eleven nutritionists had similar training and work experience, the authors suggest the variation found might be the minimum to be expected from such an exercise. In a French study thirty diet records were analysed in three different nutrition centres (Guilland *et al.* 1993). Here the focus was on group intakes and significant differences were found in mean intake for nine of fourteen nutrients, including a 29% difference between two of the estimates for PUFA and 36% for alcohol. Coding contributed less to the variability than did differences in the nutrient databases used, but differences from both sources were found. When individuals were ranked according to their nutrient intake estimated by the three French centres, considerable differences arose. For ten of fourteen nutrients, fewer than 70% of individuals were placed in the same tertile by each of the centres and 17% were placed in opposite tertiles by different centres.

Some coding decisions, considered as errors here, may not have been regarded as errors in other surveys, e.g. coding an unknown type of milk as whole rather than semi-skimmed, specified in the code book as the default code. We were aiming at complete standardisation of coding, to ensure that differences across coders had no effect on estimates of nutrient intake. Dietary data collected for the *Dietary and Nutritional Survey of British Adults* (Gregory *et al.* 1990) were processed using a code book designed for their needs; however, this was not felt to be suitable for use in INTERMAP since no definite nutrient limits are set to match codes to consumed items. The code book was designed to facilitate standardisation when foods are described in several ways and can be matched to several codes. For example, a participant may report 'two medium slices of white bloomer'; this could be coded as 'bread, white, average', 'bread, white, sliced', 'bread, white, large crusty' etc. Initially an attempt was made to use a form of matching tree or flow diagram, as might be used to classify plants, etc.; however, it was felt that the effort required would be unlikely to improve nutrient estimates appreciably. It is arguable that some differences in coding should have been allowed rather than being considered as errors, resulting in whole batches of diet records being recoded. However, if decisions about whether a difference constituted an error had been at the discretion of the site nutritionist, there may have been a tendency to permit possibly serious errors so as to avoid additional work for both the site nutritionist and the coder. In addition, making judgemental decisions about passing and failing may have created inter-personal problems, as staff were working in small teams at the local sites. While the benefits of the rigorous QC were appreciated by study staff, a consequence was that coding took much longer than expected and it was not always possible to intersperse coding with more stimulating tasks. The precise impact of the QC procedures on estimates of energy and nutrient intakes is unknown. A comparison of intakes estimated from diet records containing errors with error-free records would partly address the issue; however, it

was not possible to compare large enough numbers of diet records from matched here to allow a valid comparison.

In the UK the number of different foods eaten greatly exceeds the number of codes in the national nutrient database (Food Standards Agency, 2002) and this gap is ever increasing. Sainsbury, for example, estimate that a large store sells approximately 23 000 lines, of which 40% are 'own brand' goods and they launch an average of 3000 new products each year (J Sainsbury plc, 2004). It is inevitable that coders have to make judgement decisions when coding diet records. However skilled they are, the closest match they find for a food may not be good enough and may not have high repeatability, particularly for foods such as ready-meals and reduced-energy versions of standard foods. There are even problems with fairly standard foods like mayonnaise, muesli and sausages, since nutritional contents can vary greatly by brand and many supermarkets also have own-brand varieties. A single code in a nutrient database may be inadequate for these. Another problem we found was that composition of single brands could vary over time. Scotch eggs from a leading supermarket were found to decrease in fat content by 30 g/kg and increase in carbohydrate content by 34 g/kg between 1998 and 2000 for example. Our INTERMAP colleagues in the USA may have been able to account for such changes, as the Nutrition Coordinating Centre maintains routinely updated time-related databases (Buzzard *et al.* 1995), but we were not. Without changes in product names or barcode information, it is unlikely that it will become possible to account for such changes in the future. Since the present study was completed, the 6th edition of *McCance and Widdowson's The Composition of Foods* has been published (Food Standards Agency, 2002). This includes some new foods, yet only a limited number of foods can practically be included in such a database; therefore, even immediately following its publication, this database still houses flaws. This further reinforces the need for a standardised approach to minimise subjective judgements.

New foods needing to be coded separately are constantly coming onto the market. Although some may be available only for a limited time, it is still important to code them as accurately as possible. When it is judged that a new food needs to be added to a nutrient database, information supplied by manufacturers is usually used and fields for unlisted micronutrients are left blank leading to underestimation of micronutrient intakes. Our method of using a combination of codes (as in Table 4) overcame this problem, but the additional lines of code increased the time taken for coding and introduced additional scope for coding errors. In retrospect it may have been better to add a single new food code with values calculated from such a recipe.

We spent a great deal of time collecting information for the UK INTERMAP code book and other research teams are collecting the same information to carry out their own dietary surveys. The INTERMAP UK code book and nutrient database are available for other research groups along with the procedure manuals†. The nutrient database has been enhanced in many ways, not just to account for new foods but to include additional nutrients and to become compatible with the nutrient databases

used in the other INTERMAP countries (Schakel *et al.* 2003). We started with a blank code book, but it may have been preferable to start with a book that included basic foods. This would have reduced the number of forms being completed and processed for the same 'new' foods. We also had to rely on the accuracy of nutrient data from manufacturers; this has been shown to vary considerably from true nutrient content for some foods (International Consumer Research and Testing Ltd, 1998), but chemically analysing foods was not a feasible alternative. Many food manufacturers and supermarkets produce lists of nutrient contents of their foods. A computerised database containing a compilation of such lists would be a useful resource for researchers conducting dietary surveys. Where such lists are not available this can mean either waiting a long time for them to respond to requests for information on individual foods or searching supermarket shelves every time a new product appears on a diet record to obtain the nutrient content from the label. Supermarkets are not obliged to help us in our research and it is not a priority for them. Thus, more comprehensive lists would be of great benefit. Perhaps a generally available code book would be useful to save repetition of effort. Any such code book would have to be country specific; research groups may choose to modify it for their own purposes. Default rules, e.g. for unknown types of cooking fat, would have to be carefully considered, as they may not be appropriate for certain regions or among certain ethnic minority groups. In such instances default rules may be particularly necessary, as coders may not have the same background as participants.

Coding diet records is time consuming and attempts to standardise the process, such as those employed for INTERMAP, inevitably increase the time required. The improvements that occurred as our trainees coded more diet records and the reduction in the number of recalls failing QC checks as coders processed further batches show that accuracy in coding can be substantially enhanced. Two components of such progress were increasing familiarity with the code book and adoption of the idea that coders should not be making judgement decisions. Ongoing QC checks meant additional work as batches of diet records were coded repeatedly; this was necessary only because of unacceptably high error rates, which further highlights the need for such checks. While problems with information in food tables and from food manufacturers can be beyond the scope of many dietary surveys, it is feasible to attempt to standardise the coding process. Since our experiences show that coding errors are problems, but remediable ones, other studies need to consider including strategies in their protocols to minimise such errors. Standardisation becomes increasingly necessary with the ever-increasing number of foods available and with increasing emphasis on high quality, large, standardised, often multi-centre, studies.

Acknowledgements

It is a pleasure to thank all INTERMAP staff members involved in processing dietary data in the Belfast centre,

and the researchers in the USA and London who were involved in formatting and updating the FoodBase nutrient analysis package used in the UK.

References

- Bingham SA (1987) The dietary assessment of individuals: methods, accuracy, new techniques and recommendations. *Nutr Abstr Rev* **A57**, 705–742.
- Bingham SA (1991) Limitations of the various methods of collecting dietary intake data. *Ann Nutr Metab* **35**, 117–127.
- Black AE, Goldberg GR, Jebb SA, Livingstone MB, Cole TJ & Prentice AM (1991) Critical evaluation of energy intake data using fundamental principles of energy physiology: 2. Evaluating the results of published surveys. *Eur J Clin Nutr* **45**, 583–599.
- Buzzard IM, Schakel SF & Ditter-Johnson J (1995) Quality control in the use of food and nutrient databases for epidemiologic studies. In *Quality and Accessibility of Food Related Data*, pp. 241–252 [Greenfield H, editor]. Arlington VA: AOAC International.
- Crawley H (1993) *Food Portion Sizes*. London: H. M. Stationery Office.
- Dennis B, Ernst N, Hjortland J & Grambsch V (1980) The NHLBI nutrition data system. *J Am Diet Assoc* **77**, 641–647.
- Dennis B, Stamler J, Buzzard M, Conway R, Elliott P, Moag-Stahlberg A, Okayama A, Okuda N, Robertson C & Robinson F (2003) INTERMAP: the dietary data – process and quality control. *J Hum Hypertens* **17**, 609–622.
- Eagles JA, Whiting MG & Olsen RE (1966) Dietary appraisal: problems in processing dietary data. *Am J Clin Nutr* **19**, 1–9.
- Food Standards Agency (2002) *McCance and Widdowson's The Composition of Foods*, 6th summary ed. Cambridge: The Royal Society of Chemistry.
- Food Standards Agency (2003) Programme of Mini-surveys: Sausages Survey. Food Surveillance Information Sheets. <http://www.foodstandards.gov.uk/science/surveillance/fsis-2003/sausagesurveyfsis4103> (accessed 30 September 2003).
- Gregory J, Foster K, Tyler H & Wiseman M (1990) *The Dietary and Nutritional Survey of British Adults*. London: H. M. Stationery Office.
- Guillaud JC, Aubert R, Lhuissier M, Peres G, Montagnon B, Fuchs F, Merlet N & Astorg PO (1993) Computerized analysis of food records: role of coding and food composition database. *Eur J Clin Nutr* **47**, 445–453.
- International Consumer Research and Testing Ltd (1998) *Parallel Food Testing: Fortified Foods*. London: International Consumer Research and Testing Ltd.
- Johansson L, Solvoll K, Bjerneboe GE & Dreven CA (1998) Under- and over-reporting of energy intake related to weight status and lifestyle in a nationwide sample. *Am J Clin Nutr* **68**, 266–274.
- Ministry of Agriculture Fisheries and Food (1996) *National Food Survey 1995*. London: H. M. Stationery Office.
- National Institutes of Health (1992) *Data Preview, Baseline Survey, Sub-sample (1983–86) Part 2, Nutrition, Fall/Spring*. Washington, DC: US Department of Health and Human Services.
- Nelson M, Atkinson M & Meyer J, on behalf of the Nutritional Epidemiology Group UK (1997) *A Photographic Atlas of Food Portion Sizes*. London: MAFF.
- J Sainsbury plc (2004) Frequently Asked Questions. http://www.j-sainsbury.co.uk/media/press_questions7.htm
- Schakel SF, Buzzard IM & Gebhardt SE (1997) Procedures for estimating nutrient values for food composition databases. *J Food Compos Anal* **10**, 102–114.
- Schakel SF, Dennis BH, Wold CA, Conway R, Zhao L, Okunda N, Okayama A, Moag-Stahlberg A, Robertson C & van Heel N (2003) Enhancing data on nutrient composition of foods eaten by participants in the INTERMAP study in China, Japan, the United Kingdom, and the United States. *J Food Compos Anal* **16**, 395–408.
- Stamler J, Elliott P, Dennis B, Dyer A, Kesteloot H, Liu K, Ueshima H & Zhou BF (2003) INTERMAP: background, aims, design, methods, and descriptive statistics (nondietary). *J Hum Hypertens* **17**, 591–608.