

Development, validation and utilisation of food-frequency questionnaires – a review

Janet Cade^{1,*}, Rachel Thompson², Victoria Burley¹ and Daniel Warm²

¹Nutrition Epidemiology Group, Division of Public Health, Nuffield Institute for Health, 71–75 Clarendon Road, University of Leeds, Leeds LS2 9PL, UK; ²Institute of Human Nutrition, Southampton General Hospital, University of Southampton, Southampton SO16 6YD, UK

Submitted 13 August 2001: Accepted 6 November 2001

Abstract

Objective: The purpose of this review is to provide guidance on the development, validation and use of food-frequency questionnaires (FFQs) for different study designs. It does not include any recommendations about the most appropriate method for dietary assessment (e.g. food-frequency questionnaire versus weighed record).

Methods: A comprehensive search of electronic databases was carried out for publications from 1980 to 1999. Findings from the review were then commented upon and added to by a group of international experts.

Results: Recommendations have been developed to aid in the design, validation and use of FFQs. Specific details of each of these areas are discussed in the text.

Conclusions: FFQs are being used in a variety of ways and different study designs. There is no gold standard for directly assessing the validity of FFQs. Nevertheless, the outcome of this review should help those wishing to develop or adapt an FFQ to validate it for its intended use.

Keywords
Food-frequency questionnaire
Review
Validation
Method development

Definition of a food-frequency questionnaire

For the purposes of this document, the following definition of a food-frequency questionnaire (FFQ) was used¹:

A questionnaire in which the respondent is presented with a list of foods and is required to say how often each is eaten in broad terms such as *x* times per day/per week/per month, etc. Foods chosen are usually chosen for the specific purposes of a study and may not assess total diet.

Although food-frequency questionnaires may frequently form part of a total dietary assessment technique, as for example in the dietary history method, we have not included these types of FFQ usage in this consensus document.

Although this document assumes that a food-frequency questionnaire is appropriate for use in a particular study, it is important to be aware of the strengths and limitations of the method. No dietary method can measure dietary intake without error. Thus it is important that sources of error are taken into account².

Review methodology

A review was undertaken of all dietary studies conducted or published since 1980 in which the development,

validation or use of a food-frequency questionnaire was described. In order to identify relevant studies that described the design, evaluation and/or use of FFQs, a comprehensive search procedure was developed. Electronic databases including Medline, Embase, Cancerlit, CAB Abstracts and Dissertation Abstracts and the online Dietary Assessment Calibration/Validation Register (<http://www-dacv.ims.nih.gov/>) were searched from 1980 to September 1999. Hand searches of published conference proceedings, key nutrition journals and reference lists of retrieved articles were also undertaken.

Search terms used were based on the MESH terms and keyword searches for 'food frequency questionnaires', 'reproducibility', 'validity/validation', 'diet-study-techniques' and 'calibration'. Questionnaires that assessed only vitamin/mineral supplement intakes, alcohol or contaminants (such as heavy metals) were excluded, as were articles written in languages other than English. For the purposes of the search, we defined a food-frequency questionnaire as 'any list of one or more *foods* with frequency of intake categories'. However, since the focus of the review was the design, validation and utilisation of food-frequency questionnaires, we excluded papers in which the results of the FFQ were combined with another dietary assessment technique, as for example in the dietary history method. All references were downloaded into the computerised bibliographic program Reference Manager,

*Corresponding author: Email j.e.cade@leeds.ac.uk

which facilitated handling of the large numbers of publications obtained.

Owing to the large number of references relating to utilisation of FFQs, only references published in 1998 were included in that part of the review. A one-year sample was thought to be adequate to generate data on the way FFQs are currently being used.

A form was set up using Microsoft Access that facilitated data entry from each study. Papers were broadly divided into one of three study types: validation, reproducibility or utilisation. Data from these papers were then compiled into an Access database for analysis.

Following extraction of the data, initial results were presented at a meeting of the Nutrition Epidemiology Group, a group of UK national experts in the field, which was held at King's College in London in December 1999. The group members agreed to comment on the text of the consensus document that was in preparation. The document was also subsequently sent to a number of international experts (see list of contributors in Acknowledgements). Following receipt of these comments the review was amended taking the whole body of evidence into account. These personal experiences were particularly useful to add detail to some of the design issues discussed in the review, since this is often not included in published accounts of research using FFQs. The full document resulting from this process can be found on the web at <http://www.leeds.ac.uk/nuffield/pubs/ffq.pdf>.

Development of food-frequency questionnaires

Questionnaires may either be developed from basic principles or adapted from existing questionnaires. Of the 227 validation studies in the review, 54% used a modified version of an existing questionnaire. Of these, 25% (26/104) were adapted from a questionnaire originally devised by Block *et al.* (the NCI/Block Health Habits and History Questionnaire)³ and 27% (28/104) were adapted from one devised by Willett and colleagues (the Harvard Semiquantitative Food Frequency Questionnaire)⁴. Examples of validation studies that have used the Block questionnaire include Refs. 5–29 and those that have used the Willett questionnaire include Refs. 6, 19, 20 and 30–53. There are a number of publications that discuss the relative merits of each of these FFQs^{6,42,54–58}.

Purpose of the food-frequency questionnaire

Before selecting or designing a food-frequency questionnaire, careful consideration should be paid to its purpose. The purpose of the food-frequency questionnaire was omitted or not clearly stated in many papers. Fifty-two per cent (115/223) of food-frequency questionnaires were designed to assess foods or food groups and 74% (166/223) to assess nutrient intakes.

The group of experts felt that there were some situations when the use of food-frequency questionnaires was not

advisable. These were: in studies with small numbers of subjects; for surveillance and monitoring of current levels where *accurate absolute intakes* are required; using an FFQ developed for one country in another country unless dietary habits are very similar; and in some clinical work when precise intakes are required. There was less agreement on the suitability of food-frequency questionnaires for assessment of past diet^{59–65} and individual absolute intakes^{66,67}.

Modifying an existing questionnaire

Where time and finances are limited, the use of a pre-existing questionnaire may be particularly appealing. Although modification of an existing FFQ is a simpler and faster method than developing a questionnaire from scratch, a few points need to be considered before embarking on this approach.

1. What was the original purpose of the questionnaire?
2. Who was the target population?
3. When was the questionnaire developed?
4. Has a previous validation been carried out, and was it acceptable?

Careful scrutiny of any publications relating to the development and/or validation studies of the questionnaire is required to determine whether the original objectives of the questionnaire meet the requirements of the new study. The questionnaire may have been developed a number of years ago and thus may not cover all commonly eaten foods today: it may be literally 'out-of-date'. Furthermore, adapting the analysis program may also take more time than is generally appreciated, depending on how the program has been written.

Correlation coefficients from papers in the review were compared for newly designed food-frequency questionnaires and those adapted from other questionnaires. Newly developed questionnaires had a higher correlation for energy (0.49 vs. 0.44) and fat (0.52 vs. 0.49) than modified questionnaires. Adapted questionnaires had a higher correlation for vitamin C (0.50 vs. 0.44) and vitamin A (0.41 vs. 0.34) than newly developed questionnaires. There were no differences (new vs. adapted) for calcium (0.54 vs. 0.55) and iron (0.45 vs. 0.44). Overall, agreement between reference dietary measures and FFQs did not appear to be worse with adapted FFQs than with FFQs developed *de novo*.

Development from basic principles

Development of the food list is crucial to the success of a food-frequency questionnaire. The full variability of the population's diet, which includes many different foods, brands and preparation practices, cannot be captured fully with a finite food list.

For a food item to either contribute to absolute intake or to differentiate between individuals, it must be eaten reasonably often by an appreciable number of the

population and contain a substantial amount of the nutrient/food group of interest. Also, the use of the food under study must vary from person to person.

If the aim of the study is aetiological, then it is preferable to have a comprehensive food list enabling computation of the full range of nutrients rather than a restricted list to determine the intakes of a few nutrients. The advantages of a comprehensive list include the ability to adjust for energy intake in order to investigate fully diet–disease relationships⁶⁸. To assess energy intake requires a comprehensive food list. Also, if the nutrient of interest is highly correlated with other nutrients, unless the whole diet is assessed it may not be possible to explore this. Data obtained may have long-term use and it is difficult to predict the dietary factors of future interest, particularly if a number of research groups have potential access to the data and opportunities to re-survey the study population are limited. For example, a recent analysis was carried out of information collected on 4-year-olds in 1950 that are still being followed-up as adults⁶⁹. In order to discuss the bioavailability of a nutrient, it may be important to gather information about intakes of other nutrients that may interact with the nutrient of interest.

There are, however, certain circumstances in which the purpose of the food-frequency questionnaire may be very specific and a comprehensive food list may be unnecessary or even prohibitive. For example, as a tool for the identification of high-fat consumers for enrolment in an intervention trial^{70,71}. A number of short FFQs have been successfully developed along these lines, with a view to the assessment of intake of calcium and other nutrients thought to be related to bone health^{16,72–78}.

How to choose appropriate foods?

Previous dietary survey information collected on the appropriate population can be used to identify commonly eaten foods and recipe dishes to be included on a questionnaire^{3,79–81}. Recent data are required because of the problem of new foods. This is particularly an issue with children's diets.

Stepwise regression analysis on a dietary dataset from an appropriate population can be used to identify foods that most discriminate between individuals. There are available computerised dietary analysis packages or general statistical programs that can carry out this step^{82,83}.

Foods may also be included in the questionnaire on the basis of prior information, epidemiological or otherwise, that an association might exist (e.g. calcium and bone health^{84–86}).

The food list can be piloted in a sample of the population of interest. This may be particularly useful when working with groups whose dietary habits are not well documented. If no previous recent dietary surveys can be located, then information can be obtained from a sample of the target population using 24-hour recalls, diet histories or participant observations.

Use of individual foods versus groups of foods

Obtaining accurate reports for foods eaten both alone and in mixed dishes is particularly problematic (e.g. vegetables as a whole portion or in a mixed dish). Food-frequency questionnaires may ask the respondent to report either a combined frequency for a particular food eaten both alone and in mixed dishes (e.g. beef). Alternatively, they may ask the respondent to report separate frequencies for foods eaten alone or in combination (e.g. separate questions on different meat dishes containing beef, such as roast beef, beef casserole, chilli con carne, etc). The first approach is cognitively complex and difficult for people not involved in cooking, but the second approach may lead to double counting and overestimation of intake. In addition, when a group of foods is covered as a single question, assumptions about the relative frequencies of intakes and portion sizes of the foods must be made when calculating gram weights or nutrient intakes.

Grouping of items has been shown to lead to an underestimation of intake⁸⁷. It may be better to ask separate questions, although it has also been shown that increasing the number of items can lead to an overestimation of intake⁸⁸ (there are methods available to adjust for this – see section on Cross-check questions).

The consensus from the group of experts was that single items are better than grouped foods at least for some items in a questionnaire. The advantages being that single items can differentiate between similar foods, e.g. full-fat versions and low-fat versions of the same food. It is possible to aggregate single items but not to separate grouped items. Grouped items can complicate the question and lengthen the time and effort of completion. The food list cannot be endless and for practical reasons some grouping is often necessary. It is important that the key characteristics for grouping of foods are based on *a priori* hypothesis. This will primarily depend on the purpose of the questionnaire.

Number of food items

The number of food items listed in a food-frequency questionnaire tends to vary widely. The review found that the number of food items on a questionnaire ranged from 5 to 350. The median number was 79. Questionnaires with a more specific remit, such as an FFQ on foods rich in vitamin A, or fruit and vegetable consumption, may often be shorter than questionnaires intended to assess the whole diet. The length of the questionnaire may partly be determined by the characteristics of the target population and the number of other questionnaires the subject may need to complete.

The number of foods included should be considered along with the validity and reproducibility of the questionnaire and level of accuracy of dietary data required. If only a crude assessment of dietary data is required, it may be that a short food-frequency questionnaire is sufficient if this measures dietary intake

to the required accuracy⁸⁹. Others have shortened longer questionnaires and re-validated the resultant FFQ^{16,17,90,91}.

Willett⁶⁷, citing a study conducted by Pietinen *et al.*^{92,93} in which a 44-item FFQ was compared with a more detailed 273-item one, suggests that there is a rapidly decreasing marginal gain in information obtained with increasingly detailed questionnaires. There would therefore appear to be little to gain in unnecessarily elongating the number of food items included when developing a food-frequency questionnaire.

Assembling a list of selected foods

Once the single and grouped items have been selected it is important to consider the order of the foods in the questionnaire. To facilitate dietary reporting, food groupings should fit within respondents' conceptual framework. Related items should be clustered together, such as traditional food groups.

For closely related foods, more specific items should be placed before general items (e.g. low-calorie salad dressing before other salad dressing). Focus groups can be a useful strategy to help construct lists for culturally specific questionnaires or to provide information about which foods should be grouped together.

Food groups of particular interest should be placed near the beginning of the questionnaire but not at the start. Errors may be made in the responses to the first few questions while the participant is getting used to the format of the questionnaire. Additionally, towards the end of the questionnaire, the accuracy of responses may decline due to boredom or fatigue. Therefore it is best to start with something simple and unambiguous and place the more important items shortly after this.

Some research groups have experimented with the use of open-ended questions to record specific types or brands of foods, such as margarine or breakfast cereals. However, the limited research that has been conducted in this area suggests that there is little or no improvement in validity when subjects are allowed to specify the brand of breakfast cereal, cooking oil or multivitamin consumed compared with being asked to select from a more limited food list⁶⁵.

Frequency and portion size

Once the food list is compiled, the next step is to obtain some measure of the frequency with which each item is consumed and possibly also some indication of the amount eaten. Questions on frequency and portion size should be closed rather than open. This reduces coding time and transcription errors, and reduces the number of questionnaires that have to be rejected because responses are incomplete or cannot be adequately interpreted. If it is necessary to use open questions it is best to use well-trained interviewers so that they can ensure the questions are completed adequately.

Frequency categories should always be continuous,

with no gaps, as the sensitivity of the questionnaire will be reduced and respondents will be frustrated if they cannot find their response. The number of choices should range between 1 and 12 but will mainly depend on the intended use of the questionnaire.

The range of frequency choices should reflect the time frame of interest. The frequency categories should emphasise the more frequent end of the distribution for most foods (e.g. number of times per week). However, for foods that are eaten infrequently but make a significant contribution to nutrient intake (e.g. liver), it is important to include a less frequent option, say less than once a month. A few foods are consumed more than once a day. If there are options of more than once a day this tends to lead to gross overestimates for some people. The review found that a variety of different frequency options were used. Some used ascending whilst others used descending frequency choices. Some concentrated on foods eaten on a weekly basis, ignoring foods eaten less than once a week, whilst others were also interested in foods eaten more than once a day.

Seasonally consumed items can be problematic when reporting frequency as they may be consumed very frequently when in season and then not at all out of season. A separate section can be included that asks about consumption of seasonal items when 'in season'. The data can then be adjusted at analysis to reflect length of time in season.

Inclusion of portion size is necessary if gram weights or nutrient intakes are required. We found from the review that 22% of food-frequency questionnaires did not present portion size information, 42% specified a portion size and 36% allowed participants to describe their own portion size. Research has shown that individuals have difficulty in estimating portion sizes of foods, both when examining displayed foods and when reporting about foods previously consumed⁹⁴⁻⁹⁶. If an individual or the researcher cannot assign portion size, absolute nutrient intake cannot be calculated.

When there are no questions in the questionnaire on portion size, gram weights and nutrient intakes can be calculated using existing data on average portion sizes appropriate for the population being studied³. Gender-specific portion sizes have also been used⁹⁶.

A portion size may be specified on the questionnaire and the participants can select a frequency category according to how often they consume the specified portion size. If the frequency question is combined with a specified portion size (e.g. the Harvard Semiquantitative Food Frequency Questionnaire⁴), this presents cognitive challenges for subjects that should be addressed. This is particularly important when a subject does not consume the food item in the amounts specified. Under these circumstances, it is unclear whether the subject will just ignore the portion size or will select a different frequency category to allow for the difference in portion.

A third option is to include an additional recording option for each food to describe the usual portion. This can be achieved by asking the respondent to describe their serving as small, medium or large (where the medium portion is specified). Alternatively, photographs or food models can be used for respondents to select their own portion size. However, even within populations the use of small, medium and large as a description for portion size may not have an accepted meaning. Between populations, the value allocated to small, medium or large may be very different.

Portion sizes should reflect known consumption patterns in the population, and the questionnaire should allow for a sufficient range of expression of portion size to enable subjects with the same frequency of consumption but different portion sizes to be adequately distinguished. Use of 'standard' portions applied equally to all subjects simplifies the questionnaire but will reduce sensitivity if portion sizes vary within the population⁸⁹.

The choice of whether to include an assessment of portion size will depend on a number of factors. These include the availability of average portion size data; the variability of portion sizes in the population (if there is little variation in portion sizes then assessment may not be necessary, especially if absolute intakes are not required); and ability of the population to accurately assess portion size and level of accuracy required. Where there is little information on the usual portion sizes of a population appropriate serving sizes can be determined by work with focus groups.

The review found that measures of agreement between FFQs and a reference dietary measure were highest when subjects were able to describe their own portion size (correlation coefficients 0.5–0.6) compared with no portion size specified (use of average portion weights to compute intakes: correlation coefficients 0.2–0.5) or portion size specified on the questionnaire (correlation coefficients 0.4–0.5). In general, there was little difference between no portion size specified and whether a portion size was specified. In terms of repeatability, in general, correlation coefficients were higher when subjects were allowed to specify their own portion sizes.

Although not totally consistent across studies, the results from the review showed that some estimation of portion size rather than using average portion weights appeared advantageous. It may be, however, that the minor improvement in validity obtained when allowing subjects to specify their own portion sizes does not justify the extra cost and time involved in development⁹⁷. The issue of whether to assess portion size and the best method of doing it are still matters for discussion and further research. The group of experts was divided on their views of the usefulness of portion size estimation. On the whole there seemed to be agreement that estimation by subjects of their portion sizes was useful. However, it was acknowledged that this was not easy to do and more work

was necessary in this area. Many of the experts advocated using photographs to estimate portion size. Practical guidelines on the design and analysis of studies to validate portion size estimates and on the development and use of photographic atlases for assessing food portion size have been published⁹⁸.

Method of administration

Questionnaires may be either interviewer- or self-administered according to the needs of the study. Self-administered questionnaires require more careful preparation and pre-testing.

A useful way of overcoming limited interviewer resources is to design a questionnaire that is self-administered, but to include in the study protocol an opportunity for the responses to be reviewed and any queries clarified in a face-to-face or telephone interview⁹⁹. Computer-readable forms are useful as they can be scanned into the computer, hence eliminating data-entry errors and reducing time. One problem with self-administered food-frequency questionnaires is incomplete answers; some respondents will only complete the questionnaire for items they usually eat. Another common problem is that complete pages may be missed. A solution is to check the questionnaire for completeness soon after it is returned so that incomplete answers can be kept to a minimum. In their recent paper, Caan *et al.*¹⁰⁰ provided a list of questions for use as a food frequency review probe. These authors found that using a nutritionist to probe for correct responses on a self-administered FFQ improved agreement with a food record used as a reference method.

An alternative to the use of face-to-face interviews is to administer the FFQ by telephone. The advantages of telephone interviewing have been reviewed by Fox *et al.*¹⁰¹ and include higher response rates than postal surveys and the potential to reach large numbers of people in widely scattered geographic areas. Interview by telephone can be substantially less expensive than face-to-face interviews, but cost comparisons vary with the research setting. Posting picture booklets or other portion size estimation aids to the participants before the telephone interview can simplify the reporting of portion sizes by telephone¹⁰². However, there are further cost implications and the booklets also need to be validated.

The review found that 67% of questionnaires validated were self-administered. Correlation coefficients (interviewer vs. self-administered) between FFQs and reference measures were higher for interviewer-administered questionnaires than for self-administered questionnaires for fat (0.55 vs. 0.50), energy (0.55 vs. 0.46) and vitamin A (0.47 vs. 0.37), were similar for calcium (0.56 vs. 0.55) and slightly higher for self-administered questionnaires for vitamin C (0.45 vs. 0.49). Correlation coefficients for repeatability between interviewer-administered and self-administered FFQ were better for interviewer-administered

for fat (0.65 vs. 0.60), energy (0.67 vs. 0.63) and vitamin A (0.59 vs. 0.58), but worse for vitamin C (0.59 vs. 0.66).

Computation of food and nutrient intakes

In order to convert frequency estimates of food intake to nutrient values, an appropriate nutrient database needs to be constructed. Ritenbaugh *et al.*¹⁰³ demonstrated that the choice of nutrient database can have an impact on the strength of association between a biomarker reference method and an FFQ designed to assess carotenoids. The limitations of food tables/databases also need to be taken into consideration, particularly the extent to which missing values interfere with the aspects of diet that are to be assessed, and if and how the limitations can be addressed¹⁰⁴.

If the nutrient content of a food is not known, samples should be collected and the food analysed chemically. Alternatively, the nutrient content of mixed dishes can be estimated from recipes that include foods for which the nutrient composition is known. All recipe ingredients need to be weighed or measured. Data on weight losses associated with cooking (e.g. due to water evaporation) should be recorded to ensure accurate nutrient density of the portion size consumed. If individual foods are grouped on a questionnaire, then a composite nutrient value for the group needs to be established. This will depend on the relative frequency and portion sizes of the individual foods.

A database of portion sizes will also need to be compiled. Sources of portion size data used for estimation can be either published values, data from surveys using weighed records or estimates of specific portion sizes, ideally by the population group in question^{94–96,105–109}.

Total nutrient intake can be calculated from the sum of the products of the frequency weight and nutrient content of the portion of food. Frequency weights can be assigned to assess weekly or daily consumption (e.g. for daily/once a day = 1; four times a week, $4/7 = 0.57$). In more complex questionnaires, the nutrient content of some foods may be modified by responses to other questions, e.g. margarine by type of margarine. The edible portion of a food should also be taken into consideration in order to provide the nutrient values for weight as eaten (for example, removal of fat or bone weights from meat servings).

Missing data on food-frequency questionnaires can be treated in a number of ways. Firstly, questionnaires with a large percentage of incomplete questions should be excluded. This value needs to be decided *a priori* and will depend on the purpose and level of accuracy required. For questionnaires not exceeding the limit for incomplete data, a value of zero (food not eaten) may be used; alternatively an average value for the population could be substituted.

Additional information

Clear instructions should be given at the beginning of the questionnaire if it is to be self-administered. These are usually enhanced by the use of relevant examples.

Additional questions on the treatment of fat on meat can be used to adjust fat intake. Intuitively it might seem that this type of qualitative information would improve the validity of total fat or fatty acid estimates, but there is little evidence, in practice, to suggest that it does¹¹⁰. These types of additional question on methods of food preparation and cooking can be placed at the end of the food frequency section.

Some foods are less easy to assess via FFQ than others due to the pattern of intake (e.g. milk). Milk may be consumed frequently in small amounts (e.g. in drinks) and also less often in large amounts (on cereals, glasses of milk). It is therefore useful to ask specific questions about milk in the additional information section. Other examples that can be included in a separate cross-check section are questions on bread, sugar and alcohol. Additional questions could also be asked about key sources of the nutrients of interest to improve the accuracy of the data.

Collection of data on supplement use is potentially important and should be considered at the design stage of the FFQ^{8,111}. This is a complex area and precise details of the products consumed are required in order to assess nutrients from supplements. Furthermore, setting up a nutrient database for dietary supplements is a costly and time-consuming process due to the expanding and highly changeable market in these products.

Cross-check questions can be used to correct for misreporting of certain food groups. These are often used for fruits and vegetables as these tend to be overreported, particularly if each fruit or vegetable is listed singly in a long list. A cross-check question can be employed by asking the number of servings consumed per week of fruits and vegetables. A weighting factor may then be used to correct for any overreporting. A separate weighting can be applied for each subject, but this does assume that all items are misreported to the same extent¹¹². The newly estimated amount of foods is then used to estimate both food and nutrient intakes.

However, the cross-check method may also lead to an underestimation of intake. For example, people may not consider fruit juice when asked about portions of fruit. Although the inclusion of cross-check questions has been used successfully as a strategy to identify possible overreporting of fruit and vegetable intakes¹¹², they may not be as effective when used to assess other foods. Wolk *et al.* found that there was a negligible increase in the validity of fat estimates due to use of cross-check questions about fat¹¹⁰. If cross-check questions are used to modify the data, details of the methodology and adjusted and unadjusted food and nutrient estimates should be presented.

Pre-testing

To ensure that an FFQ is acceptable and understood by the population in which it is to be used, it is important to pre-test the questionnaire in the field. Use of cognitive interviewing techniques can help to pinpoint problems in design and comprehension of the questionnaire¹¹³.

It is important to stress that, before proceeding with an FFQ, it is recommended that the procedure for data entry, whether manual or optical, is tested. The analysis program should also be tested to ensure that there are no mistakes.

Reproducibility of food-frequency questionnaires

To determine whether a food-frequency questionnaire provides reproducible results is important for all types of study design. ‘Reproducibility’ can also be thought of as ‘reliability’. The reproducibility of FFQs has generally been assessed by administering them at two points in time to the same group of people and correlation coefficients (or some other test of association) used to assess the association between the two responses^{92,93,114–117}.

When the food-frequency questionnaire is administered by an interviewer, two aspects of reliability should be distinguished: inter-rater reliability and intra-rater reliability. Inter-rater reliability assesses whether different interviewers use the questionnaire similarly and achieve similar answers from the same subjects. Intra-rater reliability assesses whether repeat administration by the same interviewer yields the same answers, in the same way as reproducibility is assessed for self-administered questionnaires. The statistical methods used are the same for both aspects of reproducibility.

Repeatability was assessed in only 47% of validation studies in the review. It is not wise to administer a questionnaire at a very short interval as respondents may remember their previous responses. Alternatively, when a longer interval is used, true changes in dietary habit as well as variation in response contribute to reduced reproducibility¹¹⁸.

Statistical issues

Correlation

The most common method, used in 90% of studies, for assessing reproducibility was the correlation coefficient. This method has recently been shown to be flawed because it does not measure agreement between two administrations of the questionnaire, only the degree to which the two administrations are related. Since we use the same questionnaire on the same people, we would expect them to be closely related, but this is not the same as agreement^{119–123}. Other problems include the fact that the strength of the correlation is dependent on the range of values in the population (which itself can be partly influenced by size of the sample) and the characteristics of the subjects in the particular sample used. However, due

to the widespread use of correlation for assessing reproducibility, it may be helpful to use it in conjunction with another more appropriate method^{124–128}.

Where correlation is used, Pearson correlation coefficients should be used on normally distributed data and Spearman rank correlation coefficients should be used where data are not normally distributed. From the review, correlation coefficients between the two administrations of 0.5 to 0.7 were common. Correlations were somewhat higher for repeat administrations 1 month or less apart compared with those administered 6 months to 1 year apart. The time interval between repeat administrations of the food-frequency questionnaires in the review ranged from 2 hours to 15 years. In 34% the repeat administration was between 1 and 6 months later. In 31% it was between 6 and 12 months.

Within-person error can also be corrected for by estimating the correlation coefficient based on an average of a large number of replicates for each individual, though this is not commonly done¹²⁹.

Bland–Altman method

Preferable to the use of correlation coefficients is the Bland–Altman method, which assesses the agreement between the methods across the range of intakes¹²². This method was used in less than 10% of studies in the review. It can determine if there is any systematic difference between the administrations of the questionnaire (bias), and to what extent the two administrations agree (limits of agreement). It also provides a method of assessing whether the difference between the methods is the same across the range of intakes, and whether the extent of agreement differs for low intakes compared with high intakes. These may be assessed by plotting the difference between the methods against the average of the two administrations. The overall mean difference indicates if one method tends to over- or underestimate and the limits of agreement (mean difference \pm 2 standard deviations (SD)) show how well the administrations agree.

Other methods

The Kappa statistic can be used to compare categories of food intakes such as frequencies of consumption measured by two methods. Kappa statistics are not appropriate for continuous measures, unless the intention is to subsequently categorise the measure into a number of ordered groups.

Validation of food-frequency questionnaires

Validation of the FFQ method is essential, as incorrect information may lead to false associations between dietary factors and diseases or disease-related markers. For a more detailed discussion of issues important to the correct validation procedure in dietary studies, refer to publications by Burema *et al.*¹³⁰ and Nelson¹²⁹.

Validation studies may be carried out to assess whether the questionnaire is measuring what it should measure or to assess the degree to which the questionnaire agrees with a 'gold standard' or other methods of measuring diet. Alternatively, they may be undertaken to assess the level of measurement error associated with use of the food-frequency questionnaire (to allow adjustment of the results of the main epidemiological study for measurement error).

As even subtle changes in the design of food-frequency questionnaires may affect their performance, each new instrument should be validated separately, even if it is largely based on a previous questionnaire. Questionnaires may also perform differently in different demographic groups and cultures.

Sample/population selection in validation studies

In order to validate an FFQ, it needs to be tested on a sub-sample of the main study population. Age, ethnic group, gender and health status of the population can affect the outcome of the validation study¹²⁹. It is most important, therefore, that the target population should be similar to the main study population. The source of subjects for the validation study should always be stated, and their characteristics comprehensively described. These subject characteristics will affect the way that they respond to the task of completing an FFQ or undertake some other possibly more demanding method of assessing dietary intake. It has also been observed that the type of diet consumed can have an impact on the outcome of the validation study. McPherson and colleagues¹³¹ obtained a high agreement between an FFQ and food records for estimates of energy, fats and cholesterol which they attributed partly to the lack of dietary diversity in their subject sample.

Subjects who volunteer to take part in validation sub-studies are self-selected, and may therefore respond differently to an FFQ than non-volunteers. Self-selected study participants tend to provide more accurate responses to questionnaires, and they may also have different dietary habits⁹.

Sample size for validation studies

Sample size required will depend on the statistical method being used to assess reproducibility and validity. The review showed a wide range in sample sizes from 6 to 3750, with a median of 110. Expert statistical advice should be sought when deciding on the number of subjects to include in a validation study. The same issues discussed here also apply to sample size for repeatability studies.

For the Bland–Altman method, the sample size should be large enough to allow the limits of agreement to be estimated precisely. Thus a sample size of at least 50, and preferably much larger (100 or more subjects, say), is desirable. It is also valuable to take two measurements on each subject by each method to improve precision and

so that repeatability and validity can be assessed simultaneously¹³².

For the correlation coefficient, the sample size will depend on the expected association between the two measures or methods. Based on the correlation coefficient, assuming a sufficient number of days of dietary information are obtained to reasonably describe an individual's diet (typically 14 to 28 days), a sample size of no more than 100 to 200 should be sufficient (as illustrated by Refs. 65, 92, 93 and 133). Understandably, however, few studies manage to achieve such a large number of days of good quality dietary information from their subjects and therefore most use between two and five replicates (days) per subject (as illustrated by Refs. 134–136). If a strategy using a small number of replicates per subject is employed, the number of subjects needs to be increased to maintain the same precision of the corrected correlation coefficient. The sample size used will inevitably depend on resources.

Sequence of administration

Ideally, the test instrument should be administered prior to the assessment of the reference measure. Subjects would normally, in the course of the main investigation in which the test measure was to be used, encounter it independent of any other dietary assessment, and the validation process should mimic this. Secondly, completing the assessment using the reference measure may in itself draw respondents' attention to their diets.

Time frame of reference method

In order to validate the FFQ, the time frame of the reference method in relation to that of the food-frequency questionnaire needs to be taken into consideration. In theory, the food-frequency questionnaire and the reference method should assess diet over the same time span (current, past or usual intake). For example, a food-frequency questionnaire that assesses intake over a period of a year could be administered twice, a year apart, and compared with diet records collected at intervals in the intervening time. If the objective is to determine past intake by food-frequency questionnaire, this makes the validation process more difficult.

Reference method selection in validation studies

A vital component of the validation process is the selection of the appropriate reference method against which to assess the test measurement. There are considerable problems involved with measurement of true habitual dietary intake. Dietary assessments aimed at determining current intake are likely to interfere with the subject's everyday habits and cause a distortion of intake, and methods aimed at the assessment of past intake are reliant upon the memory and conceptualisation skills of the subject. Although there are now good biological measures for energy¹³⁷, nitrogen¹³⁸ and sodium intake¹³⁹, there is no

'ideal' method for the measurement of dietary intake as a whole. In conducting a validation study, food-frequency questionnaire measures are compared with an alternative, but not necessarily more accurate, method of assessing diet. Such a validation study can therefore only indicate whether the methods give related answers or not. If there is disagreement between methods the test cannot identify which method is correct or even whether it accurately assesses absolute or even relative intake.

The systematic review showed that 75% of studies validated an FFQ against another dietary method and 19% against a biomarker. Twelve per cent of studies were reports of an FFQ validation against another method, e.g. doubly labelled water, energy expenditure studies or interviews.

Dietary methods used in FFQ validation studies

In theory, the measurement errors of the food-frequency questionnaire and reference method should be independent. Possible dietary methods of choice are weighed or household records or 24-hour recalls. Weighed records, since portions are weighed, have the least correlated errors with food-frequency questionnaires. As the errors are largely independent, if anything, validity tends to be understated. If the food-frequency questionnaire results are compared with weighed records, the lack of agreement can be attributed in part to the within-subject variance that is inherent in the shorter but more accurate reference measure. It should not be assumed that the FFQ estimates true usual intake without the equivalent of random measurement error (the within-subject variance of the weighed records).

Weighed records or diet records should be the first method of choice for validating food-frequency questionnaires. Although 24-hour recalls are less demanding for the participant than diet recording and less likely to influence the actual diet of the subjects, their sources of error tend to be more correlated with the error in a dietary questionnaire (e.g. reliance upon memory, conceptualisation of portion sizes and distortion of reported diet). However, when co-operation or literacy of study subjects is limited, 24-hour recalls may be more appropriate (as illustrated in Ref. 135).

In theory, when used as a reference method, diet records or 24-hour recalls should be kept for a sufficient number of days to represent average intake and cover the interval of time corresponding to the questionnaire (typically one year). For example, four days of dietary information collected four times a year (four days for each season) to compare with a food-frequency questionnaire assessing intake over one year¹⁴⁰. There is some evidence that increasing the number of days of recording in the reference method improves the apparent validity of a questionnaire¹⁴¹. It would appear that efforts to increase the duration of recording in the reference method provide a better measure of habitual intake, which is generally more similar to the type of information generated by an FFQ.

In practice it may be better to collect a sufficient number of 'independent' replicate 24-hour recalls to allow estimation of the variance components and then use this information to statistically adjust the comparison of FFQ and reference method. Although such an approach has serious flaws (including the necessity to accept a pooled variance estimate as if it applies correctly to each individual), it goes a long way towards eliminating the impact of the random measurement error in the 24-hour recall and exposing the error term of the FFQ. Using data from two studies, Stram *et al.*¹⁴² presented calculations to determine the 'ideal' number of days of dietary recording to use in a validation study. They concluded that, in most settings, the optimal study design will rarely require more than four or five diet records per subject.

One error common to both test and reference methods is the use of national food composition tables (e.g. Ref. 143). For a full discussion of the construction, errors and use of food composition tables in epidemiological research, see West and van Staveren¹⁴⁴.

Any dietary assessment methodology is prone to a degree of mis- or underreporting. Weighed records and 24-hour recalls are not without errors; therefore it may be useful to assess their completeness. Levels of underreporting were 31% in the Second National Health and Nutrition Examination Survey (NHANES II)¹⁴⁵ and 46% for women and 29% for men in the National Diet and Nutrition Survey of British Adults¹⁴⁶. The use of the Schofield equations to predict minimum energy intakes could be employed¹⁴⁷ to eliminate participants with unfeasibly low energy intakes.

The review showed that a variety of different dietary assessment tools were used as a reference measure. Fifty-six (25%) used the weighed record; 59 (26%) used a food record/diary (not including weighed diaries); 50 (22%) used the 24-hour recall; 14 (6%) used the diet history questionnaire; and 27 (12%) used another food-frequency questionnaire. One hundred and forty-four (64%) validation studies used only one reference method (14 another FFQ, 43 weighed record, 29 24 hour-recall, 58 either a food record or diet history questionnaire). Seven (3%) validation studies used both the weighed record and 24-hour recall as reference methods.

There was little difference in correlation coefficient between the different reference measures for energy ($r = 0.47$), fat ($r = 0.51$), vitamin A ($r = 0.39$) and calcium ($r = 0.54$). For iron and vitamin C a higher correlation coefficient was found using the weighed record (iron, $r = 0.51$; vitamin C, $r = 0.50$) compared with the non-weighed record (0.41, 0.46) or 24-hour recall (0.43, 0.41, respectively).

Isotope and biochemical techniques

In recent years, there has been an increase in the use of biochemical measurements (biomarkers) of nutrients in blood or other tissues both as a general determinant of

nutritional status and also to provide a comparison with other dietary reference methods. Although biomarkers can provide an estimate of dietary intake that is independent of the subject's reported dietary intake (and therefore less prone to errors involved with underreporting or poor memory), they are often expensive, invasive and nutrient-specific, so may only be used to validate one nutrient at a time.

In general, there is a need to establish how tissue levels equate to consumption. Certain biomarkers, for example urinary nitrogen, relate directly to nitrogen intake. In others (e.g. vitamin C) the relationship is much more complex. In terms of validation studies, there is a need to be clear about just what the biomarker measures. Many, if not most, biomarkers do not permit an assessment of true absolute intake.

Biochemical reference standards are subject to three sources of error:

1. the difference between the dietary assessment and the true intake;
2. the effects of digestion, absorption, uptake, utilisation, metabolism, excretion and homeostatic mechanisms, all of which bear on the relationship between the amount ingested and the biochemical measurement; and
3. the error associated with the biochemical assay itself.

It is clear, therefore, that the biochemical marker and dietary assessment method do not measure the same thing. The errors for biochemical measures are independent of errors associated with food-frequency questionnaires.

When correlation coefficients for studies using a dietary reference method were compared with those of studies using biomarkers, little difference was observed for energy, fat, vitamin C or vitamin A. Correlations were in the region of 0.50 for all of these nutrients except vitamin A, where the correlation coefficient for FFQ vs. dietary method was 0.40 and that for FFQ vs. biomarker was 0.35.

The repeatability of the biomarker should also be evaluated. Diurnal variation estimates are available for some of the markers and should be taken into account unless blood-sampling time is standardised or incorporated as a variable in analyses.

One further factor to be taken into account when using a biomarker within a validation study is how biological variation in the biomarker relates to variation in intake. For example, it is known that, with an appropriate lag period, urinary urea tracks protein intake. Comparing protein intake estimated from daily urinary urea samples against an FFQ estimate of habitual protein intake may not be appropriate. On the other hand, with a marker that can be expected to show a wide seasonal variation, for example serum carotenoids, it is essential that the biomarker information is collected on days that are representative of the total frame of the FFQ.

Statistical issues

An absolute bias has a very limited impact in many epidemiological studies but is devastating in any attempt to assess apparent nutrient adequacy. Random measurement error has serious repercussions in epidemiological studies. Bias that is consistent within an individual but random between individuals can be misleading.

There are several statistical approaches to validation, and often several reference methods to validate the food-frequency questionnaire against. Using more than one approach demonstrates the robustness of the validation process.

Correlation, regression and the Bland–Altman method

The same arguments apply to statistical assessment of validity as to reproducibility. However, correlation and regression can be useful in helping to assess validity, because investigation of the association between different methods can be informative. Correlation coefficients were by far the most common statistical method and were used in 168 (83%) of validation studies in the review.

Regression can be used to calibrate one method compared with another. Regression analysis was undertaken by eight (4%) of the studies. Where correlation or regression is used, this should be alongside the Bland–Altman analysis and not as a replacement^{119–123}. These methods apply to continuous data; however, with ordered categorical data, Kappa should be used. Where Kappa statistics are not practical, Spearman's correlation may be used instead as the best tool available. Sensitivity, specificity, etc. may also be useful for binary data.

The important aspects of validity will vary depending on the purpose of the food-frequency questionnaire. It is not possible to produce recommendations on an ideal mean difference, limits of agreement, correlation or regression slope, as these will depend on the study objectives. However, for lower correlations, say below 0.3 or 0.4, attenuation will be so severe that it will be difficult to detect associations.

Comparison of group means

Where differences between subject groups are required or when absolute intakes are important, the validation study should assess the ability of the test measure to reflect the group mean¹²⁹. This may be achieved by using paired *t*-tests (on normally distributed data), which is the equivalent of testing the overall bias in the Bland–Altman method. For food data, distributions are less likely to be parametric and non-parametric tests may be more appropriate (such as the Wilcoxon signed rank sum test).

Statistical tests used depend on the variation in the data. For example, the unpaired *t*-test depends on the standard deviation of the differences and hence on the width of the limits of agreement from the Bland–Altman analysis. If these limits are wide, then a substantial overall bias (big difference between the two methods of assessing diet)

may well be non-significant, and therefore overlooked. *P*-values should therefore be used with caution, and a general assessment of the magnitude of possible differences made.

Classification into categories of consumption

For both the test and reference methods subjects may be divided into categories relating to the distribution of dietary intake (e.g. fifths of intake). A comparison of the subjects' categories shows whether subjects were classified in the same or different categories by the two methods. The results permit an assessment of the proportion of subjects who are classified correctly. This method gives a much clearer and undistorted picture of how well the instrument is doing compared with correlation coefficients. Data are usually divided into three or five categories. Results can be reported as an exact agreement (classified in the same category by both methods), ± 1 category and gross misclassification. Agreement can be assessed using the Kappa statistic or sensitivity/specificity can be calculated for dichotomised data.

Other statistical methods for validation studies

Other modelling-based approaches to validation have been developed. These include using components of variance¹⁴⁸ to calculate the intra-class correlation coefficient¹⁴⁹. An alternative approach is the method of triads¹⁵⁰. Thirdly, in recent years there has been increased interest in the use of more advanced modelling, e.g. using structural equation models, for dietary validation studies (for illustration see Refs. 151–154). These can be seen as a generalisation of the approaches based on components of variance and the method of triads, incorporating them in one framework. More complicated models are possible which may, in principle, resolve the problem of correlated random errors between FFQ and the reference method¹⁵¹. The statistical methods underlying these approaches are complex and expert statistical advice should be sought if these methods are to be used.

Utilisation of food-frequency questionnaires

Food-frequency questionnaires have been used in a large number of different studies. The review located 164 studies, published in 1998, in which food-frequency questionnaires were used to assess dietary intakes (not validation studies). Of the studies considered, 60% reported data on nutrient intakes and 46% reported data on food/food group consumption. Sixty-one per cent of studies used self-administered questionnaires and 58% reported an associated validation paper.

The FFQs were used in randomised control trials (2%), cohort studies (20%), case–control studies (26%) and cross-sectional studies (51%). The aim of the FFQ was to assess general dietary information only (foods, food

patterns or nutrient intakes) in 32 (20%); diet–disease relationships in 82 (50%) studies; and dietary intakes with biochemical or physical measures in 31 (19%).

There are a number of points to consider when reporting the use of an FFQ in a paper. The key aspects to include are summarised in the recommendations arising from this review. Examples of good practice in terms of description of an FFQ can be found in Refs. 155–158 and for reporting previous validation in Refs. 159–162. If as a result of a validation study the dietary results are adjusted for measurement error, then details of the method of adjustment must be given. The consensus from the group of experts was that they have reservations about adjusting for measurement error especially if the adjustments generate large changes in the dietary estimates.

Issues specific to different study designs

Food-frequency questionnaires have been designed and used in a wide range of situations, and types of dietary study. In this section, FFQ design and validation issues specific to each type of study design are discussed.

Cross-sectional surveys

Cross-sectional studies investigate relationships at a single point in time and, as such, are unable to generate information on causality. However, they have been used to provide group comparisons, ranking of individuals and an assessment of usual dietary intake^{163–166}. If the questionnaire aims to look at the percentage failing to meet nutritional requirements then issues of sensitivity and specificity also need to be addressed.

Brief questionnaires designed to measure specific dietary behaviours (e.g. fruit and vegetable consumption) may be useful in lifestyle type surveys in which the number of dietary questions needs to be limited^{167–169}. If the cross-sectional study aims to compare different subgroups of the population, for example the effect of age group or gender, then the food-frequency questionnaire should be validated for each of the important subgroups.

Case–control (retrospective) studies

Unlike cross-sectional studies, case–control studies have been used to provide support for a causal link between diet and disease (as illustrated by Refs. 170–172). Food-frequency questionnaires are a popular tool in this type of study, although the need to obtain dietary information retrospectively, i.e. before the onset of the disease, raises a number of design and validation concerns.

When designing the FFQ for use in a retrospective case–control study, the food list used should reflect dietary consumption at the relevant time point. The effect of memory is important and largely relates to the omission of foods. The number of foods recalled tends to be correlated with total intake of energy and nutrients, thus

differential misclassification will occur between those with good and poor memories.

Evidence exists that people whose dietary habits are relatively stable are more likely to be able to successfully recall past diet. Additionally, greater total diet reproducibility has been found among men with higher education, among women of less than 110% desirable weight reporting no special diet and among women reporting no medications⁶².

The presence of disease in cases may interfere with the ability to complete a questionnaire and it may be that an interview-based design would be required. Recall or memory of past diet may also vary between cases and controls due to the effects of the disease process itself or drugs used in treatment. An understanding of the typical development time of the disease may also be required in order to set the appropriate time frame of reference for the FFQ.

Validation is a particular issue in retrospective studies. It is not easy to validate a questionnaire inquiring about eating habits in the past, and the experts were divided about the ability of FFQs to function in this way. Some research groups have reported useable recall of past diet with FFQs^{173–175}, whereas others have found that the past diet correlates as well with the current diet as with the recalled diet^{173,176,177}.

The questionnaire used should ideally be validated for use with both cases and controls as the questionnaire may be handled differently in each group⁶⁴. Some research groups have looked at the reliability of response in hospital controls, where there is a danger that recall of past diet may be confused with diet consumed while in hospital. D'Avanzo *et al.*¹⁷⁸ found satisfactory comparability of dietary information from subjects interviewed at home with that provided during their original interview in the hospital, and a good reproducibility of information collected in the two settings.

Cohort (prospective) studies

In general, the sample size for a cohort study will be considerably greater than for a case–control study and, due to their ease of completion and analysis, FFQs have been used extensively in this type of study (as illustrated by Refs. 179–185).

In terms of FFQ development, a number of issues are specific to their use in cohort studies. As the duration of cohort studies generally stretches over a number of years or even decades, there may be a need to repeatedly assess diet in the cohort. However, the questionnaire may ultimately become somewhat 'out-of-date' as new foods become available over the duration of the study and dietary patterns change. If the FFQ is to be repeated during the study it may need to be adjusted to include these new items.

Since the dietary component implicated in the development of the disease may not be known at the start of the

study, and new issues may develop over time, it may be better to comprehensively measure the whole diet at the onset of the study. The FFQ should be designed to allow this. The number of times diet is measured in a cohort study will partly depend on resources and whether dietary changes are anticipated. If dietary habits do change over time and different versions of the questionnaire are used, it is important to assess whether the differences are real or a result of different questionnaires. The food-frequency questionnaire may also be used to cluster participants in terms of dietary patterns rather than just nutrient intake^{186,187}.

Validation will be undertaken at baseline, but may also be assessed at follow-ups to ensure the level of validity has not changed. However, under these circumstances it is difficult to tell if validity has changed or if there has been a change in dietary pattern, due for example to the changes in the food market⁶⁵.

Intervention studies

In an intervention study, a food-frequency questionnaire may be used to track changes in diet as a response to some form of intervention (e.g. education). As such, it must be sensitive enough to detect sometimes quite subtle dietary changes. However, food-frequency questionnaires may not be the most appropriate method to use in intervention studies, as they may not be specific enough to detect changes in diet. More importantly, the subjects may report what they consider to be the desirable responses – this would be more difficult to maintain if reporting diet prospectively over a number of days. If an intervention is trying to improve the diet as a whole, intermediate behavioural targets (such as trimming fat from meat, substituting fruit for pastry snacks, etc.) should be measured directly by including additional questions on the food-frequency questionnaire.

Dietary screening in clinical settings

The main objective of questionnaires being used in a clinical setting may be to discriminate between high and low consumers of certain foods or nutrients. Time and cost are usually constraints under these circumstances and questionnaires with a long food list may not be practical. Potentially more useful are shorter questionnaires that include foods/food groups that discriminate between high and low intakes and that are suitable for administration by staff without specialised nutrition training, as for example with the DINE questionnaire devised by Roe and colleagues⁷¹. However, the questionnaires will need to be both sensitive and specific in identifying 'at risk' patients¹⁸⁸.

In clinical settings, such questionnaires have been used to screen for low-fat diets¹¹, to assess diet in children with diabetes¹⁸⁹ or those at risk of iron-deficiency anaemia¹⁹⁰, to assess dietary behaviour in the workplace⁷⁰, and for use in practice nurse dietary assessments¹⁹¹. Food-frequency

questionnaires have also been used as a screening tool to determine study eligibility, for example, as used by Ritenbaugh and colleagues¹⁹² to exclude high fibre consumers for a cancer prevention trial.

Issues of determining absolute intakes may not apply when used as a general screening tool. If the questionnaire is to be used to identify patients who require dietary advice, for example to identify patients with high fat intakes or low fruit and vegetable intakes, then the sensitivity and specificity characteristics of the instrument are more important than absolute intakes. For example, it is important to correctly classify those with high-fat diets so that they can subsequently receive the appropriate advice. Alternatively, the questionnaire must be specific so that those with the correct amount of fat in the diet are not classified as having a high-fat diet and given inappropriate dietary advice.

Conclusions

Many studies have been devoted to the methods of measuring an individual's usual dietary intake. Currently, food-frequency questionnaires are being used in a variety of ways and different study designs. They are most commonly used to obtain estimates of an individual's food intake in relation to the development of various diseases.

This review was prepared to guide the individual about to embark on the development and/or use of a food-frequency questionnaire as a dietary assessment tool. Since the development of a new FFQ is costly, both in terms of time and resources, the issues considered to be of key importance have been summarised in the recommendations. Similarly, the adoption of a pre-existing FFQ poses particular problems according to its ultimate function, and these are also highlighted in this document.

It is well recognised that there is no gold standard for directly assessing the validity of FFQs. However, consideration has been given to the methods available and the overall design of validation studies, and this may provide guidance for those wishing to conduct a validation study on either a new or pre-existing FFQ.

Lastly, the review also provides a breakdown of the ways in which FFQs are currently being used either clinically or in research. It is hoped that these data may guide the individual who is seeking advice about the design and or/validation issues surrounding the use of FFQs under these different circumstances.

Acknowledgements

This project was funded by the Ministry of Agriculture, Fisheries and Food (Project AN0850).

Thanks to all those scientists who contributed their views and comments. In particular: George Beaton, Gladys Block, Tim Byers, Imogen Cowin, Pauline Emmett, Darren Greenwood, Allan Hackett, Rudolf Kaaks, Sara

Kirk, Christel Larsson, Jenny Matthew, Jane Pryer, Sian Robinson, Chris Sempas, Margaret Thorogood, Ailsa Welch and Walter Willett.

References

- 1 Margetts BM, Nelson M. *Design Concepts in Nutrition Epidemiology*. Oxford: Oxford University Press, 1997.
- 2 International Life Sciences Institute (ILSI). *Present Knowledge in Nutrition*. Washington, DC: ILSI Press, 1996.
- 3 Block G, Hartman AM, Dresser CM, Carroll MD, Gannon J, Gardner L. A data-based approach to diet questionnaire design and testing. *Am. J. Epidemiol.* 1986; **124**: 453–69.
- 4 Willett WC, Reynolds RD, Cottrell-Hoehner S, Sampson L, Browne ML. Validation of a semi-quantitative food frequency questionnaire: comparison with a 1-year diet record. *J. Am. Diet. Assoc.* 1987; **87**: 43–7.
- 5 Kristal AR, Feng Z, Coates RJ, Oberman A, George V. Associations of race/ethnicity, education, and dietary intervention with the validity and reliability of a food frequency questionnaire: the Women's Health Trial Feasibility Study in Minority Populations [published erratum appears in *Am. J. Epidemiol.* 1998; **148**(8): 820] [see comments]. *Am. J. Epidemiol.* 1997; **146**: 856–69.
- 6 Wirfalt AK, Jeffery RW, Elmer PJ. Comparison of food frequency questionnaires: the reduced Block and Willett questionnaires differ in ranking on nutrient intakes [see comments]. *Am. J. Epidemiol.* 1998; **148**: 1148–56.
- 7 Lemaitre RN, King IB, Patterson RE, Psaty BM, Kestin M, Heckbert SR. Assessment of trans-fatty acid intake with a food frequency questionnaire and validation with adipose tissue levels of trans-fatty acids. *Am. J. Epidemiol.* 1998; **148**: 1085–93.
- 8 Patterson RE, Kristal AR, Levy L, McLerran D, White E. Validity of methods used to assess vitamin and mineral supplement use. *Am. J. Epidemiol.* 1998; **148**: 643–9.
- 9 Riboli E, Toniolo P, Kaaks R, Shore RE, Casagrande C, Pasternack BS. Reproducibility of a food frequency questionnaire used in the New York University Women's Health Study: effect of self-selection by study subjects. *Eur. J. Clin. Nutr.* 1997; **51**: 437–42.
- 10 Shannon J, Kristal AR, Curry SJ, Beresford SA. Application of a behavioral approach to measuring dietary change: the fat- and fiber-related diet behavior questionnaire. *Cancer Epidemiol. Biomark. Prev.* 1997; **6**: 355–61.
- 11 Martin LJ, Lockwood GA, Kristal AR, Kriukov V, Greenberg C, Shatuck AL, *et al.* Assessment of a food frequency questionnaire as a screening tool for low fat intakes. *Control. Clin. Trials* 1997; **18**: 241–50.
- 12 Marshall JR, Lanza E, Bloch A, Caan B, Caggiula A, Quandt S, *et al.* Indexes of food and nutrient intakes as predictors of serum concentrations of nutrients: the problem of inadequate discriminant validity. The Polyp Prevention Trial Study Group. *Am. J. Clin. Nutr.* 1997; **65**: 1269S–74S.
- 13 Baranowski T, Smith M, Baranowski J, Wang DT, Doyle C, Lin LS, *et al.* Low validity of a seven-item fruit and vegetable food frequency questionnaire among third-grade students. *J. Am. Diet. Assoc.* 1997; **97**: 66–8.
- 14 Hartman AM, Block G, Chan W, Williams J, McAdams M, Banks WL Jr, *et al.* Reproducibility of a self-administered diet history questionnaire administered three times over three different seasons. *Nutr. Cancer* 1996; **25**: 305–15.
- 15 Bittoni MA, Wilkins JR III. Assessment of the reliability of a diet history questionnaire. *Nutr. Cancer* 1994; **21**: 143–55.
- 16 Brown JL, Griebler R. Reliability of a short and long version of the Block food frequency form for assessing changes in calcium intake. *J. Am. Diet. Assoc.* 1993; **93**: 784–9.
- 17 Block G, Hartman AM, Naughton D. A reduced dietary

- questionnaire: development and validation. *Epidemiology* 1990; **1**: 58–64.
- 18 Tucker KL, Bianchi LA, Maras J, Bermudez OI. Adaptation of a food frequency questionnaire to assess diets of Puerto Rican and non-Hispanic adults. *Am. J. Epidemiol.* 1998; **148**: 507–18.
 - 19 Eck LH, Klesges LM, Klesges RC. Precision and estimated accuracy of two short-term food frequency questionnaires compared with recalls and records. *J. Clin. Epidemiol.* 1996; **49**: 1195–1200.
 - 20 Sawaya AL, Tucker K, Tsay R, Willett W, Saltzman E, Dallal GE, *et al.* Evaluation of four methods for determining energy intake in young and older women: comparison with doubly labeled water measurements of total energy expenditure [see comments]. *Am. J. Clin. Nutr.* 1996; **63**: 491–9.
 - 21 Tylavsky FA, Sharp GB. Misclassification of nutrient and energy intake from use of closed-ended questions in epidemiologic research. *Am. J. Epidemiol.* 1995; **142**: 342–52.
 - 22 Coates RJ, Serdula MK, Byers T, Mokdad A, Jewell S, Leonard SB, *et al.* A brief, telephone-administered food frequency questionnaire can be useful for surveillance of dietary fat intakes. *J. Nutr.* 1995; **125**: 1473–83.
 - 23 Coates RJ, Eley JW, Block G, Gunter EW, Sowell AL, Grossman C, *et al.* An evaluation of a food frequency questionnaire for assessing dietary intake of specific carotenoids and vitamin E among low-income black women. *Am. J. Epidemiol.* 1991; **134**: 658–71.
 - 24 Krall EA, Dwyer JT. Validity of a food frequency questionnaire and a food diary in a short-term recall situation. *J. Am. Diet. Assoc.* 1987; **87**: 1374–7.
 - 25 Godley PA, Campbell MK, Miller C, Gallagher P, Martinson FE, Mohler JL, *et al.* Correlation between biomarkers of omega-3 fatty acid consumption and questionnaire data in African American and Caucasian United States males with and without prostatic carcinoma. *Cancer Epidemiol. Biomark. Prev.* 1996; **5**: 115–9.
 - 26 Wilkins JR III, Bunn JY. Comparing dietary recall data for mothers and children obtained on two occasions in a case-control study of environmental factors and childhood brain tumours. *Int. J. Epidemiol.* 1997; **26**: 953–63.
 - 27 Mayer-Davis EJ, Vitolins MZ, Carmichael SL, Hemphill S, Tsaroucha G, Rushing J, *et al.* Validity and reproducibility of a food frequency interview in a multi-cultural epidemiologic study. *Ann. Epidemiol.* 1999; **9**: 314–24.
 - 28 Kuriniji N, Gensler G, Milton R. Development and validation of a food frequency questionnaire in a randomised trial of eye diseases [abstract]. *Eur. J. Clin. Nutr.* 1998; **52**(Suppl. 2): S40.
 - 29 Potischman N, Carroll R, Iturra S. Comparison of the 60- and 100-item NCI-Block questionnaires with validation data. *Eur. J. Clin. Nutr.* 1998; **52**: S63.
 - 30 Green TJ, Allen OB, O'Connor DL. A three-day weighed food record and a semiquantitative food-frequency questionnaire are valid measures for assessing the folate and vitamin B-12 intakes of women aged 16 to 19 years. *J. Nutr.* 1998; **128**: 1665–71.
 - 31 Smith W, Mitchell P, Reay EM, Webb K, Harvey PW. Validity and reproducibility of a self-administered food frequency questionnaire in older people. *Aust. NZ J. Public Health* 1998; **22**: 456–63.
 - 32 Rockett HR, Breitenbach M, Frazier AL, Witschi J, Wolf AM, Field AE, *et al.* Validation of a youth/adolescent food frequency questionnaire. *Prev. Med.* 1997; **26**: 808–16.
 - 33 Smith-Warner SA, Elmer PJ, Fosdick L, Tharp TM, Randall B. Reliability and comparability of three dietary assessment methods for estimating fruit and vegetable intakes. *Epidemiology* 1997; **8**: 196–201.
 - 34 Cooper GS, Busby MG, Fairchild AP. Measurement of lactose consumption reliability and comparison of two methods. *Ann. Epidemiol.* 1995; **5**: 473–7.
 - 35 Kaskoun MC, Johnson RK, Goran MI. Comparison of energy intake by semiquantitative food-frequency questionnaire with total energy expenditure by the doubly labeled water method in young children. *Am. J. Clin. Nutr.* 1994; **60**: 43–7.
 - 36 Basch CE, Shea S, Zybert P. The reproducibility of data from a food frequency questionnaire among low-income Latina mothers and their children. *Am. J. Public Health* 1994; **84**: 861–4.
 - 37 Ajani UA, Willett WC, Seddon JM. Reproducibility of a food frequency questionnaire for use in ocular research. Eye Disease Case-Control Study Group. *Invest. Ophthalmol. Vis. Sci.* 1994; **35**: 2725–33.
 - 38 Byers T, Trieber F, Gunter E, Coates R, Sowell A, Leonard S, *et al.* The accuracy of parental reports of their children's intake of fruits and vegetables: validation of a food frequency questionnaire with serum levels of carotenoids and vitamins C, A, and E. *Epidemiology* 1993; **4**: 350–5.
 - 39 Stein AD, Shea S, Basch CE, Contento IR, Zybert P. Consistency of the Willett semiquantitative food frequency questionnaire and 24-hour dietary recalls in estimating nutrient intakes of preschool children. *Am. J. Epidemiol.* 1992; **135**: 667–77.
 - 40 Eck LH, Klesges RC, Hanson CL, Slawson D, Portis L, Lavasque ME. Measuring short-term dietary intake: development and testing of a 1-week food frequency questionnaire. *J. Am. Diet. Assoc.* 1991; **91**: 940–5.
 - 41 Tucker KL, Chen H, Vogel S, Wilson PW, Schaefer EJ, Lammi-Keefe CJ. Carotenoid intakes, assessed by dietary questionnaire, are associated with plasma carotenoid concentrations in an elderly population. *J. Nutr.* 1999; **129**: 438–45.
 - 42 Caan BJ, Slattery ML, Potter J, Quesenberry CP Jr, Coates AO, Schaffer DM. Comparison of the Block and the Willett self-administered semiquantitative food frequency questionnaires with an interviewer-administered dietary history [see comments]. *Am. J. Epidemiol.* 1998; **148**: 1137–47.
 - 43 Radimer KL, Harvey PW. Comparison of self-report of reduced fat and salt foods with sales and supply data. *Eur. J. Clin. Nutr.* 1998; **52**: 380–2.
 - 44 MacIntosh DL, Williams PL, Hunter DJ, Sampson LA, Morris SC, Willett WC, *et al.* Evaluation of a food frequency questionnaire–food composition approach for estimating dietary intake of inorganic arsenic and methylmercury. *Cancer Epidemiol. Biomark. Prev.* 1997; **6**: 1043–50.
 - 45 Bingham SA, Day NE. Using biochemical markers to assess the validity of prospective dietary assessment methods and the effect of energy adjustment. *Am. J. Clin. Nutr.* 1997; **65**: 1130S–7S.
 - 46 Brown JE, Buzzard IM, Jacobs DR Jr, Hannan PJ, Kushi LH, Barosso GM, *et al.* A food frequency questionnaire can detect pregnancy-related changes in diet. *J. Am. Diet. Assoc.* 1996; **96**: 262–66.
 - 47 Ma J, Folsom AR, Shahar E, Eckfeldt JH. Plasma fatty acid composition as an indicator of habitual dietary fat intake in middle-aged adults. The Atherosclerosis Risk in Communities (ARIC) Study Investigators. *Am. J. Clin. Nutr.* 1995; **62**: 564–71.
 - 48 Enger SM, Longnecker MP, Shikany JM, Swenseid ME, Chen MJ, Harper JM, *et al.* Questionnaire assessment of intake of specific carotenoids. *Cancer Epidemiol. Biomark. Prev.* 1995; **4**: 201–5.
 - 49 Forsythe HE, Gage B. Use of a multicultural food-frequency questionnaire with pregnant and lactating women. *Am. J. Clin. Nutr.* 1994; **59**: 203S–6S.
 - 50 Giovannucci E, Colditz GA, Stampfer MJ, Rimm EB, Litin L, Sampson L, *et al.* The assessment of alcohol consumption

- by a simple self-administered questionnaire. *Am. J. Epidemiol.* 1991; **133**: 810–7.
- 51 Stevens J, Metcalf PA, Dennis BH, Tell GS, Shimakawa T, Folsom AR. Reliability of a food frequency questionnaire by ethnicity, gender, age and education. *Nutr. Res.* 1996; **16**: 735–45.
 - 52 Bell AC, Swinburn BA, Amosa H, Scragg R, Sharpe SJ. Measuring the dietary intake of Samoans living in New Zealand: comparison of a food frequency questionnaire and a 7 day diet record. *Asia Pacific J. Clin. Nutr.* 1999; **8**: 149–54.
 - 53 Field AE, Peterson KE, Gortmaker SL, Cheung L, Rockett H, Fox MK, *et al.* Reproducibility and validity of a food frequency questionnaire among fourth to seventh grade inner-city school children: implications of age and day-to-day variation in dietary intake. *Public Health Nutr.* 1999; **2**: 293–300.
 - 54 Block G. Block vs Willett: a debate on the validity of food frequency questionnaires [letter]. *J. Am. Diet. Assoc.* 1994; **94**: 16–9.
 - 55 Block G. Invited commentary: comparison of the Block and the Willett food frequency questionnaires [editorial; comment]. *Am. J. Epidemiol.* 1998; **148**: 1160–1.
 - 56 Hankin JH. Block vs Willett: a debate on the validity of food frequency questionnaires [letter]. *J. Am. Diet. Assoc.* 1994; **94**: 16–9.
 - 57 Longnecker MP, Chen MJ, Caan B. Block vs Willett: a debate on the validity of food frequency questionnaires [letter]. *J. Am. Diet. Assoc.* 1994; **94**: 16–9.
 - 58 Willett WC. Block vs Willett: a debate on the validity of food frequency questionnaires [letter]. *J. Am. Diet. Assoc.* 1994; **94**: 16–9.
 - 59 Fraser GE, Lindsted KD, Knutsen SF, Beeson WL, Bennett H, Shavlik DJ. Validity of dietary recall over 20 years among California Seventh-day Adventists. *Am. J. Epidemiol.* 1998; **148**: 810–8.
 - 60 Lindsted KD, Kuzma JW. Long-term (24-year) recall reliability in cancer cases and controls using a 21-item food frequency questionnaire. *Nutr. Cancer* 1989; **12**: 135–49.
 - 61 Sobell J, Block G, Koslowe P, Tobin J, Andres R. Validation of a retrospective questionnaire assessing diet 10–15 years ago. *Am. J. Epidemiol.* 1989; **130**: 173–87.
 - 62 Thompson FE, Metzner HL, Lamphiear DE, Hawthorne VM. Characteristics of individuals and long term reproducibility of dietary reports: the Tecumseh Diet Methodology Study. *J. Clin. Epidemiol.* 1990; **43**: 1169–78.
 - 63 Tsubono Y, Fukao A, Hisamichi S, Tsugane S. Perceptions of change in diet have limited utility for improving estimates of past food frequency of individuals. *Nutr. Cancer* 1995; **23**: 299–307.
 - 64 Wilkens LR, Hankin JH, Yoshizawa CN, Kolonel LN, Lee J. Comparison of long-term dietary recall between cancer cases and noncases. *Am. J. Epidemiol.* 1992; **136**: 825–35.
 - 65 Willett WC, Sampson L, Browne ML, Stampfer MJ, Rosner B, Hennekens CH, *et al.* The use of a self-administered questionnaire to assess diet four years in the past. *Am. J. Epidemiol.* 1988; **127**: 188–99.
 - 66 Sempos CT. Some limitations of semiquantitative food frequency questionnaires. *Am. J. Epidemiol.* 1992; **135**: 1127–32.
 - 67 Willett WC. *Nutritional Epidemiology*. New York: Oxford University Press, 1998.
 - 68 Margetts BM, Thompson RL, Key T, Duffy S, Nelson M, Bingham S, *et al.* Development of a scoring system to judge the scientific quality of information from case–control and cohort studies of nutrition and disease. *Nutr. Cancer* 1995; **24**: 231–9.
 - 69 Prynne CJ, Paul AA, Price GM, Day KC, Hilder WS, Wadsworth ME. Food and nutrient intake of a national sample of 4-year-old children in 1950: comparison with the 1990s. *Public Health Nutr.* 1999; **2**: 537–47.
 - 70 Glasgow RE, Perry JD, Toobert DJ, Hollis JF. Brief assessments of dietary behavior in field settings. *Addict. Behav.* 1996; **21**: 239–47.
 - 71 Roe L, Strong C, Whiteside C, Neil A, Mant D. Dietary intervention in primary care: validity of the DINE method for diet assessment. *Family Practice* 1994; **11**: 375–81.
 - 72 Wilson P, Horwath C. Validation of a short food frequency questionnaire for assessment of dietary calcium intake in women. *Eur. J. Clin. Nutr.* 1996; **50**: 220–8.
 - 73 Taitano RT, Novotny R, Davis JW, Ross PD, Wasnich RD. Validity of a food frequency questionnaire for estimating calcium intake among Japanese and white women. *J. Am. Diet. Assoc.* 1995; **95**: 804–6.
 - 74 Haines CJ, Chung TK, Leung PC, Leung DH, Wong MY, Lam LL. Dietary calcium intake in postmenopausal Chinese women. *Eur. J. Clin. Nutr.* 1994; **48**: 591–4.
 - 75 Taylor RW, Goulding A. Validation of a short food frequency questionnaire to assess calcium intake in children aged 3 to 6 years. *Eur. J. Clin. Nutr.* 1998; **52**: 464–5.
 - 76 Molgaard C, Sandstrom B, Michaelsen KF. Evaluation of a food frequency questionnaire for assessing of calcium, protein and phosphorus intakes in children and adolescents. *Scand. J. Nutr./Naringsforskning* 1998; **42**: 2–5.
 - 77 Angbratt M, Moller M. Questionnaire about calcium intake: can we trust the answers? *Osteopor. Int.* 1999; **9**: 220–5.
 - 78 Rogalska-Niedzwiedz M, Charzewska J, Wajszczyk B, Lachowitz A, Gorajec M, van Erp-Baart MA. Comparison of food frequency questionnaire and a 3 day record in estimating sources of calcium intake in Polish girls and women. *Eur. J. Clin. Nutr.* 1998; **52**: S56.
 - 79 Sharma S, Cade J, Jackson M, Mbanya JC, Chungong S, Forrester T, *et al.* Development of food frequency questionnaires in three population samples of African origin from Cameroon, Jamaica and Caribbean migrants to the UK. *Eur. J. Clin. Nutr.* 1996; **50**: 479–86.
 - 80 Tsubono Y, Takamori S, Kobayashi M, Takahashi T, Iwase Y, Itoi Y, *et al.* A data-based approach for designing a semiquantitative food frequency questionnaire for a population-based prospective study in Japan. *J. Epidemiol.* 1996; **6**: 45–53.
 - 81 Cade JE, Margetts BM. Nutrient sources in the English diet: quantitative data from three English towns. *Int. J. Epidemiol.* 1988; **17**: 844–8.
 - 82 Brants HAM, Bouman M, van Erp-Baart MA, Goldbohm RA. FOFREX: a computerized system to develop food frequency questionnaires. *Eur. J. Clin. Nutr.* 1998; **52**: S66.
 - 83 Wise A. Food frequency questionnaire design by computer [abstract]. *Eur. J. Clin. Nutr.* 1998; **52**(Suppl. 2): S15.
 - 84 Silvennoinen J, Lamberg-Allardt C, Karkkainen M, Niemela S, Lehtola J. Dietary calcium intake and its relation to bone mineral density in patients with inflammatory bowel disease. *J. Intern. Med.* 1996; **240**: 285–92.
 - 85 Andon MB, Smith KT, Bracker M, Sartoris D, Saltman P, Strause L. Spinal bone density and calcium intake in healthy postmenopausal women. *Am. J. Clin. Nutr.* 1991; **54**: 927–9.
 - 86 Nelson M, Mayer AB, Rutherford O, Jones D. Calcium intake, physical activity and bone mass in pre-menopausal women. *J. Hum. Nutr. Diet.* 1991; **4**: 171–8.
 - 87 Serdula M, Byers T, Coates R, Mokdad A, Simoes EJ, Eldridge L. Assessing consumption of high-fat foods: the effect of grouping foods into single questions. *Epidemiology* 1992; **3**: 503–8.
 - 88 Krebs-Smith SM, Heimendinger J, Subar AF, Patterson BH, Pivonka E. Using food frequency questionnaires to estimate fruit and vegetable intake: association between the number of questions and total intakes. *J. Nutr. Educ.* 1995; **27**: 80–5.

- 89 Cummings SR, Block G, McHenry K, Baron RB. Evaluation of two food frequency methods of measuring dietary calcium intake. *Am. J. Epidemiol.* 1987; **126**: 796–802.
- 90 Blalock SJ, Currey SS, DeVellis RF, Anderson JJB, Gold DT, Dooley MA. Using a short food frequency questionnaire to estimate dietary calcium consumption: a tool for patient education. *Arthrit. Care Res.* 1998; **11**: 479–84.
- 91 Schaffer DM, Coates AO, Caan BJ, Slattery ML, Potter JD. Performance of a shortened telephone-administered version of a quantitative food frequency questionnaire. *Ann. Epidemiol.* 1997; **7**: 463–71.
- 92 Pietinen P, Hartman AM, Haapa E, Rasanen L, Haapakoski J, Palmgren J, *et al.* Reproducibility and validity of dietary assessment instruments. II. A qualitative food frequency questionnaire. *Am. J. Epidemiol.* 1988; **128**: 667–76.
- 93 Pietinen P, Hartman AM, Haapa E, Rasanen L, Haapakoski J, Palmgren J, *et al.* Reproducibility and validity of dietary assessment instruments. I. A self-administered food questionnaire with a portion size picture booklet. *Am. J. Epidemiol.* 1988; **128**: 655–66.
- 94 Faggiano F, Vineis P, Cravanzola D, Pisani P, Xompero G, Riboli E, *et al.* Validation of a method for the estimation of food portion size. *Epidemiology* 1992; **3**: 379–82.
- 95 Hunter DJ, Sampson L, Stampfer MJ, Colditz GA, Rosner B, Willett WC. Variability in portion sizes of commonly consumed foods among a population of women in the United States. *Am. J. Epidemiol.* 1988; **127**: 1240–9.
- 96 Conn JA, Rutishauser IHE, Wheeler CE. Portion size data for foods consumed by a randomly selected sample of Geelong adults. *Aust. J. Nutr. Diet.* 1994; **51**: 58–65.
- 97 Tjonneland A, Haraldsdottir J, Overvad K, Stripp C, Ewertz M, Jensen OM. Influence of individually estimated portion size data on the validity of a semiquantitative food frequency questionnaire. *Int. J. Epidemiol.* 1992; **21**: 770–7.
- 98 Nelson M, Haraldsdottir J. Food photographs: practical guidelines I. Design and analysis of studies to validate portion size estimates. *Public Health Nutr.* 1998; **1**: 219–30.
- 99 Tsuchida K, Mizushima S, Toba M, Soda K. Dietary soybeans intake and bone mineral density among 995 middle-aged women in Yokohama. *J. Epidemiol.* 1999; **9**: 14–9.
- 100 Caan BJ, Lanza E, Schatzkin A, Coates AO, Brewer BK, Slattery ML, *et al.* Does nutritionist review of a self-administered food frequency questionnaire improve data quality? *Public Health Nutr.* 1999; **2**: 565–9.
- 101 Fox TA, Heimendinger J, Block G. Telephone surveys as a method for obtaining dietary information: a review. *J. Am. Diet. Assoc.* 1992; **92**: 729–32.
- 102 Lyu LC, Hankin JH, Liu LQ, Wilkens LR, Lee JH, Goodman MT, *et al.* Telephone vs face-to-face interviews for quantitative food frequency assessment. *J. Am. Diet. Assoc.* 1998; **98**: 44–8.
- 103 Ritenbaugh C, Peng YM, Aickin M, Graver E, Branch M, Alberts DS. New carotenoid values for foods improve relationship of food frequency questionnaire intake estimates to plasma values. *Cancer Epidemiol. Biomark. Prev.* 1996; **5**: 907–12.
- 104 Cowin I, Emmett P. The effect of missing data in the supplements to McCance and Widdowson's food tables on calculated nutrient intakes. *Eur. J. Clin. Nutr.* 1999; **53**: 891–4.
- 105 Clapp JA, McPherson RS, Reed DB, Hsi BP. Comparison of a food frequency questionnaire using reported vs standard portion sizes for classifying individuals according to nutrient intake. *J. Am. Diet. Assoc.* 1991; **91**: 316–20.
- 106 Haraldsdottir J, Tjonneland A, Overvad K. Validity of individual portion size estimates in a food frequency questionnaire. *Int. J. Epidemiol.* 1994; **23**: 786–96.
- 107 Kassam KT, Nanchahal K, Mangtani P, Santos Silva I, McMichael A, Anderson A, *et al.* Development of an interview-administered food-frequency questionnaire for use amongst women of South Asian ethnic origin in Britain. *J. Hum. Nutr. Diet.* 1999; **12**: 7–19.
- 108 Lanham SA, Bolton-Smith C. Development of food frequency questionnaire [abstract]. *Proc. Nutr. Soc.* 1993; **52**: 330A.
- 109 Tsubono Y, Kobayashi M, Takahashi T, Iwase Y, Itoi Y, Akabane M, *et al.* Within- and between-person variations in portion sizes of foods consumed by the Japanese population. *Nutr. Cancer* 1997; **29**: 140–5.
- 110 Wolk A, Ljung H, Vessby B, Hunter D, Willett WC. Effect of additional questions about fat on the validity of fat estimates from a food frequency questionnaire. Study Group of MRS SWEA. *Eur. J. Clin. Nutr.* 1998; **52**: 186–92.
- 111 Dorant E, van den Brandt PA, Goldbohm RA, Hermus RJ, Sturmans F. Agreement between interview data and a self-administered questionnaire on dietary supplement use. *Eur. J. Clin. Nutr.* 1994; **48**: 180–8.
- 112 Calvert C, Cade J, Barrett JH, Woodhouse A. Using cross-check questions to address the problem of mis-reporting of specific food groups on food frequency questionnaires. United Kingdom Women's Cohort Study Steering Group. *Eur. J. Clin. Nutr.* 1997; **51**: 708–12.
- 113 Subar AF, Thompson FE, Smith AF, Jobe JB, Ziegler RG, Potischman N, *et al.* Improving food frequency questionnaires: a qualitative approach using cognitive interviewing. *J. Am. Diet. Assoc.* 1995; **95**: 781–8.
- 114 Bueno de Mesquita HB, Smeets FW, Runia S, Hulshof KF. The reproducibility of a food frequency questionnaire among controls participating in a case-control study on cancer. *Nutr. Cancer* 1992; **18**: 143–56.
- 115 Engle A, Lynn LL, Koury K, Boyar AP. Reproducibility and comparability of a computerized, self-administered food frequency questionnaire. *Nutr. Cancer* 1990; **13**: 281–92.
- 116 Jacobsen BK, Bonna KH. The reproducibility of dietary data from a self-administered questionnaire. The Tromso Study. *Int. J. Epidemiol.* 1990; **19**: 349–53.
- 117 Morabia A, Moore M, Wynder EL. Reproducibility of food frequency measurements and inferences from a case-control study. *Epidemiology* 1990; **1**: 305–10.
- 118 Tsubono Y, Nishino Y, Fukao A, Hisamichi S, Tsugane S. Temporal change in the reproducibility of a self-administered food frequency questionnaire. *Am. J. Epidemiol.* 1995; **142**: 1231–5.
- 119 Altman DG, Bland JM. Measurement in medicine: the analysis of method comparison studies. *Statistician* 1983; **32**: 307–17.
- 120 Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; **1**: 307–10.
- 121 Bland JM, Altman DG. Comparing methods of measurement: why plotting difference against standard method is misleading. *Lancet* 1995; **346**: 1085–7.
- 122 Bland JM, Altman DG. Measuring agreement in method comparison studies [review]. *Statist. Meth. Med. Res.* 1999; **8**: 135–60.
- 123 O'Brien E, Petrie J, Littler W, de Swiet M, Padfield PL, Altman DG, *et al.* The British Hypertension Society protocol for the evaluation of blood pressure measuring devices. *J. Hypertens.* 1993; **11**: S43–62.
- 124 Kinlay S, Elliott H, Heller RF. Identifying individuals with high fat levels and low P:S ratios, in their diets, for intensive dietary intervention. *J. Intern. Med.* 1997; **241**: 407–414.
- 125 Tokunaga S, Hirohata T, Hirohata I. Reproducibility of dietary and other data from a self-administered questionnaire. *Environ. Health Perspect.* 1994; **102**(Suppl. 8): 5–10.
- 126 Hankin JH, Wilkens LR, Kolonel LN, Yoshizawa CN. Validation of a quantitative diet history method in Hawaii. *Am. J. Epidemiol.* 1991; **133**: 616–28.

- 127 Wheeler CE, Rutishauser IHE, O'Dea K. Comparison of nutrient intake data from two food frequency questionnaires and weighed records. *Aust. J. Nutr. Diet.* 1995; **52**: 140–8.
- 128 Xing X, Burr JA, Brasure JR, Neugut AI, Marshall JR. Reproducibility of nutrient intake in a food frequency questionnaire used in a general population. *Nutr. Cancer* 1996; **25**: 259–68.
- 129 Nelson M. The validation of dietary assessment. In: Margetts BM, Nelson M, eds. *Design Concepts in Nutritional Epidemiology*. Oxford: Oxford University Press, 1997; 241–72.
- 130 Burema J, van Staveren WA, van den Brandt PA. Validity and reproducibility. In: Cameron ME, van Staveren WA, eds. *Manual on Methodology for Food Consumption Studies*. Oxford: Oxford University Press, 1988; 171–81.
- 131 McPherson RS, Kohl HW III, Garcia G, Nichaman MZ, Hanis CL. Food-frequency questionnaire validation among Mexican-Americans: Starr County, Texas [see comments]. *Ann. Epidemiol.* 1995; **5**: 378–85.
- 132 Altman DG. *Practical Statistics for Medical Research*. London: Chapman and Hall, 1991.
- 133 Bingham SA, Cassidy A, Cole TJ, Welch A, Runswick SA, Black AE, *et al.* Validation of weighed records and other methods of dietary assessment using the 24 h urine nitrogen technique and other biological markers. *Br. J. Nutr.* 1995; **73**: 531–50.
- 134 Hernandez-Avila M, Romieu I, Parra S, Hernandez-Avila J, Madrigal H, Willett W. Validity and reproducibility of a food frequency questionnaire to assess dietary intake of women living in Mexico City. *Salud Publica de Mexico* 1998; **40**: 133–40.
- 135 Hebert JR, Gupta PC, Bhonsle RB, Sinor PN, Mehta H, Mehta FS. Development and testing of a quantitative food frequency questionnaire for use in Gujarat, India. *Public Health Nutr.* 1999; **2**: 39–50.
- 136 Kristal AR, Shattuck AL, Henry HJ, Fowler AS. Rapid assessment of dietary intake of fat, fiber, and saturated fat: validity of an instrument suitable for community intervention research and nutritional surveillance. *Am. J. Health Prom.* 1990; **4**: 288–95.
- 137 Black AE, Coward WA, Cole TJ, Prentice AM. Human energy expenditure in affluent societies: an analysis of 574 doubly-labelled water measurements. *Eur. J. Clin. Nutr.* 1996; **50**: 72–92.
- 138 Isaksson B. Urinary nitrogen output as a validity test in dietary surveys [letter]. *Am. J. Clin. Nutr.* 1980; **33**: 4–5.
- 139 Bingham SA. The dietary assessment of individuals; methods, accuracy, new techniques and recommendations. *Nutr. Abstr. Rev., A (Hum. Exp.)* 1987; **57**: 705–42.
- 140 Bingham SA, Gill C, Welch A, Day K, Cassidy A, Khaw KT, *et al.* Comparison of dietary assessment methods in nutritional epidemiology: weighed records v. 24 h recalls, food-frequency questionnaires and estimated-diet records [see comments]. *Br. J. Nutr.* 1994; **72**: 619–43.
- 141 Potosky AL, Block G, Hartman AM. The apparent validity of diet questionnaires is influenced by number of diet-record days used for comparison. *J. Am. Diet. Assoc.* 1990; **90**: 810–3.
- 142 Stram DO, Longnecker MP, Shames L, Kolonel LN, Wilkens LR, Pike MC, *et al.* Cost-efficient design of a diet validation study. *Am. J. Epidemiol.* 1995; **142**: 353–62.
- 143 Holland B, Welch AA, Unwin ID, Buss DH, Paul AA, Southgate AT. *McCance and Widdowson's The Composition of Foods*. The Royal Society of Chemistry and Ministry of Agriculture, Fisheries and Food. London: HMSO, 1992.
- 144 West CE, van Staveren WA. Food consumption, nutrient intake, and the use of food composition tables. In: Margetts BM, Nelson M, eds. *Design Concepts in Nutritional Epidemiology*. Oxford: Oxford University Press, 1997; 107–22.
- 145 Klesges RC, Eck LH, Ray JW. Who underreports dietary intake in a dietary recall? Evidence from the Second National Health and Nutrition Examination Survey. *J. Consult. Clin. Psychol.* 1995; **63**: 438–44.
- 146 Pryer JA, Vrijheid M, Nichols R, Kiggins M, Elliott P. Who are the low energy reporters in the dietary and nutritional survey of British adults? *Int. J. Epidemiol.* 1997; **26**: 146–54.
- 147 Schofield WN, Schofield C, James WPT. Basal metabolic rate – review and prediction. *Hum. Nutr. Clin. Nutr.* 1985; **39**: 1–96.
- 148 Beaton GH. Interpretation of results from diet history studies. In: *The Diet History Method: Proceedings of the 2nd Berlin Meeting on Nutritional Epidemiology*. London: Smith-Gordon & Co. Ltd, 1991; 15–38.
- 149 Dunn G. *Design and Analysis of Reliability Studies*. London: Edward Arnold, 1989.
- 150 Ocke MC, Kaaks RJ. Biochemical markers as additional measurements in dietary validity studies: application of the method of triads with examples from the European Prospective Investigation into Cancer and Nutrition [review]. *Am. J. Clin. Nutr.* 1997; **65**: 1240S–5S.
- 151 Wong MY, Day NE, Wareham NJ. Measurement error in epidemiology: the design of validation studies II: bivariate situation. *Statist. Med.* 1999; **18**: 2831–45.
- 152 Kaaks R, Riboli E, Esteve J, Van Kappel AL, van Staveren WA. Estimating the accuracy of dietary questionnaire assessments: validation in terms of structural equation models. *Statist. Med.* 1994; **13**: 127–42.
- 153 Plummer M, Clayton D. Measurement error in dietary assessment: an investigation using covariance structure models. Part I. *Statist. Med.* 1993; **12**: 925–35.
- 154 Kipnis V, Carroll RJ, Freedman LS, Li L. Implications of a new dietary measurement error model for estimation of relative risk: application to four calibration studies. *Am. J. Epidemiol.* 1999; **150**: 642–51.
- 155 Baghurst KI, Baghurst PA, Record SJ. Demographic and dietary profiles of high and low fat consumers in Australia. *J. Epidemiol. Community Health* 1994; **48**: 26–32.
- 156 Franceschi S, Favero A, La Vecchia C, Negri E, Conti E, Montella M, *et al.* Food groups and risk of colorectal cancer in Italy. *Int. J. Cancer* 1997; **72**: 56–61.
- 157 Trichopoulou A, Georgiou E, Bassiakos Y, Lipworth L, Laggiou P, Proukakis C, *et al.* Energy intake and monounsaturated fat in relation to bone mineral density among women and men in Greece. *Prev. Med.* 1997; **26**: 395–400.
- 158 Schuurman AG, Goldbohm RA, Dorant E, van den Brandt PA. Vegetable and fruit consumption and prostate cancer risk: a cohort study in The Netherlands. *Cancer Epidemiol. Biomark. Prev.* 1998; **7**: 673–680.
- 159 Rimm EB, Willett WC, Hu FB, Sampson L, Colditz GA, Manson JE, *et al.* Folate and vitamin B6 from diet and supplements in relation to risk of coronary heart disease among women [see comments]. *J. Am. Med. Assoc.* 1998; **279**: 359–64.
- 160 Hertog MG, Sweetnam PM, Fehily AM, Elwood PC, Kromhout D. Antioxidant flavonols and ischemic heart disease in a Welsh population of men: the Caerphilly Study. *Am. J. Clin. Nutr.* 1997; **65**: 1489–94.
- 161 Albert CM, Hennekens CH, O'Donnell CJ, Ajani UA, Carey VJ, Willett WC, *et al.* Fish consumption and risk of sudden cardiac death [see comments]. *J. Am. Med. Assoc.* 1998; **279**: 23–8.
- 162 Andersson S-O, Wolk A, Bergstrom R, Giovannucci E, Lindgren C, Baron J, *et al.* Energy, nutrient intake and prostate cancer risk: a population-based case-control study in Sweden. *Int. J. Cancer* 1996; **68**: 716–22.
- 163 Tjonneland A, Gronbaek M, Stripp C, Overvad K. Wine

- intake and diet in a random sample of 48 763 Danish men and women. *Am. J. Clin. Nutr.* 1999; **69**: 49–54.
- 164 Bolton-Smith C, Smith WC, Woodward M, Tunstall-Pedoe H. Nutrient intakes of different social-class groups: results from the Scottish Heart Health Study (SHHS). *Br. J. Nutr.* 1991; **65**: 321–35.
- 165 Osler M, Heitmann BL, Schroll M. Ten year trends in the dietary habits of Danish men and women. Cohort and cross-sectional data. *Eur. J. Clin. Nutr.* 1997; **51**: 535–41.
- 166 Williams DE, Wareham NJ, Cox BD, Byrne CD, Hales CN, Day NE. Frequent salad vegetable consumption is associated with a reduction in the risk of diabetes mellitus. *J. Clin. Epidemiol.* 1999; **52**: 329–35.
- 167 Margetts BM, Thompson RL, Speller V, McVey D. Factors which influence healthy eating patterns: results from the 1993 Health Education Authority health and lifestyle survey in England. *Public Health Nutr.* 1998; **1**: 193–8.
- 168 Prevost AT, Whiclow MJ, Cox BD. Longitudinal dietary changes between 1984–5 and 1991–2 in British adults: association with socio-demographic, lifestyle and health factors. *Br. J. Nutr.* 1997; **78**: 873–88.
- 169 Smith AM, Smith C. Dietary intake and lifestyle patterns: correlates with socio-economic, demographic and environmental factors. *J. Hum. Nutr. Diet.* 1994; **7**: 283–94.
- 170 Slattery ML, Schumacher MC, Smith KR, West DW, Abd-Elghany N. Physical activity, diet, and risk of colon cancer in Utah [see comments]. *Am. J. Epidemiol.* 1988; **128**: 989–9.
- 171 Potischman N, Weiss HA, Swanson CA, Coates RJ, Gammon MD, Malone KE, *et al.* Diet during adolescence and risk of breast cancer among young women. *J. Natl. Cancer Inst.* 1998; **90**: 226–33.
- 172 Tzonou A, Lagiou P, Trichopoulou A, Tsoutsos V, Trichopoulos D. Dietary iron and coronary heart disease risk: a study from Greece. *Am. J. Epidemiol.* 1998; **147**: 161–6.
- 173 Bakkum A, Bloemberg B, van Staveren WA, Verschuren M, West CE. The relative validity of a retrospective estimate of food consumption based on a current dietary history and a food frequency list. *Nutr. Cancer* 1988; **11**: 41–53.
- 174 Byers T, Marshall J, Anthony E, Fiedler R, Zielezny M. The reliability of dietary history from the distant past [published erratum appears in *Am. J. Epidemiol.* 1987; **126**(6): 1220]. *Am. J. Epidemiol.* 1987; **125**: 999–1011.
- 175 Goldbohm RA, van't Veer P, van den Brandt PA, van't Hof MA, Brants HA, Sturmans F, *et al.* Reproducibility of a food frequency questionnaire and stability of dietary habits determined from five annually repeated measurements. *Eur. J. Clin. Nutr.* 1995; **49**: 420–9.
- 176 Rohan TE, Potter JD. Retrospective assessment of dietary intake. *Am. J. Epidemiol.* 1984; **120**: 876–87.
- 177 Rothenberg E. Validation of the food frequency questionnaire with the 4-day record method and analysis of 24-h urinary nitrogen. *Eur. J. Clin. Nutr.* 1994; **48**: 725–35.
- 178 D'Avanzo B, La Vecchia C, Katsouyanni K, Negri E, Trichopoulos D. An assessment, and reproducibility of food frequency data provided by hospital controls. *Eur. J. Cancer Prev.* 1997; **6**: 288–93.
- 179 Maclure M, Travis LB, Willett W, MacMahon B. A prospective cohort study of nutrient intake and age at menarche. *Am. J. Clin. Nutr.* 1991; **54**: 649–56.
- 180 Bostick RM, Potter JD, Sellers TA, McKenzie DR, Kushi LH, Folsom AR. Relation of calcium, vitamin D, and dairy food intake to incidence of colon cancer among older women. The Iowa Women's Health Study [review]. *Am. J. Epidemiol.* 1993; **137**: 1302–17.
- 181 Jacques PF, Taylor A, Hankinson SE, Willett WC, Mahnken B, Lee Y, *et al.* Long-term vitamin C supplement use and prevalence of early age-related lens opacities [see comments]. *Am. J. Clin. Nutr.* 1997; **66**: 911–6.
- 182 Aldoori WH, Giovannucci EL, Stampfer MJ, Rimm EB, Wing AL, Willett WC. Prospective study of diet and the risk of duodenal ulcer in men. *Am. J. Epidemiol.* 1997; **145**: 42–50.
- 183 Botterweck AA, van den Brandt PA, Goldbohm RA. A prospective cohort study on vegetable and fruit consumption and stomach cancer risk in The Netherlands. *Am. J. Epidemiol.* 1998; **148**: 842–53.
- 184 Hu FB, Stampfer MJ, Manson JE, Rimm E, Colditz GA, Speizer FE, *et al.* Dietary protein and risk of ischemic heart disease in women. *Am. J. Clin. Nutr.* 1999; **70**: 221–7.
- 185 Elmstahl S, Holmqvist O, Gullberg B, Johansson J, Berglund G. Dietary patterns in high and low consumers of meat in a Swedish cohort study. *Appetite* 1999; **32**: 191–206.
- 186 Hu FB, Rimm E, Smith-Warner SA, Feskanich D, Stampfer MJ, Ascherio A, *et al.* Reproducibility and validity of dietary patterns assessed with a food-frequency questionnaire. *Am. J. Clin. Nutr.* 1999; **69**: 243–9.
- 187 Greenwood DC, Cade JE, Draper A, Barrett JH, Calvert C, Greenhalgh A. Seven unique food consumption patterns identified among women in the UK Women's Cohort Study. *Eur. J. Clin. Nutr.* 2000; **54**: 314–20.
- 188 Little P, Barnett J, Kinmonth AL, Margetts B, Gabbay J, Thompson R, *et al.* Can dietary assessment in general practice target patients with unhealthy diets? *Br. J. Gen. Pract.* 2000; **50**: 43–5.
- 189 Blom L, Lundmark K, Dahlquist G, Persson LA. Estimating children's eating habits. Validity of a questionnaire measuring food frequency compared to a 7-day record. *Acta Paediatr. Scand.* 1989; **78**: 858–64.
- 190 Nelson M, White J, Rhodes C. Haemoglobin, ferritin, and iron intakes in British children aged 12–14 years: a preliminary investigation. *Br. J. Nutr.* 1993; **70**: 147–55.
- 191 Little P, Barnett J, Margetts B, Kinmonth AL, Gabbay J, Thompson R, *et al.* The validity of dietary assessment in general practice. *J. Epidemiol. Community Health* 1999; **53**: 165–72.
- 192 Ritenbaugh C, Aicken M, Taren D, Teufel N, Graver E, Woolf K, Alberts DS. Use of a food frequency questionnaire to screen for dietary eligibility in a randomised cancer prevention phase III trial. *Cancer Epidemiol. Biomark. Prev.* 1997; **6**: 347–54.

Appendix – Summary recommendations of the review

Recommendations on food-frequency questionnaire design

Consider the following points before designing a food-frequency questionnaire:

- Is information needed about foods, nutrients, dietary supplements or other food constituents, or specific dietary behaviours?
- Is frequency of consumption required?
- Is amount of consumption required?
- Is information on one food/nutrient or a range required?
- Is the population mean or individual intake required?
- Is absolute or relative intake needed?
- Is information on dietary change required?

- What level of accuracy is required?
- What is the time period of interest?
- What are the research constraints in terms of money, time, staff and respondent characteristics?
- Consult a statistician and nutritionist before embarking.

Recommendations on when the use of a food-frequency questionnaire may not be appropriate

- With small numbers of subjects.
- For surveillance and monitoring of current levels where *accurate absolute intakes* are required.
- FFQs developed in one country should not be adapted for use in another country unless dietary habits are very similar.
- In clinical work when precise intakes are required.

Recommendation on modifying existing questionnaires

- Modification of pre-existing food-frequency questionnaires for use in similar populations is useful; however, the purpose of the original and new version should be carefully considered.

Recommendation on developing the food list

- Unless the purpose of the food-frequency questionnaire is very specific a comprehensive food list is desirable.

Recommendations on grouping of food items

- Grouping of food items should be decided *a priori* according to the purpose of the questionnaire.
- Single items are preferable for food groups of main interest.
- Some grouping of foods may need to be considered to prevent excessive questionnaire length.

Recommendation on closed and open questions

- If it is necessary to use open questions, the questionnaire should be interviewer-administered rather than self-administered.

Recommendations on estimation of portion size

- Allowing subjects to estimate their portion size is more advantageous than using average portion sizes.
- Suitable methods are the use of defined small/medium/large options or estimation of portion size using photographs.

Recommendations on method of questionnaire administration

- If practical, interviewer-administered food-frequency questionnaires should be used in preference to self-administered questionnaires.
- If self-administered versions are used then checking the questionnaires for completeness should be undertaken.

Recommendation on pre-testing of food-frequency questionnaires

- Every questionnaire should be rigorously pre-tested to ensure that the meanings of the food names and the portion size descriptors are clear to the subjects, instructions are clear and that the method for recording responses is unambiguous.

Recommendation on rejection of data from food-frequency questionnaires

- The criteria for rejection of data (cut-offs) decided *a priori* should be stated in any publications, with the number (percentage) of questionnaires rejected.

Recommendations on tests for FFQ reproducibility

- Reproducibility should always be assessed.
- It should be assessed in a representative sample of the target population.
- The statistical methods used must take into account the purpose of the food-frequency questionnaire. For example, all foods and nutrients to be assessed in the main study should be assessed for reproducibility.
- The interval between repeat measurements should be chosen to minimise changes over time and recall of previous answers, and will depend on the reference period of the questionnaire.

Recommendations on use of the correlation coefficient for assessment of FFQ reproducibility

- Correlation is inappropriate to measure reproducibility.
- If correlation is used, Pearson's correlation should be used for normally distributed data, and Spearman's correlation for non-normally distributed data.

Recommendations on use of the Bland–Altman method for the assessment of FFQ reproducibility

- The methods developed by Bland and Altman should be used to assess reproducibility/repeatability, rather than correlation.
- They should be used in context and interpreted in the light of the target population and what the acceptable levels of bias and limits of agreement are in this context.

Recommendation on other statistical methods for the assessment of FFQ reproducibility

- Kappa statistics can be used instead of the Bland–Altman method for measures involving small numbers of ordered categories.

Recommendations on validation of food-frequency questionnaires

- Food-frequency questionnaires should always be validated.
- Validation studies should use similar populations to the intended main study.
- The methods of validation must take into account the purpose of the food-frequency questionnaire. For example, all foods and nutrients to be assessed in the main study should be assessed.
- Using more than one approach to validation gives added credence to the results.

Recommendation on sample/population selection in validation studies

- The subjects recruited for the validation study should be representative of the main study target population.

Recommendation on time frame of reference method in validation studies

- The period of assessment should be the same for both test and reference methods.

Recommendations on selection of dietary method for validation studies

- Both 24-hour recalls and non-weighed/weighed dietary records are suitable for a reference method.
- Multiple days of collection of dietary data should be undertaken.

Recommendations on use of biomarkers in validation studies

- Consider carefully what is being measured.
- Take into account all possible errors associated with the method.
- Take into consideration the relevant time frame, and relationship between biological variation and variation in dietary intake.

Recommendations on correlation, regression and the Bland–Altman method in validation studies

- The methods developed by Bland and Altman should be used to measure the agreement between food-

frequency questionnaires and other measures of dietary intake.

- These methods should be used in context and interpreted in the light of the target population and what the acceptable levels of bias and limits of agreement are in this context.
- Regression or correlation may be used in conjunction with the Bland–Altman method.
- If correlation is used, Pearson's correlation should be used for normally distributed data, Spearman's correlation for non-normally distributed data.
- Kappa and/or sensitivity, specificity, etc. may be appropriate if the data are ordered categorical or binary.

Recommendations on comparison of group means within validation studies

- Group means should be assessed if absolute intakes are required.
- If ranking is important, paired tests should be used, interpreting the *P*-value with caution.
- The tests used should reflect the type and distribution of the data.

Recommendations on sample size for validation studies

- Expert statistical advice should be sought to estimate required number of subjects.
- If resources are available, higher numbers of subjects will provide better estimates of reproducibility or validity.
- A sample size of at least 50 to 100 subjects for each demographic group is recommended.

Recommendation on sequence of administration

- The test instrument should be administered prior to the assessment of the reference measure.

Recommendations on description of food-frequency questionnaire in utilisation studies

- State the original objective of the questionnaire.
- If questionnaire has previously been used provide a reference.
- Describe the questionnaire in terms of number of food items, frequency choices, and how portion size has been assessed.
- State whether the questionnaire was self-administered or interviewer-administered.
- Describe the nutrient database used to calculate intake.

Recommendations on details of reproducibility and validity that should be reported in utilisation studies

- Give reference to any published validation studies.
- Describe the population, sample size and reference method used.
- Briefly report the results of the validation study for the foods/nutrients of interest (the results should reflect the purpose of the questionnaire).
- Report any adjustments to the data (e.g. use of cross-check questions).
- If there has not been a validation study, give the reason.



World Cancer
Research Fund
International

FOOD, NUTRITION AND PREVENTION OF CANCER

Grant Application Announcement

New grant cycle

Outline applications accepted up to 20 September 2002.

Further details available from www.wcrf.org

World Cancer Research Fund International is now inviting applications for grants to fund research into the effects of food and nutrition on the origins, causes and prevention of cancer. Grants will be for a maximum of £150,000 over four years, with no more than £50,000 in any one year.

In general, preference is given to applications that address the dietary and research recommendations of the WCRF Expert Report, and

- consider diets or foods as a whole and/or use whole-body systems
- use foods or dietary constituents in physiological amounts
- have direct relevance to human cancer
- address cancer prevention in preference to treatment

Specific priorities are: nutrition of the mother, fetus, neonate and young child, and lifelong cancer risk; energy intake, body composition (eg. growth and height, body weight), physical activity, and cancer risk; projects focusing on nutrition and cancer in the developing world and countries in transition from developing to developed status; and studies investigating the effectiveness of interventions or programmes aimed at changing eating habits at individual, community or population level.

Closing date for outline applications is 20 September 2002. Applications from anywhere in the world except North America will be considered.

World Cancer Research Fund International

Research Department

First Floor, 19 Harley Street, London W1G 9QJ

Telephone: +44 (0) 20 7343 4200, Fax: +44 (0) 20 7343 4220

www.wcrf.org