Food-frequency questionnaires: a review of their design, validation and utilisation

J. E. Cade1*, V. J. Burley1, D. L. Warm2, R. L. Thompson2 and B. M. Margetts2

1Nutritional Epidemiology Group, University of Leeds, Leeds LS2 9JT, UK
2Institute of Human Nutrition, University of Southampton, Southampton SO16 6YD, UK

A review of the literature concerning the design, utilisation and validation of food-frequency questionnaires (FFQ) has been carried out using a semi-systematic approach to obtaining, reviewing and extracting data from articles. Databases were searched from 1980 to 1999. The present review identified 227 validation (from 1980 to September 1999) and 164 utilisation (for 1998 only) studies. A number of design issues have been evaluated through the present review. These include: the need to consider how portion sizes have been described, self-defined giving higher mean correlations; how an FFQ was administered, interviewer-administered giving higher mean correlations for some nutrients; how many items to include on an FFQ, those with the largest number of items having higher correlations. Validation techniques were described. Most validation studies involved comparing an FFQ against another dietary assessment method; only 19% compared an FFQ to a biomarker. Measurement differences were most commonly assessed by correlation coefficients as opposed to other more appropriate methods. Mean correlation coefficients were highest for Ca and fat, and lowest for vitamin A and vegetables. The utilisation studies showed that FFQ were most commonly used in cross-sectional surveys, with ninety-three of the FFQ being designed to be disease-specific. The present review results were presented to a group of experts and a consensus arrived at concerning the development, validation and use of FFQ. Recommendations derived from the consensus arising from the literature review are presented as an appendix to the present paper.

Food-frequency questionnaires: Validation: Reproducibility

Introduction

During the 1950s and 1960s nutritionists started to develop questionnaires for the assessment of habitual food intake based on a checklist of foods consumed over a set time period because of problems encountered with 24 h recall techniques and weighed inventories. After further refinement, revision and appraisal during the 1980s and 1990s, food-frequency questionnaires (FFQ) have become one of the key research tools in nutrition epidemiology. However, despite considerable advantages in terms of ease of administration and analysis, FFQ may be limited in their usefulness and, through poor design and inappropriate use, may not yield the required information. No dietary method can measure dietary intake without error (Margetts & Nelson, 1997) so it is important that sources of error are taken into account. No formal systematic review of the FFQ method has been carried out to date. The present study has brought the results of research together to evaluate the method and assess ‘best practice’ for further studies.

Methodology

A systematic review approach (Chalmers & Altman, 1995) was developed to identify relevant studies describing the design, evaluation and/or use of FFQ. The definition of an FFQ was specified for the purposes of the search as ‘any list of one or more foods with frequency of intake categories’. However, since the focus of the present review was the design, validation and utilisation of FFQ, papers were excluded in which the results of the FFQ were combined with another dietary assessment technique, as for example in the dietary history method. Questionnaires that assessed only vitamin and mineral supplement intakes, alcohol, or contaminants (such as heavy metals) were excluded, as were articles written in languages other than English.

Abbreviations: FFQ, food-frequency questionnaire; OR, odds ratio.
* Corresponding author: Dr J. E. Cade, fax +44 113 343 3470, email j.e.cade@leeds.ac.uk
A comprehensive search procedure was developed which involved searching electronic databases including Medline (http://medline.cos.com/), Embase (www.embase.com/), CancerLit (www.cancer.gov/search/cancer_literature/), CAB Abstracts (www.cabi-publishing.org/Products/Database/Abstracts/Index.asp) and Dissertation Abstracts (http://library.dialog.com/bluesheets/html/bI0035.html) and the online Dietary Assessment Calibration/Validation Register (www.dacv.ims.nci.nih.gov/). The databases 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. Personal contact with experts was also undertaken to identify any results not in the literature to date, for example through the Nutritional Epidemiology Group and at a small number of conferences.

The papers to be included in the present review were identified using specific search terms based on the Medical Subject Headings (MeSH) terms and keywords. Searches were carried out for the terms ‘food frequency questionnaires’, ‘reproducibility’, ‘validity/validation’, ‘diet-study-techniques’ and ‘calibration’. All references were downloaded into the computerised bibliographic program Reference Manager (Research Information Systems, 1999) which facilitated the handling of the large numbers of publications obtained.

Aspects of utilisation were assessed from references published in 1998. A 1-year sample was thought to be adequate to generate data on the way FFQ are currently being used since there were a large number of references which had used FFQ.

Standard data-extraction forms were developed and included information on general aspects of the study, features of the FFQ, details of associated validation and utilisation papers, and results in terms of correlation coefficients. Information from the references identified by the search was entered onto a Microsoft Access (version 2) database which broadly categorised papers into three types: validation, reproducibility, or utilisation. One member of the team (D. W.) carried out the data extraction although any queries as to whether a study should be included or not were referred to other members of the team. Studies were not assessed for quality before inclusion in the database. Therefore all relevant studies meeting the inclusion criteria were included and quality was not judged at this stage.

The large number of studies which were identified by the search and the widely varying nature of the studies in design and presentation lead to some difficulties in terms of standardisation of data extraction. It is possible that some papers were missed out from the present review. The aim was to ensure that all the key papers (published and unpublished) were identified; nevertheless some may have been missed.

Following extraction of the data, the 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, December 1999. The group members agreed to comment on the text of the consensus document being prepared. The consensus document was also sent to a number of international experts. Following receipt of these comments the consensus document was amended, taking the whole body of evidence into account (Cade et al. 2002).

The purpose of the present paper is to describe the results of the review of the literature on FFQ.

Results

The searches yielded 1982 references of which 779 hard copies of papers were obtained. This number included review articles and a number of publications that originated from the same study. Finally, 227 validation studies (196 single studies and thirty-one groups of studies) and 164 utilisation studies (156 single studies and eight groups of studies) were identified, and data from these studies were extracted.

Validation studies

Validation of the FFQ method is important to assess the degree to which the questionnaire measures items (foods or nutrients) for which it has been designed. Incorrect information may lead to false associations between dietary factors and diseases or disease-related markers. The issues involved with validation procedures in dietary studies are discussed in more detail in Burema et al. (1988) and Nelson (1997).

Background information on the validation studies

The 227 validation studies identified originated from thirty different countries, with most (102) originating in the USA (Table 1). Of the studies, 183 (81 %) had samples drawn from the general population. The present review also showed that 45 % of validation studies used a modified version of a previously developed questionnaire. Of the 104 questionnaires adapted from previous FFQ, twenty-six were adapted from the Block questionnaire (Block et al. 1986) and twenty-eight were adapted from the Willett questionnaire (Willett et al. 1985). Examples of validation papers which have used as a basis the Block questionnaire include: Brown & Griebler (1993); Tylavsky & Sharp (1995); Eck et al. (1996); Hartman et al. (1996); Sawaya et al. (1996); Baranowski et al. (1997); Marshall et al. (1997); Riboli et al. (1997b); Kuriniji et al. (1998); Lemaitre et al. (1998); Potischman et al. (1998a); Mayer-Davis et al. (1999). Those which have used the Willett questionnaire include: Stein et al. (1992); Byers et al. (1993); Ajani et al. (1994); Basch et al. (1994); Bingham & Day (1997); Wirfalt et al. (1998); Bell et al. (1999); Tucker et al. (1999). The numbers of validation papers published increased in the mid-1990s, suggesting that researchers appreciated the need to validate their methods (Fig. 1).

Description of the food-frequency questionnaires used in the validation studies

Of the FFQ, 115 were designed to assess foods or food groups and 166 assessed nutrients. Seventy-five questionnaires assessed single nutrients or food groups (for
A large majority (110) of the FFQ were intended to measure group or population levels of intakes, a further fifty-one stated they were intended to measure absolute intakes and fifty-eight were intended to rank individuals. Some FFQ were used to measure food patterns or to generate patterns of foods rather than nutrients (Brewer et al. 1987; Randall et al. 1989; Kant et al. 1991; Frank et al. 1992; Millen et al. 1996; Barrett et al. 1998; Musaiger and Abuirmeileh, 1998; Hu et al. 1999; Tseng, 1999). It is important to consider the purpose of the questionnaire before it is designed to ensure that the questionnaire is sensitive enough to detect differences in the nutrients of interest. In general, FFQ are designed without assessing the possible need to adjust for energy intake in the analysis. If energy adjustment is needed, the FFQ should be comprehensive enough to assess energy (Margetts et al. 1995). The FFQ measured current diet in 169 studies and past diet in forty-three studies. The most common time frame for the assessment of diet was the previous 1 year.

Of the FFQ being validated, sixty-two were interviewer-led and 153 were self-administered. In nine studies parents completed the FFQ on behalf of a child (Blom et al. 1989; Stein et al. 1992; Byers et al. 1993; Hammond et al. 1993; Kaskoun et al. 1994; Hirving et al. 1995; Bellu et al. 1996; Sloan et al. 1997; Omidvar et al. 1998). Proxy reporting is always subject to uncertainty and error. For example, a parent may not be aware of all the food eaten by a child, for example at school. If the parent completes the questionnaire in collaboration with the child and school or daycare staff, this may be helpful. Parents also tend to overestimate intakes of food considered ‘healthy’ and underestimate less ‘healthy’ foods (Hammond et al. 1993). The questionnaires administered to children tended to use standard portion sizes which had been developed from adult surveys. Parental reports, on the whole, do seem to be adequate for assessing a range of food and nutrient intakes in their children.

The mean sample size for validation studies was 255 (range 6–3750) subjects. The number of items on the FFQ ranged from five to 350, with a mean of eighty-eight items. Portion sizes were specified by the researchers on eighty-five of the questionnaires. Subjects could specify their own portion size in seventy-three studies and there was no recording of portion sizes in forty-four studies. Food photographs were used in some studies to help in portion-size description (Porrini et al. 1995).

The FFQ asked about different frequency categories with a range of one to twelve divisions of time, the most common being nine categories (fifty-three studies). The range of frequency choices should reflect the time frame of interest. Both ascending and descending frequency categories were used. The format of the questions in terms of the order of the foods, either as a list or based on meals, does not appear to have a major impact on nutrient estimates from the questionnaire (Wheeler et al. 1994).

**Examples of frequency choices**

Frequency categories that FFQ used were:

- five choices (times per d, week, month, year, rarely or never);
- nine choices (never or less than once a month, one to three per month, one per week, two to four per week, five to six per week, one per d, two to three per d, four to five per d, six or more per d);
- nine choices (7, 6, 5, 4, 3, 2, or 1 d per week, monthly, rarely or never).

The FFQ was stated to include ‘additional’ questions in fifty-four studies. These included information about vita-
min and supplement use (fourteen studies) and other foods consumed (thirteen studies).

The majority of the validated FFQ were designed to be used by the general population (173 studies). However, 32% were specifically designed and validated for use in populations with or at risk of a particular disease. Of these, thirty-three out of seventy-two were designed for use in cancer studies (for example, Riboli & Kaaks, 1997; Forman et al. 1999; Kroke et al. 1999) and four in individuals with risk factors related to heart disease, including hypercholesterolaemia (van Staveren et al. 1992; Ascherio et al. 1994). Other disease-specific FFQ were validated for use in studies of osteoporosis (Nelson et al. 1988; Welten et al. 1995; Blalock et al. 1998; Angellatt & Moller, 1999), diabetes (Blom et al. 1989; Garnett et al. 1995; Riley & Blizzard, 1995) and eye disease (Molla et al. 1993; Sloan et al. 1997). Sixty-five of the studies had an associated utilisation paper. Some FFQ have been designed to include foods which are important for different ethnic population groups (Borrud et al. 1989; Wahlqvist et al. 1991; Forsythe & Gage, 1994; Kassam-Khamis et al. 1999).

Methods used in food-frequency questionnaire validation studies

Of the FFQ, 171 (75%) were validated against another dietary assessment method (for example, Block, 1982; Block et al. 1992; Suwaya et al. 1996; Bingham, 1997; Bingham et al. 1997; Ocke et al. 1997; Field et al. 1998; Fraser et al. 1998; Hebert et al. 1998a). Forty-three (19%) were validated against one or more biomarkers (for example, Byers et al. 1993; Feunekes et al. 1993; Tjonneland et al. 1993; Porrini et al. 1995; Bingham & Day 1997; Johansson et al. 1998) and twenty-seven (12%) against another method (for example, doubly labelled water or energy expenditure studies) (Kroke et al. 1999; Lindroos et al. 1999). A number of studies used adipose tissue not only to assess fatty-acid intakes but also fat-soluble-vitamin intakes (Tjonneland et al. 1993; Kardinaal et al. 1995; Godley et al. 1996).

The studies which used a weighed record as the method of comparison, collecting between 2 and 28 d of weighed intake are: (Riley & Blizzard, 1995; Barikmo et al. 1998) at 2 d; Wolk et al. (1998) and Tucker et al. (1999) using 4 × 7 d with a mode of 7 d (n 19). The 24 h recall studies used 1–28 d of recall with a mode of 1 d (n 11). Four studies used consecutive days of recall. The studies which used biomarkers measured blood (thirty-five studies) or urine (fifteen studies). Ten studies used both blood and urine (for example, Porrini et al. 1995; Bingham, 1997; Bingham & Day, 1997; Bingham et al. 1997). Blood measures of vitamin C were used in seven studies, urinary electrolytes in twelve studies and cholesterol in twelve studies. Twenty-three studies used a dietary assessment method and a biomarker measure. However, it cannot be assumed that there will be a close or simple relationship between the dietary intake of a nutrient and its biomarker given, for example, variations in absorption and metabolism.

Comparison between methods to assess measurement differences in the validation studies used correlation coefficients in 83% (168) studies. This method is flawed because it does not measure the agreement between two methods, only the degree to which the methods are related. Correlation coefficients can be useful in conjunction with the Bland–Altman method which assesses in graphical form the agreement between the methods across the range of intakes by plotting the mean of the two methods against the difference (Bland & Altman, 1986). The validation studies showed that the Bland–Altman comparison was carried out in only a small number of studies (Thompson & Margetts, 1993; Rothenberg, 1994; Saba et al. 1995; Robinson et al. 1996; Lambe et al. 1998; Molgaard et al. 1998; Thompson et al. 1998). The other main approaches used were the Kappa statistic and Student’s t test. Twenty-five studies did not state the statistical method used for comparison and twenty-eight studies did not use the four main methods of analysis. However, correlation and regression can be useful in assessing validity. Alternatively, methods which take into account the measurement error associated with FFQ may be particularly useful, such as structural equation modelling (Kaaks et al. 1994). Some FFQ have error variances higher than the true variation and there is also a tendency for errors of measurement of different nutrients to be highly correlated (Day et al. 2001). Regression can be used to calibrate one method compared with another (Gullberg & Elmstahl, 1994; Carroll et al. 1996). Correlation or regression should not be used alone to assess validity, but should be used alongside the Bland–Altman analysis (Bland & Altman, 1986, 1995, 1999). However, since correlation coefficients were used as the only method of comparing measurement differences in the majority of studies, these have been assessed by the present review and compared in a number of ways.
lowest correlation for energy with a weighed intake was 0·21 (Bell et al. 1999). This was for a study of a Samoan population living in New Zealand. It cannot be assumed that the lowest correlation represented a poor FFQ. In fact, Bell et al. (1999) concluded that the FFQ gave a better estimate of usual intake in this population than the weighed intake. This suggests that for this population at least the weighed intake may not have been an appropriate reference intake. This suggests that for this population at least the usual intake in this population than the weighed intake. This suggests that for this population at least the usual intake may not have been an appropriate reference method. For Fe and vitamin C a higher correlation coefficient was found using the weighed record compared with the unweighed record or 24 h recall (Table 3).

**Design issues in validation studies**

The size of the validation study did not make an appreciable difference, on average, to the study results. The present review showed that correlation coefficients comparing the FFQ with a reference method were no higher for larger validation study sample sizes than smaller studies.

The issue of how portion sizes were assessed by the FFQ was considered. Correlation coefficients were highest when subjects were able to describe their own portion size compared with no portion size specified (use of average portion weights to compute intakes) or portion size specified on the questionnaire. There was little difference between no portion size specified and whether a portion size was specified (Table 4).

Portion sizes were specified by the researchers in eighty-five of the questionnaires. This presupposes knowledge about portion sizes in the population of interest. If portion sizes are unknown this information may need to be collected separately. In one study which collected portion sizes, the amounts typically consumed differed by 50% or more from published ‘standard’ portion weights for more than one-third of the foods considered (Conn et al. 1994). There were also differences in portion weights between men and women and by age group and socio-economic status. For most foods the variation in portion size within individuals exceeded that between individuals. It may be appropriate to use sex-specific ‘typical’ portion weights rather than ‘standard’ portions to estimate nutrient intake from frequency data.

The method of administration of the FFQ in validation studies was assessed. Of the questionnaires validated, 67% were self-administered. Correlation coefficients (interviewer v. self-administered) between FFQ and reference measures were higher for interviewer-administered questionnaires than self-administered questionnaires for fat (0·55 v. 0·50), energy (0·55 v. 0·46), and vitamin A (0·47 v. 0·37). Correlation coefficients were similar for Ca (0·56 v. 0·55) and slightly higher for self-administered questionnaires for vitamin C (0·45 v. 0·49). The use of interviewers may be an advantage in some situations and allows for immediate checking by the interviewer of improbable or unlikely responses. Against this is the cost of employing interviewers, ensuring standardised training processes and their presence may increase the likelihood of social desirability bias in responses from subjects.

Telephone administration of an FFQ has been carried out and compared with face-to-face interviewing. The results obtained were comparable. However, both these studies were carried out in the USA where telephone ownership is particularly high (Schaffer et al. 1997; Lyu et al. 1998). Telephone interviews have been used to supplement missing data collected from questionnaires administered by mail (Caan et al. 1991).

Most validation studies included men and women. Four studies with correlations presented included just men (Bakkum et al. 1988; Pietinen et al. 1988; Feskanich et al. 1994; Jain et al. 1996). Sixteen studies presenting correlations used only women (for example, Martin-Moreno et al. 1993; Arnold et al. 1995; Robinson et al. 1996; Friis et al. 1997; Martin et al. 1997; Baumgartner et al. 1998; Hernandez-Avila et al. 1998; Wirfalt et al. 1998; Patterson et al. 1999). The studies that had used only men showed slightly higher mean correlations for fat (0·49 v. 0·46) and

### Table 2. Mean correlations between food-frequency questionnaire and reference method

<table>
<thead>
<tr>
<th>Nutrient</th>
<th>n</th>
<th>Range of correlations</th>
<th>Mean correlation</th>
<th>Median correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total fat</td>
<td>86</td>
<td>−0·16 to 0·86</td>
<td>0·51</td>
<td>0·51</td>
</tr>
<tr>
<td>Energy</td>
<td>77</td>
<td>−0·14 to 0·92</td>
<td>0·47</td>
<td>0·46</td>
</tr>
<tr>
<td>Vitamin C</td>
<td>73</td>
<td>0·14 to 0·84</td>
<td>0·49</td>
<td>0·50</td>
</tr>
<tr>
<td>Vitamin A</td>
<td>58</td>
<td>0·05 to 0·83</td>
<td>0·39</td>
<td>0·37</td>
</tr>
<tr>
<td>Ca</td>
<td>63</td>
<td>0·20 to 0·89</td>
<td>0·55</td>
<td>0·56</td>
</tr>
<tr>
<td>Fe</td>
<td>41</td>
<td>0·08 to 0·83</td>
<td>0·45</td>
<td>0·47</td>
</tr>
<tr>
<td>Fruit</td>
<td>16</td>
<td>−0·01 to 0·71</td>
<td>0·49</td>
<td>0·51</td>
</tr>
<tr>
<td>Vegetables</td>
<td>17</td>
<td>0·16 to 0·72</td>
<td>0·39</td>
<td>0·38</td>
</tr>
</tbody>
</table>

### Table 3. Mean correlation coefficients for food-frequency questionnaire compared against different reference methods

<table>
<thead>
<tr>
<th>Nutrient</th>
<th>Weighed record (n)</th>
<th>Unweighed record (n)</th>
<th>24 h recalls (n)</th>
<th>Biomarkers (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy</td>
<td>0·47 (29)</td>
<td>0·47 (29)</td>
<td>0·48 (23)</td>
<td>0·51 (14)</td>
</tr>
<tr>
<td>Fat</td>
<td>0·51 (36)</td>
<td>0·52 (22)</td>
<td>0·49 (22)</td>
<td>0·54 (16)</td>
</tr>
<tr>
<td>Vitamin C</td>
<td>0·50 (27)</td>
<td>0·46 (18)</td>
<td>0·41 (20)</td>
<td>0·46 (15)</td>
</tr>
<tr>
<td>Vitamin A</td>
<td>0·39 (23)</td>
<td>0·38 (13)</td>
<td>0·39 (16)</td>
<td>0·35 (13)</td>
</tr>
<tr>
<td>Ca</td>
<td>0·54 (23)</td>
<td>0·53 (15)</td>
<td>0·54 (22)</td>
<td>0·53 (8)</td>
</tr>
<tr>
<td>Fe</td>
<td>0·51 (16)</td>
<td>0·41 (10)</td>
<td>0·43 (15)</td>
<td>0·51 (8)</td>
</tr>
</tbody>
</table>
energy (0.42 v. 0.38) compared with the studies that only included women. However, somewhat lower mean correlations were seen for men v. women for vitamin C (0.45 v. 0.51) and vitamin A (0.29 v. 0.43). These results imply a possible differential misclassification for men and women using an FFQ compared with another reference method. (The results of numbers were too small to break this down further by type of reference method used.)

The length of the reference period for the FFQ (for current-use questionnaires) was explored. FFQ asking for recall over the previous month had slightly higher correlations with the reference method than those recalling over the previous year. A small number of questionnaires were designed to measure nutrient intake in the past. Correlations with a reference method were similar for these questionnaires compared with FFQ designed to assess current intake. The studies which have used FFQ to assess diet over the previous month had slightly higher correlation coefficients than those with the smallest number of items. (For example, the top quartile of items had somewhat higher correlation coefficients than those with the smallest number of items. For example, the top quartile of items v. the lowest quartile of items: fat 0.53 v. 0.47; vitamin C 0.50 v. 0.46.) Although a number of the shorter questionnaires were assessing single nutrients such as Ca, many of the longer questionnaires were also assessing a single nutrient. For example, questionnaires with eight items (Angrrett & Moller, 1999), nine items (Nelson et al. 1988), fifteen items (Blalock et al. 1993), forty or ninety-eight items (Brown & Griebler, 1993) have all been designed to assess Ca intake. The study by Brown & Griebler (1993) looked at two versions of a modified Block questionnaire and found that the short (forty-item) FFQ provided reliable estimates of usual Ca intake in milligrams and captured more than 75 % of the Ca intake measured by the long (ninety-eight-item) FFQ. The shortest questionnaire had five items relating to fruit and vegetable intake (Kristal et al. 1998) and the longest questionnaire had 350 items plus a booklet with 120 photographs to estimate portion sizes (Riboli et al. 1997a). This underlines the fact that FFQ are all different.

How foods are organised on questionnaires may have an impact on the ease of completion for subjects. One study found that subjects preferred to be asked about food intake using a questionnaire arranged by meal rather than broad food group (Boutron et al. 1989). However, foods tend to be organised on FFQ in terms of food groups.

Correlation coefficients were compared for newly designed FFQ and those adapted from other questionnaires. Newly developed questionnaires had a higher correlation for energy (0.49 v. 0.44) and fat (0.52 v. 0.49) than modified questionnaires. Adapted questionnaires had a higher correlation for vitamin C (0.50 v. 0.44) and vitamin A (0.41 v. 0.34) than newly developed questionnaires. There were no differences (new v. adapted) for Ca (0.54 v. 0.55) and Fe (0.45 v. 0.44).

**Repeatability studies**

The issue of whether an FFQ can produce repeatable or reproducible results is important for all types of study design. Repeatability is generally assessed by administering the same FFQ twice to the same group of subjects and analysing the association between the two responses.

The repeatability of dietary intake was assessed in 107 of the validation studies (45 %), the most common time interval between repeat measures being 1 year. One study looked at the long-term reproducibility of diet, with repeat measures 15 years apart (Thompson et al. 1990). The shortest time interval for repeating the questionnaire was 2 h (Frank et al. 1992). In 34 % the repeat administration was between 1 and 6 months later. Correlation coefficients were used to compare the differences between repeated measures in 90 % of studies. Less than 10 % used the Bland–Altman method. The difference in absolute intakes was assessed in 39 % of studies.

In general, correlation coefficients were higher in repeatability studies when subjects were allowed to specify their own portion sizes. The repeatability of commonly consumed foods tends to be better than that for less commonly consumed items (Wiecha et al. 1994). The season of repeating the questionnaire appears to have little impact on the results (Hartman et al. 1996).

When the FFQ is repeated in relation to disease outcome, this can affect the results. It is not possible to distinguish real differences in reported intakes from differences in reporting of intakes and so apparent associations that may

<table>
<thead>
<tr>
<th>Nutrient</th>
<th>Describe own portion size (n)</th>
<th>No portion size specified (n)</th>
<th>Portion size specified (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy</td>
<td>0.54 (31)</td>
<td>0.44 (9)</td>
<td>0.42 (31)</td>
</tr>
<tr>
<td>Fat</td>
<td>0.57 (35)</td>
<td>0.42 (12)</td>
<td>0.47 (31)</td>
</tr>
<tr>
<td>Vitamin C</td>
<td>0.53 (28)</td>
<td>0.45 (10)</td>
<td>0.46 (29)</td>
</tr>
<tr>
<td>Vitamin A</td>
<td>0.46 (23)</td>
<td>0.21 (5)</td>
<td>0.37 (26)</td>
</tr>
<tr>
<td>Ca</td>
<td>0.61 (23)</td>
<td>0.49 (8)</td>
<td>0.53 (28)</td>
</tr>
<tr>
<td>Fe</td>
<td>0.53 (14)</td>
<td>0.38 (5)</td>
<td>0.40 (20)</td>
</tr>
</tbody>
</table>
be found in a case-control study cannot necessarily be attributed to real dietary differences (Bunn 1993). One study explored fat intake in relation to the risk of breast cancer using repeated dietary questionnaires administered both before and after the diagnosis of breast cancer. A baseline FFQ was completed by all subjects. Breast cancer cases and age-matched controls were sent another FFQ in 1989 inquiring about their diet in 1985. Age-adjusted analysis using the prospective (1986) questionnaire demonstrated no association between breast cancer incidence and intakes of total fat (odds ratio (OR) between the highest and lowest quintiles of 0·87; 95 % CI 0·54, 1·40) and saturated fat (OR 0·97; CI 0·64, 1·46). Age-adjusted analysis using the retrospective (1989) questionnaire suggested positive associations between breast cancer incidence and intakes of total fat (OR 1·43; 95 % CI 0·90, 2·27) and saturated fat (OR 1·38; 95 % CI 0·89, 2·13). Adjustment for total energy intake gave similar results. These findings suggest that case-control studies of diet and breast cancer using retrospective FFQ may yield biased associations between fat intake and the risk of breast cancer (Giovannucci et al. 1993).

Correlation coefficients between two administrations of an FFQ of 0·5 to 0·7 are common. Repeat administrations of the FFQ at 1 month or less tended to give higher correlation coefficients than repeat administrations further apart (Table 5). Random error assessed by repeat administrations of an FFQ seems to only moderately attenuate observed diet–disease relationships; i.e., an observed OR of 1·5 and a correlation coefficient of 0·70 yield an unattenuated OR of 2·1 (Bueno de Mesquita et al. 1992).

The repeatability of nutrient intakes estimated using semi-quantitative FFQ has been shown to be high in the elderly. Older subjects may be more established in their dietary habits than younger subjects, so any tendency for repeatability to decrease due to impaired memory associated with advanced age is offset by a lower intra-individual variability in dietary habit (Nelson et al. 1988; Lazarus et al. 1995). Individuals with a stable diet, those with a vegetarian diet, and those with more education, are able to recall their diet–disease relationships; i.e., an observed OR of 1·5 and a correlation coefficient of 0·70 yield an unattenuated OR of 2·1 (Bueno de Mesquita et al. 1992).

The repeatability of nutrient intakes estimated using FFQ has been shown to be high in the elderly. Old subjects may be more established in their dietary habits than younger subjects, so any tendency for repeatability to decrease due to impaired memory associated with advanced age is offset by a lower intra-individual variability in dietary habit (Nelson et al. 1988; Lazarus et al. 1995). Individuals with a stable diet, those with a vegetarian diet, and those with more education, are able to recall their past dietary practices with reasonable reliability (Kuzma & Lindsted, 1990; Stevens et al. 1996). Repeatability studies in minority population groups were in the range usually reported in evaluations of the performance of questionnaires in non-minority populations (0·5–0·8) (Shea et al. 1991; Coates & Monteith, 1997). However, some studies in the USA have found higher correlation coefficients between repeated FFQ for white Americans compared with African-Americans (Stevens et al. 1996).

Most repeatability studies have used subjects who are reporting their own intakes. One study showed good test–retest reliability for parents reporting on behalf of their preschool children (Treiber et al. 1990). Repeated FFQ have been obtained from proxy reporters of diet for the second occasion. In one study, husbands were questioned about their deceased wife’s food intake with limited success (Hislop et al. 1992). The repeatability of dietary patterns rather than simply nutrients has been assessed in a study using a sub-sample of the Health Professionals Follow-up Study (Hu et al. 1999). Their findings indicated a reasonable reproducibility of the major dietary patterns defined by factor analysis with data from their FFQ.

In nutrition-intervention research, it is important to consider the sensitivity of dietary assessment instruments to the changes in nutrient intake or dietary behaviour under study. This aspect of repeatability has been called ‘responsiveness’ and is an index of an instrument’s sensitivity to change. Illustrations of this measure from two randomised dietary-intervention trials that targeted reductions in fat intake, the Women’s Health Trial and the Eating Patterns Study, suggest that short, inexpensive measures such as FFQ or questionnaires that assess dietary habits can be as responsive as multiple-day diet records (Kristal et al. 1994). It is suggested that intervention studies should include at least two types of dietary assessment tools and the reliability and responsiveness of these tools should be reported as part of the study outcome (Kristol et al. 1994).

### Table 5. Correlation coefficient by time interval between repeat administrations of the food-frequency questionnaire

<table>
<thead>
<tr>
<th>Nutrient</th>
<th>1 month or less</th>
<th>1 to 6 months</th>
<th>6 months to 1 year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy</td>
<td>0·68</td>
<td>0·67</td>
<td>0·60</td>
</tr>
<tr>
<td>Fat</td>
<td>0·68</td>
<td>0·63</td>
<td>0·57</td>
</tr>
<tr>
<td>Vitamin C</td>
<td>0·60</td>
<td>0·64</td>
<td>0·54</td>
</tr>
<tr>
<td>Vitamin A</td>
<td>0·60</td>
<td>0·64</td>
<td>0·54</td>
</tr>
<tr>
<td>Ca</td>
<td>0·67</td>
<td>0·70</td>
<td>0·55</td>
</tr>
</tbody>
</table>

A total of 179 publications were identified which had used FFQ and were published in 1998. These were 156 single-study reports and eight groups of studies, making a total of 164 studies. Of these studies, forty-eight had been adapted from a previous FFQ, including nine from Block and ten from Willett questionnaires. Sixty-one of the studies were carried out in the USA, twenty-five in the UK, thirty-one from the rest of Europe and six from Australia and New Zealand.

The FFQ used in these studies were similar in design to those used in the validation studies. The mean number of items on the FFQ was ninety-nine. In eighty-eight of the studies, the FFQ was designed to assess current diet, most commonly over 1 year. Ninety-nine of the studies used self-administered FFQ. The FFQ were most commonly used in cross-sectional surveys (eighty-three studies).

Associated validation papers were mentioned in ninety-one of the 164 studies. Of these, eighteen referred to the Willett FFQ as the basis for the questionnaire and twenty referred to the Block–National Cancer Institute FFQ. Results were rarely modified as a result of a validation study. For example, FFQ data were cross-checked against food-diary data. Validity was accepted if correlation coefficients were > 0·4 (Samaras et al. 1998).

The presentation of results sometimes included the adjustment of nutrients for total energy intakes using regression techniques (Attili et al. 1998; Curhan et al. 1998; Franceschi et al. 1998; Giovannucci et al. 1998, Vahon et al. 1998). These methods allow comparison of nutrients to be made between groups independent of the effect of different energy intakes. The ability to carry this out presupposes that the questionnaire is comprehensive.
enough to determine energy intake. Under-reporting, particularly of energy, can be a problem in dietary assessment studies. Energy adjustment appears to minimise the bias generated by under-reporting with respect to particular nutrients and their association with various disease outcomes (Gnardellis et al. 1998).

Of the FFQ, ninety-three were stated to be disease-specific, and thirty-six of them were designed to assess risks associated with cancers. A list of the diseases or conditions which FFQ were used to assess and associated references are shown in Table 6. Their uses have ranged from assessing diet in studies of various types of cancer to diet and cognitive function. A qualitative assessment of each tool has not been made. In order to assess quality there are a number of aspects which need to be considered, including whether there is an associated validation paper and whether the tool has been validated in the population on which it is being used. Even if a previous study has validated the FFQ, it should not be assumed that the tool will be applicable in all situations. An in-depth exploration of this issue is outside the scope of the present review. Nevertheless, for example, for the eight papers relating to breast cancer only three had reported previous published validation studies, which were not necessarily undertaken on a similar population to which the tool was currently being applied.

Conclusion
The FFQ is a useful tool for assessing diet in a number of different types of study. Specific design and validation issues have been highlighted by the present review. If these issues are taken into account in future research, this should improve the quality of the data collected. It is important to remember that there is no standard FFQ. Each questionnaire should be judged for its ability to provide the information for which it was intended. The consensus document has made recommendations (see Appendix) which should be considered when planning and analysing studies utilising an FFQ technique in order to optimise its performance.

Acknowledgements
The study was supported by the Ministry of Agriculture, Fisheries and Food (MAFF project no. AN0850).

Table 6. Diseases and references for disease-specific food-frequency questionnaires (FFQ)

<table>
<thead>
<tr>
<th>Which diseases are FFQ designed for?</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cancer (general)</td>
<td>Ajani et al. (1998); Caderni et al. (1998); Chatenoud et al. (1998); Gnardellis et al. (1998); Grievink et al. (1998); Heerstrass et al. (1998); Jorga et al. (1998); Talamini et al. (1998); Voorrips et al. (1998); Voss et al. (1998)</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>Ambrosone et al. (1998); Cade et al. (1998); De Stefani et al. (1998a); Hebert et al. (1998b); Potschman et al. (1998b); Rohan et al. (1998); Thorand et al. (1998); Zhang et al. (1998)</td>
</tr>
<tr>
<td>Colon and colorectal cancer</td>
<td>Baron et al. (1998); Braga et al. (1998); Chen et al. (1998); De Stefani et al. (1998e); Franceschi et al. (1997, 1998); Gianz et al. (1998); Hyman et al. (1998); Keku et al. (1998); Negri et al. (1998); Singh &amp; Fraser (1998); Slattery et al. (1998)</td>
</tr>
<tr>
<td>Lung cancer</td>
<td>De Stefani et al. (1998b); Garcia-Closas et al. (1998); Jatoi et al. (1998); Sinha et al. (1998)</td>
</tr>
<tr>
<td>Oesophageal cancer, nasopharyngeal cancer</td>
<td>Brown et al. (1998); Farrow et al. (1998); De Stefani et al. (1999)</td>
</tr>
<tr>
<td>Oral leukoplakia</td>
<td>Gupta et al. (1998)</td>
</tr>
<tr>
<td>Ovarian cancer</td>
<td>Webb et al. (1998)</td>
</tr>
<tr>
<td>Cervical dysplasia</td>
<td>Kantesky et al. (1998); Wideroff et al. (1998)</td>
</tr>
<tr>
<td>Prostate cancer</td>
<td>Bairati et al. (1998); Giovannucci et al. (1998); Hartman et al. (1998); Schuurman et al. (1998)</td>
</tr>
<tr>
<td>Renal cell cancer</td>
<td>De Stefani et al. (1998d); Yuan et al. (1998)</td>
</tr>
<tr>
<td>Stomach cancer</td>
<td>Galanis et al. (1998); Ji et al. (1998)</td>
</tr>
<tr>
<td>Non-Hodgkin’s lymphoma</td>
<td>De Stefani et al. (1998c)</td>
</tr>
<tr>
<td>CHD</td>
<td>Almendingen et al. (1998); Hu et al. (1998); Kitamura et al. (1998); Swinburn et al. (1998); Thorand et al. (1998); Tzonou et al. (1998)</td>
</tr>
<tr>
<td>Blood pressure</td>
<td>Dwyer et al. (1998)</td>
</tr>
<tr>
<td>Cardiovascular disease</td>
<td>Cox &amp; Whichelow (1998)</td>
</tr>
<tr>
<td>Hypercholesterolaemia</td>
<td>Romero et al. (1998)</td>
</tr>
</tbody>
</table>
Table 6. Continued

<table>
<thead>
<tr>
<th>Which diseases are FFQ designed for?</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bone health</td>
<td>Melhus et al. (1998); New et al. (1998b)</td>
</tr>
<tr>
<td>Osteoporosis</td>
<td>Patton et al. (1998)</td>
</tr>
<tr>
<td>Renal stones</td>
<td>Curhan et al. (1998); Hirvonen et al. (1998); Leonetti et al. (1998); Sowers et al. (1998)</td>
</tr>
<tr>
<td>Appendicitis</td>
<td>Naaeder &amp; Archampong (1998)</td>
</tr>
<tr>
<td>Cerebral palsy</td>
<td>Petridou et al. (1998)</td>
</tr>
<tr>
<td>Cognitive function</td>
<td>Munger et al. (1998); Paleologos et al. (1998)</td>
</tr>
<tr>
<td>Type 2 diabetes, diabetic eye disease, insulin resistance</td>
<td>Kao et al. (1998); Liu et al. (1998); Millen et al. (1998)</td>
</tr>
<tr>
<td>Drug use</td>
<td>Himmelgreen et al. (1998)</td>
</tr>
<tr>
<td>Dental caries</td>
<td>Arnadottir et al. (1998)</td>
</tr>
<tr>
<td>Gallstones</td>
<td>Attili et al. (1998)</td>
</tr>
<tr>
<td>Helicobacter pylori</td>
<td>Malaty et al. (1998)</td>
</tr>
<tr>
<td>Hansen’s disease</td>
<td>Oh et al. (1998)</td>
</tr>
<tr>
<td>Hyperemesis</td>
<td>Signorello et al. (1998)</td>
</tr>
<tr>
<td>Fe status</td>
<td>Al-Othman et al. (1998)</td>
</tr>
<tr>
<td>Low birth weight</td>
<td>Santos et al. (1998)</td>
</tr>
<tr>
<td>Maculopathy</td>
<td>Shoff et al. (1998)</td>
</tr>
<tr>
<td>Menopause</td>
<td>Nagata et al. (1998)</td>
</tr>
<tr>
<td>Multiple sclerosis</td>
<td>Ghadarian et al. (1998)</td>
</tr>
<tr>
<td>Neural-tube defects</td>
<td>Shaw et al. (1998)</td>
</tr>
<tr>
<td>Respiratory disease</td>
<td>Fluge et al. (1998)</td>
</tr>
<tr>
<td>Rheumatic fever</td>
<td>Zaman et al. (1998)</td>
</tr>
<tr>
<td>Rheumatoid arthritis</td>
<td>New et al. (1998a); Volker et al. (1998)</td>
</tr>
<tr>
<td>Ventilatory function</td>
<td>Carey et al. (1998)</td>
</tr>
</tbody>
</table>

References


Sawaya AL, Tucker K, Tsay R, Willett W, Saltzman E, Dallas GE & Roberts SB (1996) Evaluation of four methods for determining energy intake in young and older women: comparison with...


Appendix

These are the recommendations of the consensus document (for more detail, see Cade et al. 2002).

Recommendations on food-frequency questionnaire development

Questionnaire design

Consider the following points before designing an FFQ:

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

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.

FFQ 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.

Modifying existing questionnaires

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

Developing the food list

Unless the purpose of the FFQ is very specific a comprehensive food list is desirable.

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 single foods may need to be considered to prevent excessive questionnaire length.

Closed and open questions

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

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 and large options, and estimation of portion size using photographs.

Method of administration

If practical, interviewer-administered FFQ should be used in preference to self-administered questionnaires.

If self-administered versions are used then checking the questionnaires for completeness should be undertaken.

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.
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 (and percentage) of questionnaires rejected.

Recommendations on tests for food-frequency questionnaire 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 FFQ. 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. Use of the correlation coefficient for the assessment of food-frequency questionnaire 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.

Use of the Bland–Altman method for the assessment of reproducibility

The methods developed by Bland and Altman should be used to assess reproducibility and 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.

Other methods for the assessment of food-frequency questionnaire 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

FFQ 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 FFQ. 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.

Sample or population selection in validation studies

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

Time frame of reference method in validation studies

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

Selection of dietary method for validation studies

Both 24 h recalls and unweighed and weighed records are suitable for a reference method. Multiple days of collection of dietary data should be undertaken.

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 the relationship between biological variation and variation in dietary intake.

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 FFQ 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 is ordered, categorical or binary.

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.

Sample size for validation studies

Expert statistical advice should be sought to estimate the 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 fifty to 100 subjects for each demographic group is recommended.

Sequence of administration

The test instrument should be administered before the assessment of the reference measure.
**Recommendations on description of food-frequency questionnaires in utilisation studies**

- State the original objective of the questionnaire.
- If the questionnaire has previously been used, give a reference.
- Describe the questionnaire in terms of the number of food items and 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.

**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 or nutrients of interest (the results should reflect the purpose of the questionnaire).
- Report any adjustments to the data (for example, use of cross-check questions).
- If there has not been a validation study, give the reason.