

## Evaluating the quality of dietary intake validation studies

Lluís Serra-Majem<sup>1,2\*</sup>, Lene Frost Andersen<sup>3</sup>, Patricia Henríque-Sánchez<sup>1</sup>, Jorge Doreste-Alonso<sup>1</sup>, Almudena Sánchez-Villegas<sup>1</sup>, Adriana Ortiz-Andrelluchi<sup>1</sup>, Eva Negri<sup>4</sup> and Carlo La Vecchia<sup>4,5</sup>

<sup>1</sup>Nutrition Research Group, Department of Clinical Sciences, Centre for Health Sciences, University of Las Palmas de Gran Canaria, PO Box 550, 35080 Las Palmas de Gran Canaria, Spain

<sup>2</sup>Community Nutrition Research Centre of the Nutrition Research Foundation, University of Barcelona Science Park, Baldri Reixac 4, 08028 Barcelona, Spain

<sup>3</sup>Department of Nutrition, Institute of Basic Medical Sciences, University of Oslo, PO Box 1046 Blindern, 0316 Oslo, Norway

<sup>4</sup>Istituto di Ricerche Farmacologiche 'Mario Negri', Via La Masa 19, I-20156 Milan, Italy

<sup>5</sup>Istituto di Statistica Medica e Biometria, 'G. A. Maccacaro', University of Milan, Milan, Italy

(Received 15 June 2009 – Revised 2 October 2009 – Accepted 1 November 2009)

Within the EUROpean micronutrient RECommendations Aligned Network of Excellence (EURRECA), a scoring system was developed to assess the quality of dietary intake validation studies. The scoring system included three steps. The first step was to give each study a quality score, which included five components: sample size, statistics used, data collection procedure, consideration of seasonality and supplement use. Scores ranged from 0 to 7, and validation studies classified as very good ( $\geq 5$ ), good (5–3.5), acceptable/reasonable (3.5–2.5) and poor ( $< 2.5$ ). The second and third steps included an adjustment/weighting of the correlation coefficient according to the quality score and moreover a rating of the adjusted/weighted correlation. The scoring system was tested in 124 validation studies that included at least one vitamin. Only 5.6% of the 124 studies were judged to be of very good quality according to the quality score, 41.9% of good quality and 16.9% had a poor rating. When adjusting for the study quality scores, crude and adjusted mean correlations of vitamins A, C, D and E intakes were similar, but the percentage of correlation values classified as poor or very good was higher after adjustment. These results show the importance of considering the quality of studies validating dietary assessment methods and the correlations obtained for the micronutrient of interest when interpreting effects observed in epidemiological studies using dietary assessment methods. Without a doubt, this subject constitutes a key topic for research in nutritional epidemiology.

### Validation studies: Nutrient intake: Quality: Score

The evaluation of a population's micronutrient consumption constitutes a true challenge for nutrition research and, in general, has received less attention than the consumption of energy and macronutrients. Food and nutrient intakes are estimated through the administration of dietary assessment methods that usually differ according to study objectives, available resources, the population under study and the design of the epidemiological study<sup>(1–3)</sup>. As such, in cross-sectional studies with the aim of evaluating food consumption and nutritional status for a given population or group of individuals, traditionally daily methods have been utilised, such as a single or multiple 24 h recalls (24HR) or food records. In contrast, epidemiological studies have mostly employed diverse variations of a Food Frequency Questionnaire (FFQ) that has been validated typically by daily methods or biochemical indicators.

Daily methods like 24HR and food records, which are less often validated than FFQ, have both advantages and disadvantages that have been amply described, and given their open-ended nature, by and large their intrinsic value has

been accepted if they are administered in adequate conditions<sup>(2,3)</sup>. However, their utility has been challenged due to the increasing knowledge of the problem with misreporting in these methods, which especially affect the selective underreporting of energy and certain food groups among given population groups<sup>(4)</sup>. On the other hand, FFQ used in epidemiological studies have been validated or compared to other methods. Hence, notable validation studies of FFQ can be found in the literature, initiated in a rigorous and thorough manner by the groups of Walter Willett at Harvard University<sup>(5,6)</sup>; and of Rohan and Potter at the University of South Australia<sup>(7)</sup>, among others.

Generally speaking, FFQ have usually been validated for the consumption of energy, macronutrients, fatty acids, fibre and a given micronutrient. In fact, the first validation studies of these types of questionnaires addressed only a few vitamins and minerals such as vitamin C, vitamin A, Ca and Fe<sup>(1,5–7)</sup>. For the most part, correlations for micronutrients were usually lower than those for macronutrients, especially for those

On behalf of EURRECA's RA 1.1 'Intake Methods' members: Serra-Majem L (Coordinator), Cavelaars A, Dhonukshe-Rutten R, Doreste JL, Frost Andersen L, García-Álvarez A, Glibetic M, De Groot L, De Vries J, Gurinovic M, Henríquez-Sánchez P, Naska A, Ngo J, Novakovic R, Ortiz-Andrelluchi A, Ørverby NC, Pijls L, Ranic M, Ribas-Barba L, Ristic-Medic D, Román-Viñas B, Ruprich J, Saavedra-Santana P, Sánchez-Villegas A, Tabacci G, Tepsic J, Trichopoulou A, van't Veer P, Vucic V, Wijnhoven TMA.

\* **Corresponding author:** Lluís Serra-Majem, fax +34 928 453475, email lserra@dcc.ulpgc.es

micronutrients whose intake level depended on the consumption of a great number of foods<sup>(1,5,6)</sup>. Since the initial validation studies conducted in the 1980s to the present, hundreds of these types of studies have been published and, despite the fact that consensus-based criteria for how they should be administered are lacking, in most cases, the experience of Willett & Lenart<sup>(1)</sup> and the Nurse's Health Study has been followed.

Within the context of the EUROpean micronutrient RECOmmendations Aligned Network of Excellence<sup>(8)</sup> (EURRECA), we have conducted an in-depth review of all validation studies<sup>(9–14)</sup> with the aim of analysing the utility of distinct dietary questionnaires, and particularly FFQ, for evaluating micronutrient intake. To achieve this, systematic reviews of all published validation articles that evaluated micronutrients or *n*-3 fatty acids were realised so as to assess the degree of validity of these questionnaires. In general, the questionnaires under study evaluated a set of various micronutrients, and occasionally, the focus was to assess a specific micronutrient such as Ca or vitamin A. Given the EURRECA project's objective of subsequently conducting distinct meta-analyses for each of the micronutrients, it was deemed necessary to make a tool available that could be used to estimate the quality of validation studies that evaluated dietary questionnaires. To our knowledge, the development of such a quality criteria tool has not been carried out to date.

The objective of the present study was to provide a tool for evaluating the quality of studies validating FFQ and to explore its applicability in a sample of validation studies. This would allow for weighting-observed correlations with obtained quality estimations when combining a given number of studies for the evaluation of a specific micronutrient in different population groups<sup>(9–14)</sup>.

### Material and methods

In order to identify the most accurate method for assessing the different micronutrient and *n*-3 intakes in the adult population and among specific population groups, a scoring system was needed to assess the quality of the different validation studies.

Based on tasks for the present research activity on intake methods, which consisted of conducting systematic reviews of validation papers, for all reviews, the papers were classified into three different categories, according to the reference method or gold standard applied in the validation study:

- (1) The reference method was another dietary assessment method evaluating short-term intake, including both 24HR, estimated and weighed records of less than 7 d.
- (2) The reference method was another dietary assessment method of long-term intake, including more than 7 d of dietary collection and
- (3) The reference method was a biomarker. In this case, a discussion on the selected/available biomarker and its characteristics as a recovery or concentration biomarker was needed.

The studies included in categories (1) and (2) may be called calibration and correlation rather than validation studies, since the errors of both dietary assessment methods could be correlated or might not be independent. Sources of errors

tend to be replicated in both dietary assessment methods: the one being evaluated and the gold standard<sup>(1,2)</sup>. The use of biomarkers as the reference method is usually not feasible since there is a lack of markers for many micronutrients and usually they do not reflect a pure estimation of only diet.

After an initial meeting in Prague on 15–16 May 2008, where a draft list of variables to be included in the scoring system was discussed, a working group composed by the authors of the present article made decisions about the different variables and score values to be included in the tool, using a consensus-based methodology. The proposed tool was discussed at the EURRECA Integrating Meeting in Montenegro (9–13 June 2008) and also at the Steering Committee in Milano (29–30 September 2008). The final version was ready by October 2008. The variables considered were:

- (1) *Sample and sample size* of the study, with a maximum of 1 point; 0.5 points allocated when the sample was not homogeneous for certain variables such as sex, socio-economic status, smoking and obesity, and 0.5 points when the sample size was of more than 100 individuals (fifty individuals when using biomarkers as the gold standard).
- (2) *Statistics* to assess validity. A maximum of 3 points was allocated; 1 for comparisons between methods' means, medians or difference; from 0.5 to 1.5 according to the correlation used (crude, energy adjusted, deattenuated or intraclass); plus 0.5 when statistics to assess agreement or misclassification were utilised.
- (3) *Data collection*. 1 point if data were gathered by personal interview.
- (4) *Seasonality*. Only when considered in the validation design. Addition of 0.5 points.
- (5) *Supplements* included and validated. Addition of 1.5 points when the validation study considered supplements.

Scores could range from 0 (poorest quality) to a maximum of 7 (highest quality). This allowed for the classification of validation studies according to their methodological quality (Table 1 step 1):

- (1) Very good/excellent, score  $\geq 5.0$ .
- (2) Good,  $3.5 \leq \text{score} < 5$ .
- (3) Acceptable/reasonable,  $2.5 \leq \text{score} < 3.5$ .
- (4) Poor, score  $< 2.5$ .

For the simultaneous analysis of several validation papers with the aim of estimating a mean combined correlation of various studies per micronutrient for a given diet assessment method, the correlation coefficient value of each study was multiplied by its quality score. Then, the sum of the weighted correlations was divided by the sum of the validation studies' quality scores. This provided us with a correlation coefficient adjusted for the study's methodological quality (see Table 1 step 2).

In order to analyse the implications of using the scoring system, a systematic review of the validation studies addressing the intake of at least one vitamin was performed, and the average of correlations was calculated without (crude) and with (adjusted) the scoring system. As explained in the paper from Henríquez-Sánchez *et al.*<sup>(9)</sup>, the following number of studies addressing each vitamin was included: 76 articles

**Table 1.** Scoring system to assess the quality of the nutrient intake validation studies

Variables in general	Specific variable	Points	Score
Step 1 Quality of the validation study			
Score the studies according to:			
1. Sample and sample size	Non-homogeneous sample (sex, SES, smoking and obesity) $n > 100$ ( $n > 50$ for biomarkers)	0.5	
2. Statistics:		0.5	
Group level	Compare/test mean or median or difference	1	
Correlations: (select only one of the three types, choosing type with the highest score)	Correlation	0.5	
	Adjusted correlations (energy)	1	
	Deattenuated or intraclass correlations	1.5	
Agreement	Classification or Bland & Altman plot	0.5	
3. Data collection	Gathered by face to face interview	1	
4. Seasonality	Considered	0.5	
5. Supplements	Included and data considered in analysis	1.5	
		Total score (S)	= $\Sigma$

Total score 'S' may range from a minimum of 0 (for all studies) to a maximum of 7

#### Step 2 Correlation reporting

Weight the correlation coefficient of the study by the score obtained in step 1 (quality of the validation study) and calculate the average for all the studies for the nutrient under study:

$$X_n = \frac{\sum(R_n \cdot S_n)}{\sum S_n}$$

Where,  $R_n$  = correlation coefficient;  $S_n$  = quality score of validation study;  $X_n$  = weighted correlation average per reference method category

#### Step 3 Correlation rating

Classify each weighted correlation average ( $X_n$ ) according to the following:

$X_n < 0.3$	Poor
$0.3 \leq X_n < 0.5$	Acceptable
$0.5 \leq X_n < 0.7$	Reasonable/good
$X_n \geq 0.7$	Very good

SES, Socioeconomic status.

for vitamin A, 108 for vitamin C, 21 for vitamin D, 75 for vitamin E, 47 for folic acid, 19 for vitamin B<sub>12</sub>, 21 for vitamin B<sub>6</sub>, 49 for thiamine, 49 for riboflavin and 30 for niacin, extracted from a total of 124 different studies. The correlation values obtained for every vitamin in each of the studies were summed up, first without consideration of the quality of the study or scoring system, and second taking this factor into account. Subsequently, the proportion of correlations for vitamins A, C, D and E were classified into four categories: very good ( $\geq 0.7$ ), good (0.5–0.69), acceptable (0.3–0.49) and poor ( $< 0.3$ ; Table 1 step 3), and comparisons between crude and adjusted figures were made using a McNemar  $\chi^2$  test in SPSS-PC<sup>(15)</sup>.

## Results

Table 1 presents the scoring system administered to evaluate the quality of the validation studies. Among the 124 validation studies analysed, the quality scores ranged from 0.5 to 6 (highest score obtained by Sudha *et al.*<sup>(16)</sup>). The average was 3.19 and median 3.0 (Table 2). Table 2 shows the classification of the validation studies, with 47.5% of the studies rating as good (41.9%) or very good (5.6%) quality, and 16.9% having a poor quality rating. Table 3 shows the mean and classification of crude and quality-adjusted correlation coefficients obtained for vitamins A, C, D and E. Although no significant differences were found, correlation averages were similar, and the percentages of correlations

below 0.3 (poor) were considerably higher after adjusting for the quality of the validation study. Additionally and in contrast, for vitamins A and D, the percentages of correlations  $\geq 0.7$  (very good) were substantially higher after adjustment for quality.

## Discussion

It is quite surprising that to date a systematic review has not been conducted, which evaluates validation studies of FFQ. Over the last few decades, nutritional epidemiology has developed and increased in an exponential fashion, due in part to the innovation of simplified, properly validated questionnaires for evaluating the consumption of food and nutrients. The development of these types of instruments accommodates the inherent needs of epidemiological studies in which larger population samples make it a necessity to utilise data collection methods that are valid, accessible, reproducible and less costly. In general, the validation of FFQ has been achieved by comparing with various records/registers, which adequately distributed over a long period of time, typically 1 year, reflect usual intake. Multiple 24HR and biomarkers have also been utilised.

When developing the quality scoring system, five principal variables were taken into account: sample and sample size, statistics, data collection, seasonality and supplements. Sample size is an aspect that has been given minimal

**Table 2.** Quality score of the validation studies of FFQ including any vitamin

Quality score	Studies of FFQ v. dietary record, long term (n 53)		Studies of FFQ v. dietary record, short term (n 20)		Studies of FFQ v. 24HR, long term (n 19)		Studies of FFQ v. 24HR, short term (n 19)		Studies of FFQ v. bio-markers (n 33)		Total studies (n 124)	
	n	%	n	%	n	%	n	%	n	%	n	%
Very good ( $\geq 5.00$ )	5	9.4	0	0.0	1	5.3	1	5.3	1	3.0	7	5.6
Good ( $3.5 \leq \text{score} < 5$ )	22	41.5	6	30.0	8	42.1	7	36.8	15	45.5	52	41.9
Acceptable ( $2.5 \leq \text{score} < 3.5$ )	15	28.3	7	35.0	8	42.1	9	47.4	12	36.4	44	35.5
Poor ( $< 2.5$ )	11	20.8	7	35.0	2	10.5	2	10.5	5	15.2	21	16.9
Mean												3.19
Median												3.00

24HR, 24 h recall.

consideration in validation studies. According to Willett & Lenart<sup>(1)</sup>, selecting an appropriate number of subjects for a validation study is less than straightforward as correlations for many nutrients are likely to be examined and the precision required remains arbitrary. These authors have estimated that a reasonable sample size for a validation study should comprise between 100 and 200 people. We chose the sample of 100 individuals as a study quality criterion when using other dietary assessment methods as a gold standard; when using biochemical markers as the reference methods, a sample size of just fifty individuals could be considered satisfactory, since measurement errors in biomarkers are essentially uncorrelated with errors in any dietary assessment method<sup>(1)</sup>. We added 0.5 when the sample size was of more than 100 people and 0.5 when the sample was not homogeneous. Validating a questionnaire in a homogeneous sample (i.e. very healthy people) could reduce its external validity and usefulness when administering it in other samples (for example, in obese or low-income populations, among others). Other possibilities such as weighting the study for the inverse of the variance could also have been considered, instead of using a specific sample size to assess quality.

Regarding the statistical methods used in epidemiological studies, although it is useful to compare means or medians for the two methods, it is more important to provide data on the associations between the intakes measured by the two methods<sup>(1)</sup>. Correlations are the most applied statistical procedures that should be presented, and to the extent possible, be adjusted for different variables, particularly energy, age and sex. Deattenuating the correlation is critical to reduce its dependency on between-person variation.

The correlation between the questionnaire measurement and the subjects' true long-term intake of micronutrients is then estimated from the observed correlation with mean reference measurements, with correction for attenuating effects due to random errors in the reference measurements themselves<sup>(17)</sup>. For this reason, deattenuated correlations were assigned a higher rating than crude or adjusted correlations in the scoring system. Using biochemical markers may solve the problems of correlation errors between two dietary questionnaires, but only for those micronutrients where a recovery marker is available<sup>(18)</sup>. However, as only a limited number of recovery markers are available, adjustment for energy intake could be an alternative, since it may diminish the problem of correlated random errors. For this reason, energy adjustment and the deattenuating of the correlation are critical points in the statistical methods used in the validation studies of dietary intake. Bland & Altman<sup>(19,20)</sup> recommended not to use the correlation analysis in these types of studies, but rather to analyse the standard deviation of the difference between the two methods, since this is not influenced by between-person variation. Additionally, classification within the same tertiles of consumption may be useful as well. Statistical procedures account for 3 out of 7 points of the quality score as they are critical to a sound analysis. Moreover, since there is no single method to relate a surrogate measure to a measure of truth that conveys all the available information, it is probably best to analyse and present the data in several ways<sup>(21,22)</sup>.

As for data collection procedures, FFQ and food records can be administered by personal interview or by telephone or can be self-administered, using electronic means or postal mailings. In general, the accuracy and quality of data

**Table 3.** Comparison between crude and quality adjusted correlation coefficients obtained for vitamins A, C, D and E

Range	Vitamin A					Vitamin C					Vitamin D					Vitamin E				
	$R_{\text{crude}}$		$R_{\text{adjusted}}$		P	$R_{\text{crude}}$		$R_{\text{adjusted}}$		P	$R_{\text{crude}}$		$R_{\text{adjusted}}$		P	$R_{\text{crude}}$		$R_{\text{adjusted}}$		P
	%	n	%	n		%	n	%	n		%	n	%	n		%	n	%	n	
<0.3	16.3	16	24.5	24	ns	8.8	9	15.7	16	ns	9.5	2	19.0	4	ns	19.7	15	26.3	20	ns
0.3–0.49	46.9	46	42.9	42	ns	40.2	41	38.2	39	ns	33.3	7	33.3	7	ns	40.8	31	35.5	27	ns
0.5–0.69	32.7	32	23.5	23	ns	35.3	36	30.4	31	ns	42.9	9	19.0	4	ns	21.1	16	19.7	15	ns
$\geq 0.7$	4.1	4	9.2	9	ns	15.7	16	15.7	16	ns	14.3	3	28.6	6	ns	18.4	14	18.4	14	ns
Total	100	98	100	98		100	102	100	102		100	19	100	19		100	76	100	76	

collection are greater when using personal interviews, particularly in cross-sectional studies<sup>(23)</sup>. In addition, self-administered questionnaires are generally less reliable in low income populations and those with less education<sup>(24,25)</sup>. Taking the afore-mentioned into account, in our opinion, collecting dietary information for the gold standard or reference method by personal interview, independently if the validated questionnaire is self-administered or not, may provide better comparisons with less possibility of correlating errors among the two estimated methods. A score of 1 was included for those validation studies using information gathered by face to face interview. The qualification of the person administering the interview was a variable not taken into specific consideration for scoring, but is also a factor that impacts on quality data collection i.e. previous training, being a dietitian, or bilingual or cross-cultural interviewer<sup>(26)</sup>.

Seasonality is an important issue in nutrition epidemiology, particularly as it is directly related to the intake of certain vitamins. For this reason, it is important to take it into consideration when validating a questionnaire; usually, authors tend to distribute the food records over different time periods throughout the year<sup>(27)</sup>. As such, we have included seasonal variation as a quality criterion, with the addition of 0.5 points for studies taking this factor into account.

Finally, the inclusion of supplements was considered an important characteristic in validation studies for micronutrients, unless the questionnaire was intentionally designated for non-supplement users only. With the exception of certain vitamins from the B complex group, for the majority of micronutrients, validation study correlations improved when supplement intake was taken into consideration<sup>(28)</sup>, which seems evident for those populations whose intake of supplements is qualitatively and quantitatively important. In fact, in a systematic review on vitamins<sup>(9)</sup>, correlations were higher when supplements were taken into account in the validation studies, especially for vitamins D, E, B<sub>6</sub> and folic acid, but less for vitamin B<sub>12</sub> or vitamin A.

Some authors have developed scoring systems for evaluating dietary methodology (quality) in epidemiological studies<sup>(29)</sup>, and the quality criteria included, among others, the number of FFQ items, the form of administration and the application of feasibility testing or pre-testing. It is often difficult to come across these quality parameters in epidemiological studies, and particularly so in the abstract of the article. Nevertheless, our objective in the present analysis was not so much to judge the quality of nutrition information in epidemiological studies applying *a priori* defined criteria, but rather to assess the quality of validation and calibration studies of FFQ, with the aim of including, excluding or weighting the study or studies that utilise a given questionnaire in possible reviews or meta-analyses for a specific micronutrient. To our knowledge, this factor has not been taken into consideration in systematic reviews or meta-analyses of nutrition epidemiological studies before.

In the present study, we found that less than half of FFQ validation studies, evaluating any given vitamin, were of 'good' or 'very good' quality and that 17% were of 'poor' quality. Quality scores were lower in those validation studies where FFQ was compared to dietary records reflecting short-term intake. When applying the quality score to adjust the correlation obtained in these validation studies for vitamins,

the mean correlations were nearly the same, but the percentage of studies with correlations falling to the lowest category (poor) increased for all vitamins included in this analysis. Moreover, after adjusting for the study quality score, an increase in the percentage of studies whose correlations shifted to the highest category (very good) was seen for vitamins A and D. This information should be considered when interpreting results from studies using FFQ to evaluate vitamin intake.

The impact of measurement errors in dietary assessment instruments in the design, analysis and interpretation of nutrition studies may be much greater than previously estimated, particularly in prospective cohort studies<sup>(30)</sup>. Recently, discussions about FFQ limitations compared to food records or recalls have been pointed out, particularly due to results emerging from the EPIC and other studies<sup>(31–38)</sup>. Without a doubt, this has led to the erroneous utilisation of some FFQ, not necessarily due to the nature of the instrument in and of itself, but rather to its inappropriate application (i.e. FFQ validated in a population or for dietary components different from that in the epidemiological study); albeit, a considerable percentage of these tools may also be of low-quality design. As such, and especially when considering micronutrients, it is critical to utilise high-quality instruments in which we have obtained an adequate estimation of intake for a given micronutrient (drawn from the correlation coefficient of the validation study). The issue should not be limited to the validation of a questionnaire, but rather it should address the utility of the instrument for measuring intakes of a given micronutrient and whether the design and analysis are adequate<sup>(39)</sup>.

As pointed out by Block & Hartman<sup>(40)</sup>, negative outcomes may often be the result of poor instruments and faulty data impede the progress of research, which may all be particularly relevant when studying micronutrients. According to those authors, the factors that may affect the validity of a diet questionnaire are: (1) respondent characteristics, (2) questionnaire design and quantification, (3) adequacy of the reference data and (4) quality control of data management.

Judging the quality of dietary exposure assessment in epidemiological studies is a crucial element. The scoring system proposed in the present paper or other tools that may follow can contribute to increasing the quality of the evidence in nutrition research, due to its capacity to serve as guidance for validation studies of diet questionnaires as well as for its utility in the selection and weighting of results from already conducted nutritional epidemiological studies.

### Acknowledgements

The studies reported herein have been carried out within the EURRECA Network of Excellence ([www.eurreca.org](http://www.eurreca.org)), financially supported by the Commission of the European Communities, specific Research, Technology and Development Programme Quality of Life and Management of Living Resources, within the Sixth Framework Programme, contract no. 036196. This report does not necessarily reflect the Commission's views or its future policy in this area. L. S.-M. developed the initial validation instrument, coordinated the discussion and consensus, designed the study and wrote the paper. L. F. A. took part in the discussions of the

initial validation instrument and contributed to the discussion, consensus and manuscript content. P. H. conducted some statistical analysis and contributed to the discussion and consensus. J. D.-A. conducted some analysis, contributed to the discussion, consensus and manuscript content. A. S.-V. contributed to the discussion, consensus and manuscript content. A. O.-A. contributed to the discussion, consensus and manuscript content. E. N. contributed to the discussion, consensus and manuscript content. C. L. V. contributed to the discussion, consensus and manuscript content. The authors have no conflict of interests to report.

## References

1. Willett WC & Lenart E (1998) Reproducibility and validity of food-frequency questionnaires. In *Nutritional Epidemiology*, 2nd ed., pp. 101–147 [WC Willett, editor]. Oxford: Oxford Medical Publications.
2. Bingham SA & Nelson M (1991) Assessment of food consumption and nutrient intake. In *Design Concepts in Nutritional Epidemiology*, pp. 153–191 [BM Margetts and M Nelson, editors]. Oxford: Oxford Medical Publications.
3. Cameron ME & Van Staveren WA (1988) *Manual on methodology for Food Consumption Studies*. Oxford: Oxford University Press.
4. Poslusna K, Ruprich J, de Vries JH, *et al.* (2009) Misreporting of energy and micronutrient intake estimated by food records and 24 hour recalls, control and adjustment methods in practice. *Br J Nutr* **101**, Suppl. 2, S73–S85.
5. Willett WC, Sampson L, Stampfer MJ, *et al.* (1985) Reproducibility and validity of a semiquantitative food frequency questionnaire. *Am J Epidemiol* **122**, 51–65.
6. Willett WC, Reynolds RD, Cottrell-Hoehner S, *et al.* (1987) Validation of a semi-quantitative food frequency questionnaire: comparison with a 1-year diet record. *J Am Diet Assoc* **87**, 43–47.
7. Rohan TE & Potter JD (1984) Retrospective assessment of dietary intake. *Am J Epidemiol* **120**, 876–887.
8. Ashwell M, Lambert JP, Alles MS, *et al.* (2008) How we will produce the evidence-based EURRECA toolkit to support nutrition and food policy. *Eur J Nutr* **47**, Suppl. 1, 2–16.
9. Henríquez-Sánchez P, Sánchez-Villegas A, Doreste-Alonso J, *et al.* (2009) Dietary assessment methods for micronutrient intake: a systematic review on vitamins. *Br J Nutr* **102**, Suppl. 1, S10–S37.
10. Serra-Majem L, Pfrimer K, Doreste-Alonso, S, *et al.* (2009) Dietary assessment methods for intakes of Iron, Calcium, Selenium, Zinc and Iodine. *Br J Nutr* **102**, Suppl. 1, S38–S55.
11. Øverby NC, Serra-Majem L & Frost Andersen L (2009) Dietary assessment methods on *n*-3 fatty acids intake: a systematic review. *Br J Nutr* **102**, Suppl. 1, S56–S63.
12. Ortiz-Andrellucchi A, Doreste-Alonso J, Henríquez-Sánchez P, *et al.* (2009) Dietary assessment methods for micronutrient intake in pregnant women: a systematic review. *Br J Nutr* **102**, Suppl. 1, S64–S86.
13. Ortiz-Andrellucchi A, Henríquez-Sánchez P, Sánchez-Villegas A, *et al.* (2009) Dietary Assessment Methods for micronutrient intake in infants, children and adolescents: a systematic review. *Br J Nutr* **102**, Suppl. 1, S87–S117.
14. Ortiz-Andrellucchi A, Sánchez-Villegas A, Doreste-Alonso J, *et al.* (2009) Dietary Assessment Methods (DAME) for micronutrient intake in elderly people: a systematic review. *Br J Nutr* **102**, Suppl. 1, S118–S149.
15. Bland M (2000) *An introduction to medical statistics*. 3rd ed. pp. 247–250. Oxford: Oxford University Press.
16. Sudha V, Radhika G, Sathya RM, *et al.* (2006) Reproducibility and validity of an interviewer-administered semi-quantitative food frequency questionnaire to assess dietary intake of urban adults in southern India. *Int J Food Sci Nutr* **57**, 481–493.
17. Kaaks RJ (1997) Biochemical markers as additional measurements in studies of the accuracy of dietary questionnaire measurements: conceptual issues. *Am J Clin Nutr* **65**, Suppl. 4, 1232S–1239S.
18. Kaaks R, Ferrari P, Ciampi A, *et al.* (2002) Uses and limitations of statistical accounting for random error correlations, in the validation of dietary questionnaire assessments. *Public Health Nutr* **5**, 969–976.
19. Bland JM & Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* **1**, 307–310.
20. Bland JM & Altman DG (1999) Measuring agreement in method comparison studies. *Stat Methods Med Res* **8**, 135–160.
21. Ambrosini GL, de Klerk NH, Musk AW, *et al.* (2001) Agreement between a brief food frequency questionnaire and diet records using two statistical methods. *Public Health Nutr* **4**, 255–264.
22. Negri E, Franceschi S, La Vecchia C, *et al.* (1994) The application of different correlation coefficients to assess the reproducibility of a food frequency questionnaire. *Eur J Cancer Prev* **3**, 489–497.
23. Serra-Majem L, Ribas-Barba L & Aranceta-Bartrina J (2006) Evaluación del consumo de alimentos en poblaciones. Encuestas alimentarias (Evaluation of food consumption in populations: dietary surveys). In *Nutrición y Salud Pública: métodos, bases científicas y aplicación (Nutrition and Public Health Methods Applications and Scientific Basis)*, 2nd ed., pp. 136–145 [L Serra-Majem and J Aranceta-Bartrina, editors]. Barcelona: Masson.
24. Ngo J, Gurinovic M, Frost-Andersen L, *et al.* (2009) How dietary intake methodology is adapted for use in European immigrant population groups – a review. *Br J Nutr* **101**, Suppl. 2, S86–S94.
25. Thompson FE & Subar AF (2008) Dietary assessment methodology. In *Nutrition in the Prevention and Treatment of Disease*, 2nd ed., pp. 3–39 [A Coulston and C Boushey, editors]. Washington, DC: Academy Press.
26. Kohlmeier L (1994) Gaps in dietary assessment methodology: meal- vs list-based methods. *Am J Clin Nutr* **59**, Suppl. 1, 175S–179S.
27. Cade JE, Burley VJ, Warm DL, *et al.* (2004) Food frequency questionnaires: a review of their design, validation and utilization. *Nutr Rev* **17**, 5–22.
28. Willett WC, Sampson L, Browne ML, *et al.* (1988) The use of a self-administered questionnaire to assess diet four years in the past. *Am J Epidemiol* **127**, 188–199.
29. Dennis LK, Snelelaar LG, Nothwehr FK, *et al.* (2003) Developing a scoring method for evaluating dietary methodology in reviews of epidemiologic studies. *J Am Diet Assoc* **103**, 483–487.
30. Kipnis V, Midthune D, Freedman L, *et al.* (2002) Bias in dietary-report instruments and its implications for nutritional epidemiology. *Public Health Nutr* **5**, 915–923.
31. Bingham SA, Gill C, Welch A, *et al.* (1994) Comparison of dietary assessment methods in nutritional epidemiology: weighed records v. 24 h recalls, food-frequency questionnaires and estimated-diet records. *Br J Nutr* **72**, 619–643.
32. Bingham S, Luben R, Welch A, *et al.* (2008) Associations between dietary methods and biomarkers, and between fruits and vegetables and risk of ischaemic heart disease, in the EPIC Norfolk Cohort Study. *Int J Epidemiol* **37**, 978–987.
33. Freedman LS, Schatzkin A, Thiebaut ACM, *et al.* (2007) Abandon neither the Food Frequency Questionnaire nor the dietary fat-breast cancer hypothesis. *Cancer Epidemiol Biomarkers Prev* **16**, 1321–1322.

34. Willett WC & Hu FB (2007) The food frequency questionnaire: letter to the editors. *Cancer Epidemiol Biomarkers Prev* **16**, 182–183.
35. Kipnis V, Subar AF, Midthune D, *et al.* (2003) The structure of dietary measurement error: results of the OPEN biomarker study. *Am J Epidemiol* **158**, 14–21.
36. Schatzkin A, Kipnis V, Carroll RJ, *et al.* (2003) A comparison of a food frequency questionnaire with a 24-hour recall for use in an epidemiological cohort study: results from the biomarker-based Observing Protein and Energy Nutrition (OPEN) study. *Int J Epidemiol* **32**, 1054–1062.
37. Kristal AR & Potter JD (2006) Not the time to abandon the food frequency questionnaire: counterpoint. *Cancer Epidemiol Biomarkers Prev* **15**, 1759–1760.
38. Kipnis V, Midthune D, Freedman LS, *et al.* (2001) Empirical evidence of correlated biases in dietary assessment instruments and its implications. *Am J Epidemiol* **153**, 394–403.
39. Willett WC (2008) Commentary: flawed study designs are not salvaged by large samples. *Int J Epidemiol* **37**, 987–989.
40. Block G & Hartman AM (1989) Issues in reproducibility and validity of dietary studies. *Am J Clin Nutr* **50**, Suppl. 5, 1133–1138, Discussion 1231–1235.