Issues in assessing the validity of nutrient data obtained from a food-frequency questionnaire: folate and vitamin B_{12} examples

Victoria M Flood[1], Wayne T Smith[2], Karen L Webb[1,3] and Paul Mitchell[4,5,*]

[1]Department of Public Health, University of Sydney, Sydney, Australia: [2]Centre for Clinical Epidemiology and Biostatistics, Newcastle University, Newcastle, Australia: [3]Department of Molecular and Microbial Biosciences, University of Sydney, Sydney, Australia: [4]Department of Ophthalmology (Centre for Vision Research), University of Sydney, Sydney, Australia: [5]Westmead Millennium Institute, Centre for Vision Research, Westmead Hospital, Westmead, New South Wales 2145, Australia

Abstract

Objective: To compare methods used to assess the validity of nutrient intake data obtained from a food-frequency questionnaire (FFQ), using folate and vitamin B_{12} as nutrient examples.

Design: Cross-sectional sample from a population cohort.

Setting: Two postcode areas west of Sydney, Australia.

Subjects: In total, 2895 people aged 49 years and older provided dietary data using a semi-quantitative FFQ (79% of 3654 subjects examined). The validity of the FFQ was assessed against three 4-day weighed food records (WFRs) completed by 78 people (mean age 70 years).

Results: Folate and vitamin B_{12} validity data were assessed using different methods. The Spearman ranked correlations (energy-adjusted) were 0.66 for folate and 0.38 for vitamin B_{12}. Using the Bland–Altman method, following $\log_e$ transformation, no linear trend existed between the differences and means for folate and vitamin B_{12}. Large differences existed between the FFQ and WFR in individual cases, particularly for vitamin B_{12}. Finally, data were divided into quintile categories for the test and reference method: 79% classified folate within one quintile, 65% classified vitamin B_{12} within one quintile; there was no gross misclassification for folate and only 3% misclassification for vitamin B_{12}.

Conclusions: Different methods of analysis provided different information about the validity of the FFQ. Correlation coefficients should not be used alone to assess the validity of nutrient data, but should be used in conjunction with Bland–Altman analyses. Depending on the use of the data, additional assessment of classification categories is recommended. This worked example demonstrates that absolute intakes of folate and vitamin B_{12} should be used with caution.

The validity of a measuring instrument or tool is the degree to which it produces a true and accurate assessment of what it intends to measure. Knowledge about the validity of dietary data used in research is paramount to a thorough evaluation and interpretation of nutritional research findings. The food-frequency questionnaire (FFQ) is a tool commonly used in large epidemiological studies, because it is relatively inexpensive and easy to administer and analyse for large numbers of people[1]. Dietary assessment methods all involve some degree of measurement error. Thus, rather than evaluating the validity of a questionnaire against a 'gold standard', the relative validity or 'inter-method reliability' is measured, which is the ability of two methods to provide similar results[2]. Methods used to assess and interpret the relative validity of FFQs

have varied in the literature, but have tended to rely on correlation analysis of nutrients and/or foods measured by two or more dietary assessment methods[3]. A recent consensus document on FFQs[3] recommends that the relative validity of FFQs be assessed using a variety of statistical approaches; the selection of these approaches should be guided by the purpose of measuring diet with the FFQ.

This study illustrates an assessment of the validity of an FFQ with weighed food records (WFRs), using different statistical approaches, from the practical examples of folate and vitamin B_{12}. Smith *et al.*[4] conducted a previous validation study of the FFQ used in the Blue Mountains Eye Study (BMES); however, folate and vitamin B_{12} were not assessed at that time, because food composition data

Corresponding author: Email paul_mitchell@wmi.usyd.edu.au

for folate and vitamin B_{12} were not available in the Australian nutrient database. The 1994 validity study assessed agreement between the FFQ and the WFR using two approaches: correlation analysis and an assessment of misclassification error, after subdividing the data into quintiles (including weighted kappa statistics). The methods used in the present paper include correlation analysis (using Pearson and Spearman correlations), quintile categories (weighted kappa) and the Bland–Altman method[5].

Methods

Study population

The BMES is a population-based cohort study of common eye diseases among residents of a defined area, west of Sydney, who were aged 49 years or older. This population is representative of older Australians in most respects[6]. In summary, of 4433 eligible people, 3654 (82.4%) attended an eye examination during the period 1992–1994.

Food-frequency questionnaire

The 145-item semi-quantitative FFQ was modified for the Australian diet and vernacular from an earlier FFQ of Willett *et al.*[7], and included portion size estimates and the usual frequency, strength, brand and type of supplements. Participants attempting the FFQ numbered 3267 (89%) and, of these, 2895 were usable (79% of those examined, 89% of those who attempted the FFQ)[4]. FFQs with more than 12 items missing or with implausible, extreme values were excluded. Respondents of the FFQ were asked about the foods eaten in the previous 12 months, and an allowance for seasonal variation of fruit and vegetables was made during analyses by weighting seasonal fruits and vegetables.

Subjects in the validation study

A random selection of 186 BMES subjects, weighted to include more older people (aged 65–85 years), were invited to take part in the validation study in 1994. Each subject was required to complete three 4-day WFRs approximately four months apart[4], which therefore incorporated seasonal variation and thus reflected the allowance for seasonal variation applied to the FFQ analysis. Of the 150 people who agreed to participate, 139 began recording food intake and 78 subjects (52% of those who agreed to participate) completed all three 4-day WFRs, which were included in the secondary analysis of folate and vitamin B_{12}.

Nutrient database

Folate and vitamin B_{12} nutrient composition values for the analyses of FFQs and WFRs were obtained from the AUSNUT nutrient database[8] and the UK tables of food composition[9], respectively. In these secondary analyses, the data were entered using the software package

SERVE[10]. Mean daily estimates were calculated for each FFQ and the three 4-day WFRs.

Statistical methods

Analyses used SPSS, version 9.0 for Windows[11]. Concurrent validity of the FFQ compared with the WFRs was assessed using several methods:

1. Pearson product–moment and Spearman ranked correlations.
2. Bland–Altman limits of agreement (LOA)[5], in which the mean agreement between the two methods was calculated, i.e. FFQ − WFR. The LOA define the limits within which 95% of these differences are expected to fall (mean ± two standard deviations of the differences). The differences between the two methods were plotted against the average of the two methods. Any dependency between the two methods was tested by fitting the regression line of differences (H_0: $\beta = 0$, $\alpha = 0.05$), i.e. ideally if the two methods are equally variable, the correlation between the differences would equal zero. Natural-log (ln) transformation was performed since the dietary data were skewed, as recommended by Bland and Altman[5]. To permit further interpretation of the ln-transformed data, the antilogs of the limits were taken, providing a ratio FFQ/WFR of the data. The ratios were multiplied by 100 and are therefore expressed as percentages, 100% representing ideal agreement.
3. Joint classification of nutrient intake assessed by the FFQ and the average of the three WFRs was assessed using quintiles of intake for each nutrient from the FFQ and WFR, respectively. The proportion grossly misclassified applied when one dietary assessment method classified the individual's intake into the lowest quintile and the other method classified it into the highest quintile. Quadratic weighted kappa values were calculated comparing quintiles of intake for each nutrient from the FFQ and WFR[12].

In all methods, energy-adjusted intakes were calculated using the method suggested by Willett and Stampfer[13]. Nutrient intake values were replaced with their respective residuals from a regression model with the nutrient intake as the dependent variable and the total energy intake as the independent variable. A constant, the expected nutrient value for the entire population, was added to the residual for each nutrient. Energy adjustment was done separately for each FFQ and the average of the three WFRs.

Results and discussion

Subjects in the validation study

Subjects who completed a usable FFQ were more likely to be younger, with a mean age of 1 year younger than the

whole study population, but were no more likely to have serious eye disease (data not shown). The proportion of men and women participating in the validation and BMES studies was similar (about 45% men and 55% women). By design, those participating in the validation study were older than those participating in the BMES cohort study, by a mean age of around 5 years (65 years in BMES and 70 years in the validation study; $P < 0.0001$). The generalisability of this validation study to the whole study population is reasonable. A slightly older age range was selected to conduct the validation study because we were mainly interested in age-related diseases uncommon in younger people and we wanted to be sure of a valid instrument for subjects likely to be cases. The participation rate in the validation study was acceptable and the loss to follow-up was not unduly large for such a long-term study with high subject burden[4].

Comparison of group means

Group means should be assessed if absolute intakes are examined[3]. If the data are normally distributed, paired t-tests can be used. However, it is common for nutrient data to not be normally distributed (as was the case in our dataset), so other non-parametric tests should be used, such as the Wilcoxon signed-rank sum test to compare group means. The mean daily intakes of folate and vitamin B_{12} for the average of the three 4-day WFRs and the FFQs are shown in Table 1. The FFQ provided higher mean estimates of nutrient intake than the WFRs (all were statistically significantly different using the Wilcoxon signed-rank sum test: folate, $P < 0.0001$; vitamin B_{12}, $P < 0.0001$).

Correlation

Past validation studies have commonly calculated the correlation coefficient (r) between two methods of dietary assessment. This function measures the strength of the relationship between two variables, but not the agreement between them[5]. Correlation of a nutrient may be high, but one method may consistently give higher estimates than the other method. Furthermore, the test of significance for a correlation is irrelevant, as we would expect a relationship to exist between two methods of measuring the same thing[5]. However, the correlation may have some value as it can be compared with measurements evaluated in past studies and is relatively easy to interpret. The consensus document on FFQs recommends using correlation for continuous data, in conjunction with the Bland–Altman analysis. Pearson's correlation should be used for normally distributed data and Spearman's ranked correlation for non-normally distributed data.

Pearson product–moment correlation and Spearman rank correlation (adjusted for energy) for comparisons of nutrient intakes from the FFQ and the three 4-day WFRs are shown in Table 1. The estimated folate intake gave a better correlation between the FFQ and WFR (0.67 Pearson) than the estimated intake for vitamin B_{12} (0.18 Pearson). The correlation for folate was similar to that observed for vitamin C (0.69 Pearson) and thiamine (0.56 Pearson) as previously reported by Smith et al.[4]. The correlation for vitamin B_{12} was similar to those found previously for protein (0.18 Pearson) and zinc (0.10 Pearson). This is not surprising because these nutrients are found in the same types of food (mainly meats). In the previous validation study[4], the frequency of meat intake from the FFQ was similar to that observed in the WFRs. However, while the mean serving size of meat was similar between the FFQ and WFR, a considerable proportion of subjects reported serving sizes in the WFR that differed substantially from those reported in the FFQ (ranging from half to almost double the standard serving sizes used in the FFQ). This is a well-known problem associated with the

Table 1 Results of various methods to assess agreement between the FFQ and the WFR for intakes of folate and vitamin B_{12} in the Blue Mountains Eye Study dataset

	Folate	Vitamin B_{12}
Intake (μg), mean (SD)		
Average of three WFRs ($n = 78$)	238 (67)	3.2 (2.2)
FFQ ($n = 78$)	329 (114)	4.4 (2.7)
Overall population reliably completing the FFQ ($n = 2895$)	329 (105)	4.9 (3.4)
Pearson correlation*	0.67	0.18
Spearman ranked correlation*	0.66	0.38
Mean difference (FFQ − WFR)*†	0.40	0.40
LOA*†	0.14, 0.65	0.22, 1.04
Antilog (%), mean (95% CI)*	149 (145−154)	150 (146−167)
Antilog of LOA (%)*	115, 192	80, 282
Percentage correctly classified into the same quintile*	46	37
Percentage classified within one quintile*	79	65
Percentage grossly misclassified*‡	0	3
Weighted kappa	0.33	0.22

FFQ – food-frequency questionnaire; WFR – 4-day weighed food record; SD – standard deviation; LOA – limits of agreement; CI – confidence interval.
* Energy-adjusted, using the residuals from regression models in which energy intake is the independent variable and nutrient intakes are the dependent variables.
† Energy-adjusted and ln-transformed.
‡ One dietary method classifies intake into the bottom quintile; the other method classifies intake into the top quintile.

use of FFQs and suggests that subjects were not appropriately adjusting their reported frequency by serving size.

Bland–Altman analysis

The Bland–Altman analysis[5] assesses the agreement between two dietary assessment methods across the range of intakes. This analysis is able to assess if there is any bias between the two methods, i.e. whether there is any systematic difference between the two methods and the extent to which the two methods agree. This is achieved by plotting the difference between the two dietary assessment methods against the average of the two methods. The bias and the LOA need to be interpreted in the context of the use of the questionnaire[3].

Application of the Bland–Altman analysis to the assessment of the validity of FFQs is relatively new, with only a few published studies comparing FFQs with diet records by this means[14,15].

Figures 1 and 2 illustrate the findings of a Bland–Altman analysis for folate and vitamin B$_{12}$, respectively. After energy adjustment and ln transformation of the data, the difference in nutrient intake between the two methods, FFQ − WFR, was plotted on the y-axis and the average of the two methods, (FFQ + WFR)/2, on the x-axis. Bland and Altman[5] recommended log transformation on skewed data in order to narrow the LOA and assist interpretation. The mean difference between the FFQ and WFR for folate, prior to ln transformation, was 171.5 µg, and a fitted regression line indicated a significant linear trend ($P < 0.0001$). That is, a dependency existed between the

difference of the two methods and the average of the two methods; as the folate intake of individuals increased, so did the magnitude of the error between the FFQ and WFR (data not shown). However, after ln transformation of the skewed data (Fig. 1), the regression line for folate no longer indicated a significant linear trend. To further assist the interpretation of this ln-transformed data, it is useful to consider the antilog of the data. The antilog of the ln-transformed data represents the energy-adjusted ratio FFQ/WFR. In 95% of cases, the FFQ will provide 15–92% higher folate estimates than the WFR (Table 1). All estimates of folate from the FFQ were greater than the WFR estimates. The estimates of dietary intake from the FFQ were obtained from a relatively extensive list of 145 items. Folate from natural sources, particularly fruit and vegetables, may have been overestimated because of the known tendency of FFQs to overestimate intakes of these foods when the list of individual fruits and vegetables is long[16].

For vitamin B$_{12}$, after ln transformation of the skewed data, the LOA were −0.22 and 1.04 (Fig. 2). Information about antilogs in Table 1 informs interpretation. In 95% of cases, the vitamin B$_{12}$ estimates from the FFQ differed from those from the WFR: from 20% below to 182% above. This large variation could be due to the variation from standard meat servings noted in the WFRs, as discussed earlier.

Classification into categories of consumption

This is a useful method if the data are divided into quintiles and compared to the likelihood of an association with a
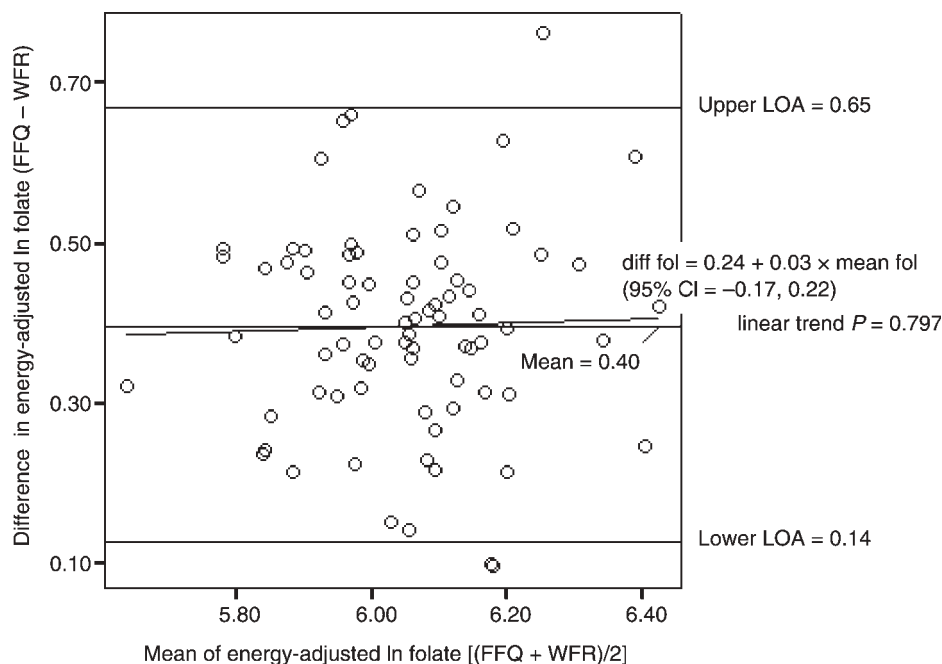

Fig. 1 Bland–Altman method of assessing agreement between the FFQ and WFR for dietary folate intake, after ln transformation, applied to data from the Blue Mountains Eye Study ($n = 78$). FFQ − food-frequency questionnaire; WFR − 4-day weighed food record; LOA − limit of agreement; CI − confidence interval
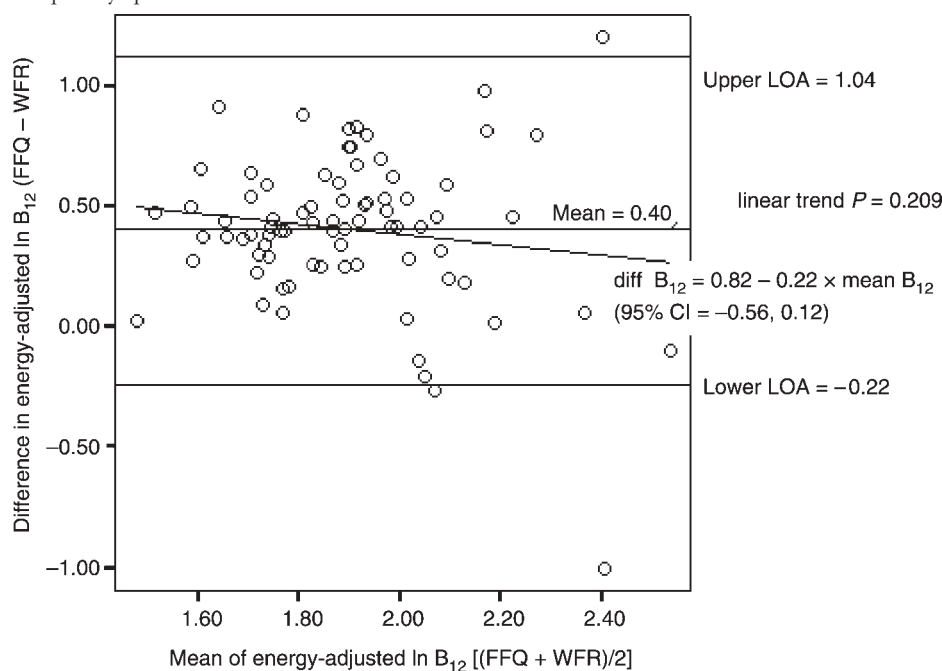
Fig. 2 Bland–Altman method of assessing agreement between the FFQ and WFR for dietary vitamin B_{12} intake, after ln transformation, applied to data from the Blue Mountains Eye Study ($n = 78$). FFQ – food-frequency questionnaire; WFR – 4-day weighed food record; LOA – limit of agreement; CI – confidence interval

disease or other outcome, as is commonly performed in large epidemiological studies.

In this study, the proportions of subjects correctly classified within one quintile category for folate and vitamin B_{12} were 79% and 65%, respectively (Table 1). There was no gross misclassification for folate and only 3% misclassification for vitamin B_{12}. Weighted kappa values were fair to poor (0.33 for folate and 0.22 for vitamin B_{12}). The relatively high proportion of people classified within one quintile and the small number grossly misclassified for folate and vitamin B_{12} illustrate that the FFQ is capable of reasonably ranking vitamin B_{12} and folate intakes.

Conclusions

The Bland–Altman analysis used to assess the agreement between the two dietary assessment methods in this study provides additional information that cannot be obtained from correlation coefficients only. Although folate has a higher correlation coefficient than vitamin B_{12}, the Bland–Altman plots provide further information about the validity of the FFQ for these nutrients. Prior to natural-log transformation, the fitted regression line for folate indicated that as subjects consumed higher quantities of folate, the FFQ had a tendency to overestimate their folate intake by a greater amount. Although the difference for vitamin B_{12} was not statistically different from zero, the graph demonstrates a large difference between the two methods in some individuals; the FFQ provided vitamin B_{12} estimates up to almost three times greater than the WFR.

Our 1994 validation study concluded that, overall, the BMES FFQ was a valid instrument for classifying older, community-based subjects, and performed well for most macronutrients and micronutrients, with the exceptions of protein, retinol and zinc. The current validation study for the nutrients folate and vitamin B_{12} indicates similar findings. Overall, folate has a reasonable agreement with an alternative dietary assessment method and performed very well when ranking individuals by quintiles. The FFQ has a tendency to overestimate folate intake, and this appears to be greatest in those with higher intakes. Poorer agreement was noted between the two dietary assessment methods for vitamin B_{12}, possibly reflecting the difficulty in estimating meat servings when using the FFQ. However, the FFQ was reasonable for ranking individuals according to their vitamin B_{12} intake. Thus, the estimates of absolute intakes of folate and vitamin B_{12} should be interpreted with caution.

Acknowledgements

References

1 Thompson FE, Byers T. Dietary assessment resource manual. *Journal of Nutrition* 1994; **124**: 2245S–317S.
2 Armstrong BK, White E, Saracci R. *Principles of Exposure Measurement in Epidemiology*. Oxford: Oxford University Press, 1992.

3   Burley V, Cade J, Margetts B, Thompson R, Warm D. *Consensus Document on the Development, Validation and Utilisation of Food Frequency Questionnaires.* London: Ministry of Agriculture Fisheries and Food, 2000.

4   Smith W, Mitchell P, Reay EM, Webb K, Harvey PWJ. Validity and reproducibility of a self-administered food frequency questionnaire in older people. *Australian and New Zealand Journal of Public Health* 1998; **22**(4): 456–63.

5   Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical assessment. *Lancet* 1986; **1**(8476): 307–11.

6   Mitchell P, Smith W, Attebo K, Wang JJ. Prevalence of age-related maculopathy in Australia: The Blue Mountains Eye Study. *Ophthalmology* 1995; **102**: 1450–60.

7   Willett W, Sampson L, Browne M, Stampfer M, Rosner B, Hennekins C, *et al*. The use of a self-administered questionnaire to assess diet four years in the past. *American Journal of Epidemiology* 1988; **127**: 188–99.

8   Australia and New Zealand Food Authority (ANZFA). *National Nutrition Survey Nutrient Database (AUSNUT)* [software]. Cariberra: ANZFA, 1999.

9   Holland B, Welch AA, Unwin ID, Buss DH, Paul AA, Southgate DAT. *McCance & Widdowson's The Composition of Foods*, 5th ed. London: The Royal Society of Chemistry and Ministry of Agriculture, Fisheries and Food, 1991.

10 H & M Williams. *SERVE Nutrition Management System, Version 3.95* [software]. Sydney: H&M Williams, 2001.

11 SPSS, Inc. *SPSS for Windows, Version 9.0*. Chicago, IL: SPSS, Inc., 1989-1996.

12 Streiner DL, Norman GR, eds. Reliability. In: *Health Measurement Scales: A Practical Guide to their Development and Use*. New York: Oxford University Press, 1989; 94–5.

13 Willett W, Stampfer MJ. Total energy intake: implications for epidemiologic analyses. *American Journal of Epidemiology* 1986; **124**(1): 17–27.

14 Ambrosini GL, de Klerk NH, Musk AW, Mackerras D. Agreement between a brief food frequency questionnaire and diet records using two statistical methods. *Public Health Nutrition* 2001; **4**(2): 255–64.

15 Thompson RL, Margetts BM. Comparison of a food frequency questionnaire with a 10-day weighed record in cigarette smokers. *International Journal of Epidemiology* 1993; **22**(5): 824–33.

16 Willett WC. Invited commentary: comparison of food frequency questionnaires. *American Journal of Epidemiology* 1998; **338**: 1157–9.