Public Health Nutrition: 4(2), 255–264 DOI: 10.1079/PHN200062

Agreement between a brief food frequency questionnaire and diet records using two statistical methods

GL Ambrosini[1,*], NH de Klerk[1], AW Musk[1,2,3] and D Mackerras[4]
[1]Department of Public Health, University of Western Australia, Nedlands Campus, Perth, W.A. 6907, Australia: [2]Department of Respiratory Medicine, Sir Charles Gairdner Hospital, Perth, W.A., Australia: [3]Department of Medicine, University of Western Australia, Perth, W.A., Australia: [4]Menzies School of Health Research, Darwin, N.T., Australia

Submitted 1 November 1999: Accepted 1 July 2000

Abstract

Objective: To compare intra- and inter-method reliability of a semi-quantitative food frequency questionnaire (FFQ) designed specifically to measure beta carotene (BC) and retinol intake, using two methods – the limits of agreement (LOA) and the correlation coefficient.
Design: A cross-sectional study of dietary intake.
Setting: A randomized trial of vitamin A supplements in 2769 subjects with past asbestos exposure.
Subjects: Data from 57 men and 26 women, aged 28–72 years, living in Western Australia.
Methods: The FFQ was administered at baseline (FFQ1) and repeated 1 year later (FFQ2). Four 1-week diet records (DRs) were completed during the year.
Results: Mean agreement between FFQ2 and FFQ1 was 120% for BC and 98% for retinol. LOA were 47–306% and 21–459%, respectively. Mean agreement between FFQ2 and the DR was 149% for BC and 63% for retinol; LOA were 50–447% and 11–349%, respectively. Mean agreement and LOA varied across energy intakes. Between the DR and FFQ2, correlation coefficients were 0.36 for BC and 0.51 for retinol. These varied considerably across age, gender and energy intakes and were not in accordance with limits of agreement findings.
Conclusion: Although correlation coefficients were positive and significant, there was less than ideal intra-method and inter-method reliability shown by the limits of agreement method. Bias was uneven across the range of intakes, the LOA were wide and, compared with the DR, the FFQ significantly over-estimated BC and under-estimated retinol. This shows the limitations of calculating correlation coefficients alone, for assessing reliability and validity.

Keywords
Food frequency questionnaire
Diet records
Reliability
Validity
Reproducibility
Nutritional epidemiology
Vitamin A

The self-administered food frequency questionnaire (FFQ) is one of the most popular tools for the assessment of dietary intakes in epidemiological research. It is highly practical and provides estimates that are more representative of usual intake than a 24-h dietary recall[1]. However, the inaccuracies in estimating dietary intake over the long term often lead to large errors, loss of statistical power, and bias in measures of association[2]. Sufficiently large errors can completely mask existing diet–disease associations[3,4].

The FFQ has become widely accepted in epidemiology, mostly because its estimates have been shown to correlate with other dietary methods[1]. Many of the published validation studies have generally calculated the correlation coefficient between a questionnaire and a reference method. This may give an over-optimistic indication of agreement, as a positive correlation is expected between two methods supposedly measuring the same thing. Statisticians have repeatedly commented that because the correlation coefficient is dependent on the range and variation in the sample, it reflects association but not agreement[5–8]. In its failure to describe bias, it may give a misleading impression about the extent of agreement[9].

A more informative approach is to calculate the limits of agreement, which provide information about the direction of bias, whether or not bias is constant across levels of intake, and the magnitude of errors between the measurements[6]. This method can be applied when examining the repeated administration of an instrument or when comparing it with a reference method. Using the

*Corresponding author: Email ginaa@dph.uwa.edu.au

limits of agreement method, we illustrate the limitations of using correlation coefficients to describe the intra-method reliability of an FFQ and its inter-method reliability compared with average intake from four 1-week diet records (DRs).

Methods

For the purpose of this paper, we refer to reliability as the repeatability of a method, or the degree of stability in its repeated measures[10]. Validity is the degree to which a measurement actually measures what it was designed to, usually the 'true' measure[10]. In the assessment of dietary methods, and indeed many other methods, only the relative validity may really be estimated, as there is no gold standard for measuring 'true' dietary intake. We therefore refer to the comparison of dietary methods as inter-method reliability[8].

Study subjects

A reliability study was conducted within a randomized clinical trial (the Vitamin A Program) which was designed to examine the efficacy of beta carotene and retinol (preformed vitamin A) supplements in reducing the risk of lung cancer, mesothelioma and possibly other malignancies after asbestos exposure[11]. Participants for the Vitamin A Program ($n = 2769$) were recruited from two large cohorts of former workers and residents of an asbestos mining and milling town in Western Australia[12]. A semi-quantitative FFQ specially designed to estimate average daily intake of dietary and supplementary beta carotene (BC) and retinol was administered to all trial participants at baseline, to check for similar intakes in the two randomized groups and to investigate diet–disease relationships.

Adults entering the trial in its sixth month were asked if they would provide four 1-week DRs and repeat the FFQ over the following year. Most of the 570 people enrolling during this time were invited, except where language, reading or writing difficulties were apparent at their interview. Subjects were actively recruited until a sample of at least 100 was obtained.

Ethical approval

All subjects gave informed consent in participating in the Vitamin A Program, and the study was approved by the Human Rights Committee of the University of Western Australia and the Clinical Drug Trials Committee of the Sir Charles Gairdner Hospital, Perth, Western Australia.

Dietary methods

Vitamin A food frequency questionnaire

The semi-quantitative FFQ was posted to all trial candidates to complete at home and checked for completeness at their interview for induction into the Vitamin A Program (FFQ1). The questionnaire asked about usual dietary habits and vitamin supplement use during the previous year. Those in the reliability study repeated the FFQ 1 year later (FFQ2), after completing their fourth DR. To reduce the likelihood of vitamin A toxicity, subjects were advised at induction (after already completing FFQ1) to cease taking all forms of vitamin A supplements except their assigned trial supplements. Therefore, the reliability of supplementary intakes was not examined in this study, as alterations in supplement use were actively encouraged.

Foods in the FFQ were chosen using a ranking analysis of 1-week DRs from 60 control subjects participating in a local case–control study of renal calculi disease (unpublished data). The ranking criteria were based upon the food's BC and retinol contents, its ability to discriminate these intakes, and any other factors of nutritional significance, e.g. cruciferous vegetables. The final list of foods included (Table 1) was comparable to the major contributors of vitamin A identified by large UK and US dietary surveys[13,14].

The questionnaire specified a medium serving of each food and asked subjects to describe their usual serving as small, medium or large, in relation to this quantity. The serving sizes used in the FFQ were derived from the renal calculi study. There were eight options for the frequency of consumption ranging from 'never or less than once per month' up to '2 or more times per day'. Responses were collected separately for summer and winter consumption of fruits and vegetables to discern seasonal differences in intake. Mean daily retinol and beta carotene intakes were calculated using the Australian Food Composition Tables[15].

Diet record

To provide a 'reference' method for comparison with the FFQ, four 1-week DRs were completed in the year following FFQ1. The first DR was completed in early December (summer), followed by one every 3 months, representing each season. The food type and source (e.g. fresh, frozen, canned), cooking method, added cooking fats and subject's recipes for mixed dishes were obtained wherever possible. All DRs were entered by the same nutritionist (GLA) using Diet-1[16] and linked with the Australian Food Composition Tables[15]. Mean daily retinol and BC intakes were calculated by combining all four DRs.

Statistical analysis

Internal validity

The characteristics of those subjects who took part in the reliability study were compared with those joining the trial in the same period but who declined to take part ($n = 487$), to assess the generalizability of the reliability study results. Gender distributions and the prevalence of

https://doi.org/10.1079/PHN200062 Published online by Cambridge University Press

Table 1 Food items listed in the brief vitamin A food frequency questionnaire

Liver	Cheddar type cheese	Green beans or peas	Apples
Liverwurst/Pate	Cream cheese	Broccoli	Bananas
Meat stew including onion and carrot	Parmesan cheese	Carrot, raw	Oranges
Meat pie	Ice cream	Carrot, cooked	Apricots, raw or canned
Tomato based meat sauce, e.g. bolognese	Cooked eggs	Spinach or silverbeet	Apricots, dried
Homemade sweet pastries	Butter (not in cooking)	Pumpkin	Peaches, fresh
Homemade cake	Margarine (not in cooking)	Zucchini	Peaches, canned
Homemade biscuits	Vegetable soup including tomatoes or carrot	Sweet potato	Rockmelon
Cheesecake	Plain tomato sauce without meat	Tomato, raw	Figs, raw
Type of milk used	Mixed dishes containing cheese or eggs, including sauces	Olives	Orange juice, 100%
Milk added to cereals		Capsicum	
Milk added to drinks		Vegetable juice, including carrot	

smoking and supplement use at baseline were tested using the chi-square or Fisher's Exact Test, as appropriate. Average age and cigarettes smoked per day at baseline, and dietary intakes estimated by FFQ1 were tested using *t*-tests on log-transformed data.

Agreement

Using the limits of agreement method, the average agreement (or bias) between two methods or repeated measurements using the same method is calculated as the mean of their differences, e.g. mean (test − reference). The limits of agreement (LOA) define the boundaries within which 95% of these differences are expected to fall, and are calculated as (average agreement) $\pm t_{(n-1,\ 0.025)}$ (standard deviation of differences)[6]. The width of the LOA should be judged according to the required sensitivity of the measure or outcome, e.g. in view of diagnostic consequences or misclassification effects[6].

To detect uneven biases, the differences between two measures were plotted against the average of the two, and any dependency between them was formally tested by fitting the regression line of differences (H_o: $\beta = 0$, $\alpha = 0.05$). These steps were followed to estimate intra-method agreement between FFQ1 and FFQ2, and inter-method agreement between FFQ2 and the DR. Overall agreement was assessed by the degree of relative bias (mean of differences), the random error (width of LOA), and the dependence of differences on the magnitude of estimates. Even if the bias and dependency was small, extremely wide LOA indicate the potential for very large differences between methods or repeated measures, and agreement was considered poor.

Log transformation of skewed data is recommended before calculating the LOA, and this was necessary for the dietary intake estimates[6]. Subsequent anti-logging rendered mean agreement, LOA and 95% confidence limits as ratios, with FFQ2 expressed as a multiple of FFQ1 for intra-method reliability, and FFQ2 expressed as a multiple of the DR for inter-method reliability. Multiplying these ratios by 100 allows them to be expressed as percentages, with 100% representing ideal agreement[6].

Agreement was also assessed by gender, age (less than 55 years vs. 55 years and older) and energy intake strata. Subjects were divided into tertiles of energy intake using data from the DR. As the FFQ did not assess total energy intakes, adjustment of retinol or beta carotene intakes for energy intake was not possible.

Pearson's correlation coefficients

Intra- and inter-method Pearson correlation coefficients were calculated using log-transformed data with PC-SAS[17]. These were also estimated by gender, age group, and tertiles of energy intake.

Results

Generalizability

Out of 570 people enrolling in the trial during recruitment, 118 (21%) agreed to take part in the reliability study. Of these, 83 people (70%; 57 men and 26 women) successfully completed four DRs and 76 people (55 men, 21 women) completed FFQ1, FFQ2 and the DRs. Over half of those withdrawing from the reliability study did so before finishing their first DR. The most common reasons for not completing all four diet records or subsequent questionnaires were the commitment required ($n = 31$) and withdrawal from the trial because of non compliance or reported side effects ($n = 11$).

The source of subjects for the trial was a mining town population, and the majority of subjects for the trial and reliability study were men. There were no significant differences ($P < 0.05$) between trial and reliability study subjects in average age, prevalence of current smokers, or dietary beta carotene and retinol intake as reported by FFQ1, although the small number of subjects prevented reliable testing in some cases (Table 2). There were more female reliability subjects who had never smoked ($P = 0.059$), but those smoking at the time appear to have been consuming more cigarettes per day than trial subjects ($P = 0.20$) (Table 2). The number of subjects taking additional supplements was small, and did not differ greatly between reliability and trial subjects. There were

https://doi.org/10.1079/PHN200062 Published online by Cambridge University Press

Table 2 Baseline characteristics of reliability study and other trial subjects

	Women		Men	
	Reliability subjects	Trial subjects	Reliability subjects	Trial subjects
n	26	107	57	380
%	31	22	69	78
Age in years, mean ± SD	51 ± 11	50 ± 14	56 ± 11	54 ± 11
Never smoked (%)	70	50	28	22
Ex-smoker (%)	15	30	49	52
Current smoker (%)	15	20	23	26
Cigarettes smoked day^{-1} by current smokers, mean ± SD (*n*)	26 ± 16 (4)	17 ± 9 (21)	14 ± 13 (13)	19 ± 12 (99)
Dietary beta carotene (mg, mean ± SD)	4771 ± 2676	5776 ± 4104	5531 ± 3318	5957 ± 3944
Dietary retinol (mg, mean ± SD)	463 ± 676	414 ± 791	505 ± 539	559 ± 975
% taking beta carotene supplement	4	4	0	2
% taking retinol supplement	8	9	5	8

Dietary and supplementary intakes measured by FFQ1.

no consistent patterns indicating that those who volunteered for the reliability study were different in terms of their dietary and smoking behaviours.

Mean daily nutrient intakes

In both men and women, FFQs yielded the highest intake of dietary BC, and the DR the lowest (Table 3). FFQs gave lower retinol intakes than the DR in men but there was no consistent pattern for women, possibly owing to their small numbers. Additional supplements were not commonly taken, and were not reported on the DR even when reported on the FFQ.

Intra-method reliability

After log transformation of the data, there was no dependency between the average difference or bias between FFQ2 and FFQ1 and increasing intakes, for either BC or retinol (Figs 1 and 2). It is clear that the mean difference for retinol is approximately 0 on the log scale (or 100%) but this is significantly different from 0 for BC (Table 4). The LOA for both nutrients are wide, for example FFQ2 gives retinol estimates that are between 21% and 459% of FFQ1 even though, on average, the two estimates are the same (Table 4). In other words, even though the FFQ retinol results agree on average, the

Table 3 Mean daily nutrient intakes from FFQs and the average of four 1-week diet records

	FFQ1			FFQ2			Diet record		
	N	Mean ± SD	Range	*N*	Mean ± SD	Range	*N*	Mean ± SD	Range
Dietary beta carotene (mg)									
Men	57	5531±3318	1643–15995	55	6556±3201	1886–16495	57	4082±1770	1105–9700
Women	26	4771±2676	1382–13142	21	5352±3135	2317–13460	26	3826±1983	1145–10865
Supplementary beta carotene (mg)									
Men	0	–	–	0	–	–	0	–	–
Women	1	6000±0	–	0	–	–	0	–	–
Total beta carotene (mg)									
Men	57	5531±3318	1643–15995	55	6556±3201	1886–16495	57	4082±1770	1105–9700
Women	26	5002±2800	1381–13142	21	5352±3135	2317–13460	26	3826±1983	1145–10865
Dietary retinol (μg)									
Men	57	505±539	75–2846	55	617±1102	42–6588	57	939±1631	122–10334
Women	26	463±676	41–3417	21	370±394	41–1551	26	417±297	198–1407
Supplementary retinol (μg)									
Men	3	2561±411	2184–3000	6	2496±515	1500–3000	0	–	–
Women	2	1500±0	1500–1500	5	2367±865	1500–3515	0	–	–
Total retinol (μg)									
Men	57	640±755	75–3345	55	906±1328	42–6588	57	939±1631	122–10334
Women	26	578±759	41–3417	21	933±1048	41–3641	26	417±297	198–1407
Energy (kJ)							57	9618±1885	5413–12890
Men							26	6695±1173	4433–8718
Women									

https://doi.org/10.1079/PHN200062 Published online by Cambridge University Press

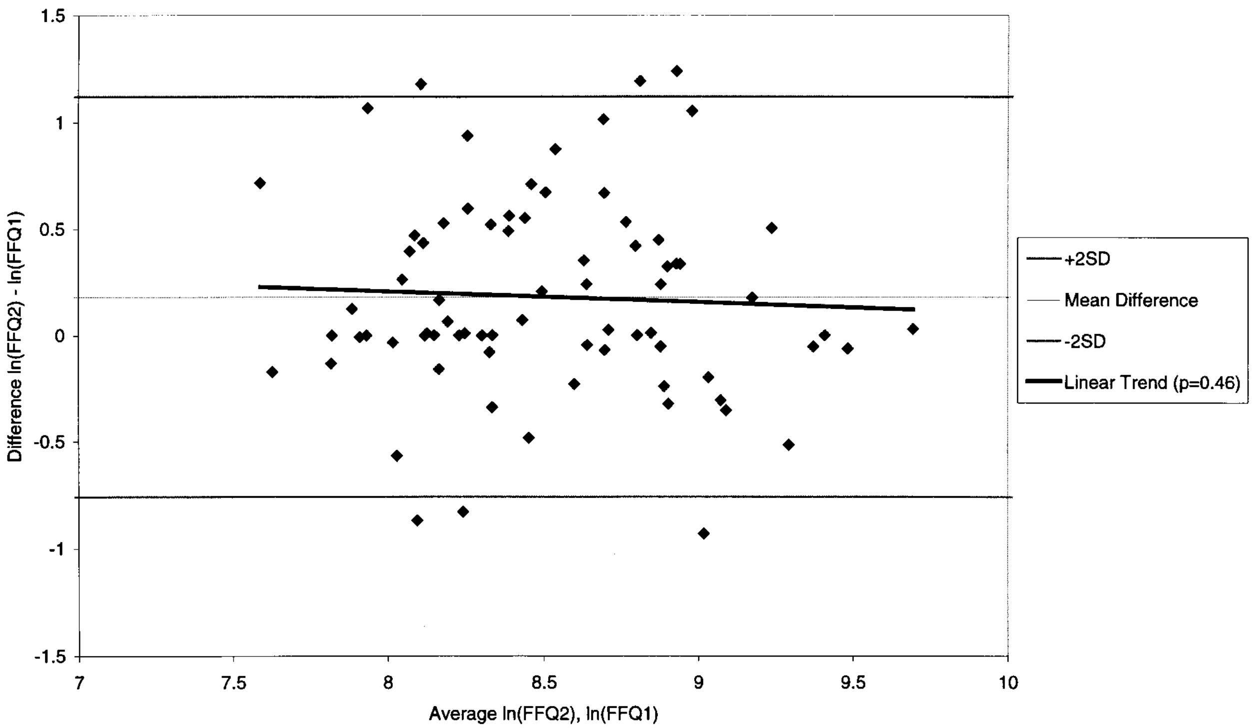

Fig. 1 Agreement between FFQ2 and FFQ1, dietary beta carotene (mg)

intakes reported by individuals could vary nearly five-fold in either direction from one occasion of reporting to another. On average, FFQ2 significantly over-estimates FFQ1 intakes of BC, and individual intakes vary as much as almost four-fold between occasions.

When examined by gender, age and energy intake, there is no important variation by gender for BC, in either the mean agreement or LOA (Table 4). Although the mean agreement for BC in women is no longer significantly different from 100%, this is possibly due to their smaller

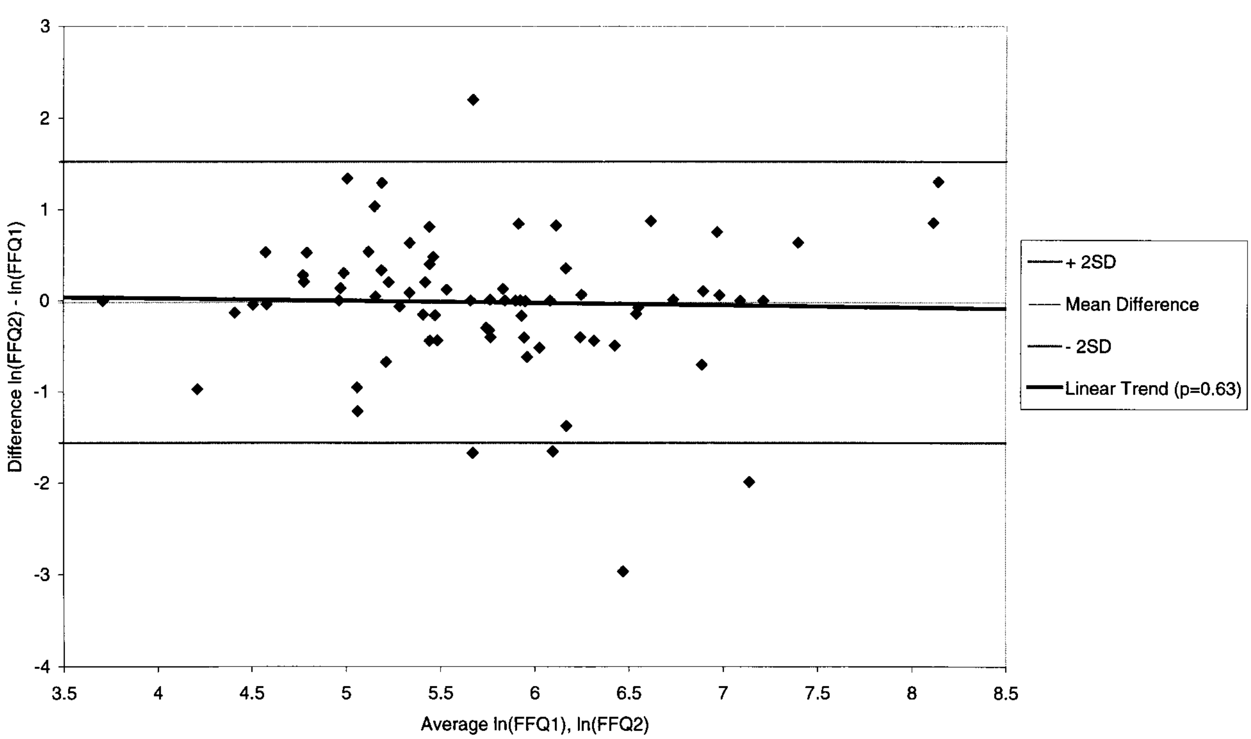

Fig. 2 Agreement between FFQ2 and FFQ1, dietary retinol (μg)

https://doi.org/10.1079/PHN200062 Published online by Cambridge University Press

Table 4 Intra-method agreement, limits of agreement (LOA) and correlation coefficients between FFQ1 and FFQ2

	N	Mean agreement* (%) (95% CI)	LOA† (%)	*r*(95% CI)‡
Dietary beta carotene				
Non-stratified	76	120 (108–133)	47–306	0.59 (0.42–0.72)
Men	55	120 (106–136)	48–301	0.60 (0.40–0.75)
Women	21	120 (97–147)	47–308	0.49 (0.08–0.76)
<55 years of age	35	106 (101–112)	78–146	0.70 (0.48–0.84)
55+ years of age	41	111 (102–120)	68–180	0.49 (0.21–0.69)
Low energy intake	25	116 (95–143)	42–326	0.55 (0.18–0.79)
Mid energy intake	25	131 (107–161)	47–371	0.44 (0.03–0.72)
High energy intake	26	113 (96–131)	50–255	0.72 (0.45–0.87)
Dietary retinol				
Non-stratified	76	98 (82–117)	21–459	0.66 (0.51–0.77)
Men	55	102 (83–126)	22–474	0.62 (0.48–0.99)
Women	21	88 (64–121)	20–383	0.73 (0.44–0.88)
<55 years of age	35	98 (90–107)	59–163	0.68 (0.45–0.83)
55+ years of age	41	101 (88–116)	45–229	0.56 (0.31–0.74)
Low energy intake	25	80 (54–117)	11–554	0.69 (0.39–0.86)
Mid energy intake	25	121 (91–162)	29–510	0.65 (0.33–0.84)
High energy intake	26	98 (76–125)	28–338	0.62 (0.28–0.82)

* Exp(mean (FFQ2 − FFQ1)), i.e. multiple of FFQ1.
† Mean differences $\pm t_{(n-1,\ 0.025)}$ (standard deviation of differences).
‡ Pearson's correlation coefficient.

numbers. Mean retinol agreement in women was lower than in men, although this was not significant, and there was marked reduction in the width of their LOA. Stratification by age reduced the bias for BC by moving mean agreement towards 100% for both age groups and substantially reducing the width of the LOA. For retinol, stratification by age also led to a substantial improvement in the LOA. Compared with non-stratified results, stratification by energy intake had a variable effect on the mean agreement but did not reduce the width of LOA for either nutrient.

The simple correlations for the unstratified and stratified results were all statistically significant (Table 4). However, the variation in the correlations did not necessarily reflect the changes in agreement. For example, the correlations for BC in women and those aged 55 years or more are essentially the same, and yet the mean agreement and LOA for the older group is much better than that for women. For retinol, the highest correlation occurs in the low-energy group, which has the most biased mean agreement and the largest degree of variation in agreement, as shown by the wide LOA. Correlations in the vicinity of 0.60 would often be regarded as indicating quite good repeatability. The extra information provided by the limits of agreement calculations shows that these can be obtained when individuals report a five-fold variation in intake between repeated estimations.

Inter-method reliability

Even after log transformation of the data, there was a pronounced dependency between retinol intake and the mean difference in retinol intakes between the FFQ2 and the DR; there was a similar but weaker pattern with BC intakes (Figs 3 and 4). On average, FFQ2 gave significantly higher intakes of BC and significantly lower intakes of retinol than did the DR (Table 5). Individual differences between the DR and FFQ2 varied as much as almost seven-fold for BC and six-fold for retinol (Table 5).

Stratifying by gender improved the mean agreement for BC intake in women but substantially worsened the LOA, and slightly improved the LOA in men. Stratification by gender did not alter mean retinol agreement but did improve the LOA in women. Stratification by age or energy intake had a small effect on mean retinol agreement and highly variable effects on the width of LOA for both BC and retinol.

The unstratified and many of the stratified correlation coefficients were statistically significant (Table 5). Again the patterns between the correlation coefficients do not necessarily reflect the patterns seen in mean agreement. For example, in energy strata a very poor correlation is seen with retinol in the highest tertile of energy intake and yet this is not the group with the widest LOA. The correlations for retinol by age are substantially different even though the agreement results are almost identical. Similarly for BC, correlation

https://doi.org/10.1079/PHN200062 Published online by Cambridge University Press

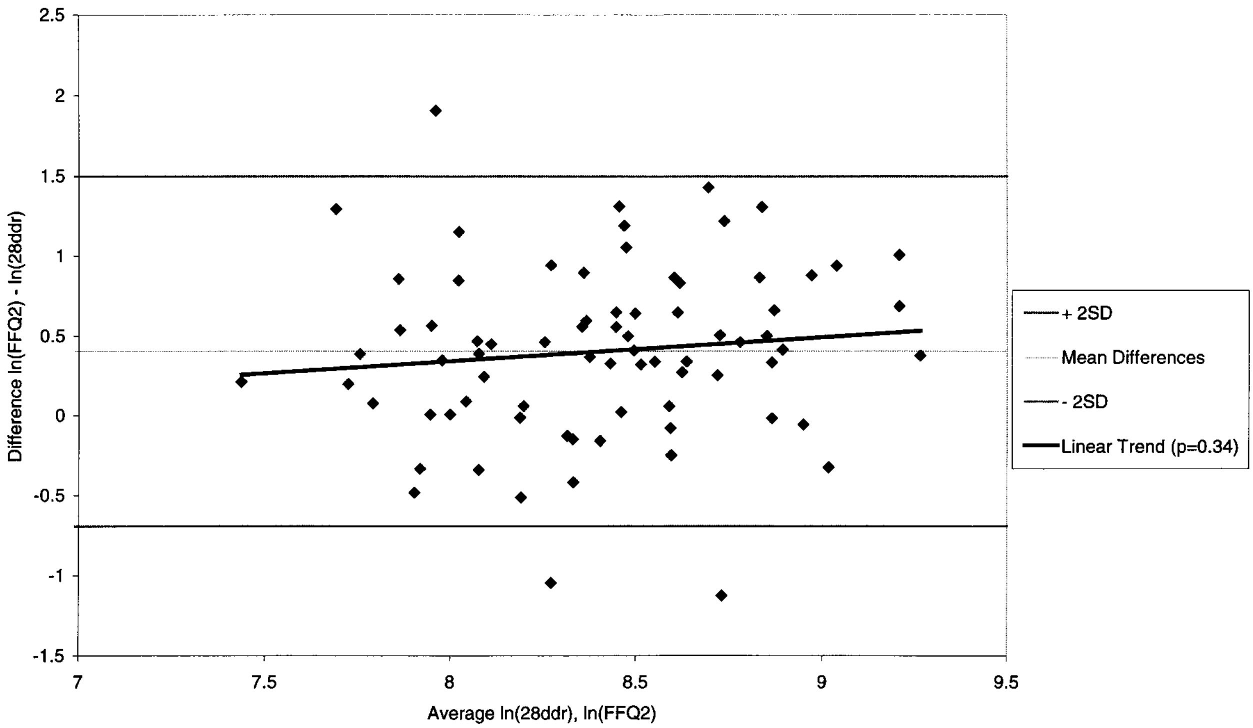

Fig. 3 Agreement between FFQ2 and four 1-week diet records, dietary beta carotene (mg)

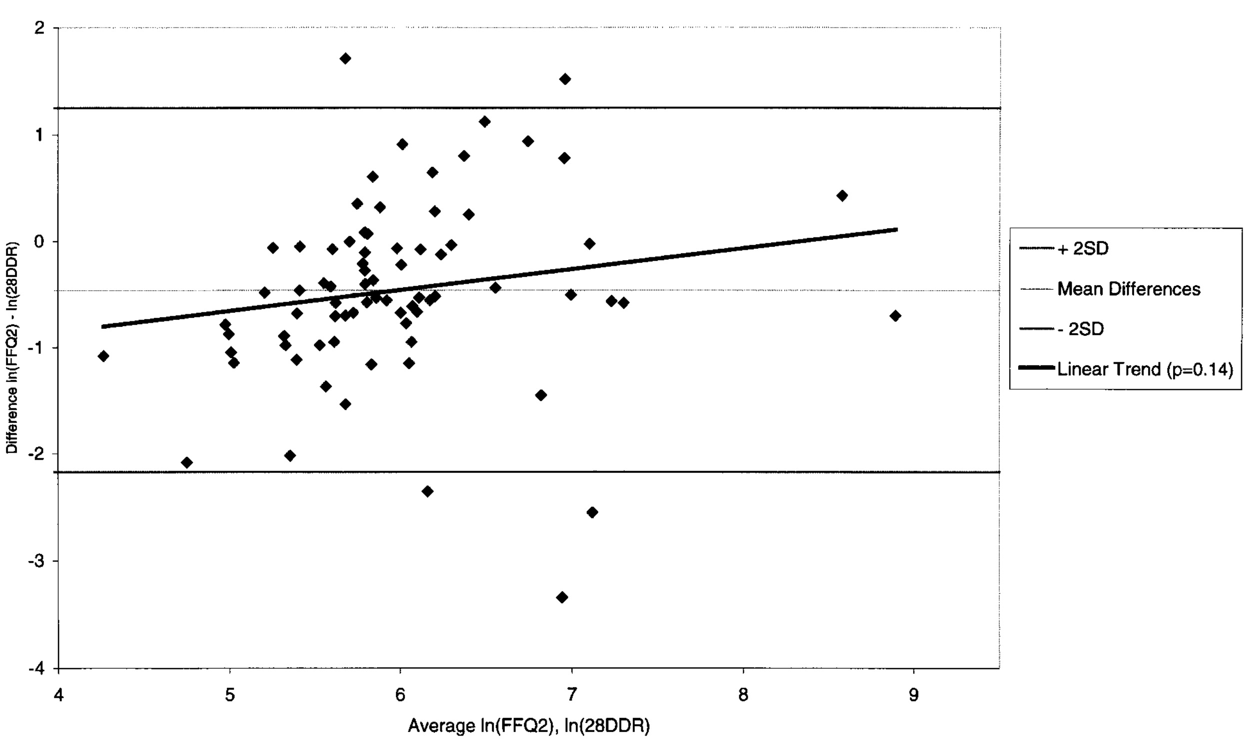

Fig. 4 Agreement between FFQ2 and four 1-week diet records, dietary retinol (μg)

https://doi.org/10.1079/PHN200062 Published online by Cambridge University Press

Table 5 Inter-method agreement, limits of agreement (LOA) and correlation coefficients between FFQ2 and four 1-week diet records

	N	Mean agreement* (%) (95% CI)	LOA† (%)	*r* (95% CI)‡
Dietary beta carotene				
Non-stratified	76	149 (132–170)	50–447	0.36 (0.14–0.54)
Men	55	156 (138–177)	61–398	0.49 (0.26–0.67)
Women	21	133 (99–177)	34–526	0.04 (−0.41–0.46)
<55 years of age	37	144 (119–174)	47–444	0.31 (−0.01–0.58)
55+ years of age	39	142 (118–169)	47–428	0.38 (0.07–0.62)
Low energy intake	25	147 (108–201)	31–694	0.10 (−0.33–0.49)
Mid energy intake	25	153 (128–183)	62–373	0.46 (0.06–0.73)
High energy intake	26	150 (126–175)	64–345	0.68 (0.38–0.85)
Dietary retinol				
Non-stratified	76	63 (52–77)	11–349	0.51 (0.32–0.66)
Men	55	63 (50–80)	11–367	0.48 (0.24–0.66)
Women	21	63 (46–87)	15–278	0.59 (0.26–0.67)
<55 years of age	37	67 (51–87)	13–331	0.39 (0.08–0.64)
55+ years of age	39	60 (45–80)	10–370	0.57 (0.31–0.75)
Low energy intake	25	68 (45–103)	8–573	0.53 (0.15–0.77)
Mid energy intake	25	71 (56–89)	22–223	0.76 (0.52–0.89)
High energy intake	26	52 (36–75)	8–329	0.12 (−0.30–0.50)

* Exp (mean (FFQ2 − DR)), i.e. multiple of DR.
† Mean differences ± $t_{(n-1,\ 0.025)}$ (standard deviation of differences).
‡ Pearson's correlation coefficient.

coefficients for mid and high tertiles of energy intake are very different, but agreement is practically the same for these two levels. It is clear that if correlations had been used to evaluate differences between strata, different conclusions would have been drawn in a number of instances about the comparison of methods, than if the LOA information were used.

Discussion

Bland and Altman were the first to illustrate poor agreement in a study that had reported strong correlation coefficients[5]. The LOA method clearly identifies the shortcomings of the FFQ in this study in comparison with the correlation coefficients, which give no information about potential errors or the directions of bias. Where there was significant over-estimation and under-estimation of intakes, positive correlation coefficients were significant in almost every case. Correlation coefficients also varied considerably across gender, age and thirds of energy intake, but this was not reflected in the limits of agreement findings. While the LOA method does not provide a single number representing agreement, but requires interpretation of the mean bias, direction of bias and the LOA, the correlation coefficient does not provide any such detail. The LOA approach reveals that a questionnaire's performance may not be assessed so highly if depending on correlation coefficients to indicate intra-method or inter-method reliability.

Correlation coefficients were calculated in this study to illustrate their limitations in the evaluation of reliability, but are comparable to those found in other studies. Studies comparing a repeated food frequency questionnaire report correlation coefficients ranging from 0.58 to 0.66 for beta carotene, and 0.57–0.66 for retinol[18–20]. Others comparing a diet record of at least 4 days with a food frequency questionnaire report correlation coefficients from 0.34 to 0.48 for beta carotene, and 0.22–0.44 for retinol[18–21]. This implies that the proportion of variance common to both true intake and the questionnaire appears to be similar across studies. However, we have shown that similar correlation coefficients can be obtained when other measures of agreement differ substantially (Tables 4 and 5). It is therefore not possible to say whether our questionnaire has a comparable level of agreement with other questionnaires.

It is difficult to compare the limits of agreement found in this study with others, as we are aware of only one other published study comparing a food frequency questionnaire with a diet record using Bland and Altman's method. However, this did not calculate limits of agreement for beta carotene or retinol intakes[22]. 'Agreement' is commonly reported in terms of the significance of the mean of differences, the median difference or the mean percent difference, e.g. mean{(FFQ − DR)/DR}, using non log-transformed data. A misleading impression of agreement may result from these methods, as these indices of overall bias may be close to zero even when the limits of agreement are very wide.

The response rate for participation in this reliability

https://doi.org/10.1079/PHN200062 Published online by Cambridge University Press

study appears low (21%). However, many dietary studies do not report how many people they attempted to recruit and often recruit subjects who are not part of the main study, which can lead to uncertainty about the generalizability of the results obtained. Volunteer samples are thought suitable for validation studies as long as they do not differ from the source population in age or gender[22]. There was no substantial evidence of a healthy 'volunteer effect' that could be shown by greater supplement use, fewer smokers or greater dietary intakes of vitamin A among our reliability study subjects[23]. For the purpose of validating the questionnaire for use within the population it was designed for, the sample captured is suitably representative of the trial subjects and there is no reason for not generalizing the findings of this reliability study to the overall study.

The LOA findings show that mean FFQ intakes estimated on two occasions were not necessarily the same. Recording DRs between FFQ1 and FFQ2 may have affected the intra-method reliability of our FFQ in two ways. The process of recording the diet may have influenced the reporting of intake through attention bias[24] which often leads to under-estimation of true consumption[25]. Alternatively, genuine changes in the diet may have occurred due to a heightened awareness of dietary patterns. Both would contribute to reducing the reliability of the repeated questionnaire, which was completed almost immediately after the final DR.

It is possible that induction into the trial (after completing FFQ1) may have encouraged dietary alterations. Smoking cessation was discussed where appropriate and dietary advice was freely available. As a consequence, subjects may have become more aware of the importance of diet and the cessation of smoking for the prevention of malignancies, both of which may lead to changes in dietary intake. In view of these likely interventions, the reliability of the FFQ may have been better examined later in the course of the trial, after efforts to change dietary patterns initiated either by participating in the trial or by recording diet records may have occurred, or been less likely. However, this would have had no effect on the inconsistencies found between correlation coefficients and LOA findings.

Over-estimation by food frequency questionnaires in comparison with other dietary methods has been cited in many other studies[19,26–28], particularly for Vitamins A and C, and fruits and vegetables[19,27], and is a probable consequence of the long lists of fruits and vegetables that are often included in them[27]. The majority of foods in this FFQ are beta carotene sources, namely fruits and vegetables, which may have contributed to the over-estimation of beta carotene. Fewer foods in the questionnaire represent retinol sources, which may have limited the opportunity for subjects to report their usual retinol intake, leading to greater under-estimation. Wider LOA among those with low energy intakes in their DR highlights the difficulty of estimating truly lower retinol intakes, and supports the likelihood that retinol sources in the questionnaire under-represented usual intake.

No two dietary methods can possibly be in perfect agreement because of the inherent variation in the human diet and the practical difficulties in measuring diet. Although anti-logging the differences between methods tends to over-estimate the LOA which at least reduces the likelihood of accepting poor methods of measurement[6], it is clear that the FFQ studied here significantly over-estimates beta carotene and under-estimates retinol dietary intakes, and the degree to which this occurs varies across levels of intakes. This differential bias will impact on predictions of diet–disease relationships. The correlation coefficients are unable to show this.

There have been many terms used for validity, reliability, agreement and reproducibility that have been interchangeably associated with different statistical procedures. The term 'agreement' has been used loosely to describe the misclassification of individuals, the differences between means, and Pearson, intra-class and regression correlation coefficients. These statistical functions actually represent different capabilities of the methods being examined. The universal adoption of standardized definitions and terminology for validation techniques would help reduce the use of inappropriate statistics and support suitable comparisons of validation studies[29]. These results not only emphasize the necessity of validating food frequency questionnaires but also, by using an alternative method, highlight the importance of the statistical method used to assess reliability and validity.

Acknowledgements

The authors wish to thank the study participants for their co-operation and the following staff for their assistance with data collection for this project: C. Hickling, I. Foundas, F. Smith, N. Olsen, L. Watts and J. Sleith. This study was supported by the Western Australian Government Worker's Compensation and Rehabilitation Commission and the Western Australian Cancer Foundation.

References

1 Willett WC. *Nutritional Epidemiology*. 2nd edn. New York: Oxford University Press, 1998.
2 Kaaks R, Riboli E. Validation and calibration of dietary intake measurements in the EPIC project: methodological considerations. *Int. J. Epidemiol.* 1997; **26**((Suppl. 1)): S15–25.
3 Kipnis V, Carroll RJ, Freedman LS, Li L. Implications of a new dietary measurement error model for estimation of relative risk: application to four calibration studies. *Am. J. Epidemiol.* 1999; **150**: 642–51.
4 Prentice RL. Measurement error and results from analytic epidemiology: dietary fat and breast cancer. *JNCI* 1996; **88**((23)): 1738–47.
5 Altman DB, Bland JM. Measurement in medicine: the analysis

https://doi.org/10.1079/PHN200062 Published online by Cambridge University Press

of method comparison studies. *The Statistician* 1983; **32**: 307–17.
6 Bland MJ, Altman DG. Statistical methods for assessing agreement between two methods of clinical assessment. *Lancet* 1986; (Feb 8): 307–11.
7 Hebert J, Miller D. The inappropriateness of conventional use of the correlation coefficient in assessing validity and reliability of dietary assessment methods. *Eur. J. Epidemiol.* 1991; **7**(4): 339–43.
8 Armstrong BK, White E, Saracci R. *Principles of Exposure Measurement in Epidemiology*. New York: Oxford University Press, 1992.
9 Chinn S. The assessment of methods of measurement. *Stat. Med.* 1990; **9**: 351–62.
10 Last JM. ed. *A Dictionary of Epidemiology*. 2nd edn. New York: Oxford University Press, 1988.
11 De Klerk NH, Musk AW, Ambrosini GL, Eccles JL, Hansen J, Olsen N, *et al.* Vitamin A and cancer prevention II: comparison of the effects of retinol and beta-carotene. *Int. J. Cancer* 1998; **75**: 362–7.
12 Musk AW, de Klerk NH, Ambrosini GL, Eccles JL, Hansen J, Olsen NJ, *et al.* Vitamin A and cancer prevention I: observations in workers previously exposed to asbestos at Wittenoom, Western Australia. *Int. J. Cancer* 1998; **75**: 335–61.
13 Block G, Dresser A, Hartman M, Carroll M. Nutrient sources in the American diet: quantitative data from the NHANES II survey. *Am. J. Epidemiol.* 1985; **122**: 13–26.
14 Cade J, Margetts B. Nutrient sources in the English diet: quantitative data from three English towns. *Int. J. Epidemiol.* 1988; **17**: 844–8.
15 Lewis J, Holt R. NUTTAB91–92: Commonwealth Government of Australia. Canberra: National Food Authority, 1992.
16 Goodsell C. Diet/1 Nutrient Analysis Software. Version 3.10. Brisbane: Xyris Software, 1991.
17 SAS System for Windows. Version 6.12. Calgary: SAS Institute, Inc, 1989–1996.
18 Kristal AR, Feng Z, Coates RJ, Oberman A, George V. Associations of race/ethnicity, education and dietary intervention with the validity and reliability of a food frequency questionnaire. *Am. J. Epidemiol.* 1997; **146**: 856–69.
19 Mannisto S, Virtanen M, Mikkonen T, Pietinen P. Reproducibility and validity of a food frequency questionnaire in a case–control study on breast cancer. *J. Clin. Epidemiol.* 1996; **49**(4): 401–9.
20 Rimm E, Giovannucci E, Stampfer M, Colditz G, Litin L, Willett W. Reproducibility and validity of an expanded, self-administered, semi-quantitative food frequency questionnaire among male health professionals. *Am. J. Epidemiol.* 1992; **135**: 1114–26.
21 Feskanich D, Marshall J, Rimm E, Litin L, Willett W. Simulated validation of a brief food frequency questionnaire. *Ann. Epidemiol.* 1994; **4**: 181–7.
22 Thompson RL, Margetts BM. Comparison of a food frequency questionnaire with a 10-day weighed record in cigarette smokers. *Int. J. Epidemiol.* 1993; **22**(5): 824–33.
23 Patterson RE, Neuhouser ML, White E, Hunt JR, Kristal AR. Cancer-related behaviour of vitamin supplement users. *Cancer Epidemiol. Biomark. Prev.* 1998; **7**(1): 79–81.
24 Lissner L, Heitman BL, Lindroos AK. Measuring intake in free-living human subjects: a question of bias. *Proc. Nutr. Soc.* 1998; **57**: 333–9.
25 Johansson L, Solvoll K, Bjorneboe GEA, Drevon CA. Under- and overreporting of energy intake related to weight status and lifestyle in a nationwide sample. *Am. J. Clin. Nutr.* 1998; **68**: 266–74.
26 Hebert JR, Thurley TG, Chiriboga DE, Barone J. A comparison of selected nutrient intakes derived from three diet assessment methods used in a low-fat maintenance trial. *Public Health Nutr.* 1998; **1**(3): 207–14.
27 Harlan L, Block G. Use of adjustment factors with a brief food frequency questionnaire to obtain nutrient values. *Epidemiology* 1990; **1**: 224–31.
28 Munger R, Folsom A, Kushi L, Kaye S, Sellers T. Dietary assessment of older Iowa women with a food frequency questionnaire: nutrient intake, reproducibility, and comparison with 24-hour dietary recall interviews. *Am. J. Epidemiol.* 1992; **136**: 192–200.
29 Kohlmeier L, Overview of validity, quality control and measurement error issues in nutritional epidemiology. *Eur. J Clin. Nutr.* 1993; **47**(Suppl. 2): S1–5.