

Validity and reproducibility of the food-frequency questionnaire used in the Shanghai Men's Health Study

Raquel Villegas¹, Gong Yang¹, DaKe Liu², Yong-Bing Xiang², Hui Cai¹, Wei Zheng¹ and Xiao Ou Shu^{1*}

¹Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt University Medical Center and Vanderbilt-Ingram Cancer Center, Medical Center East, Suite 6000, 1215 21st Avenue South, Nashville, TN 37232-8300, USA

²Department of Epidemiology, Shanghai Cancer Institute, 2200/25# Xie Tu Road, Shanghai 200032, PR China

(Received 12 May 2005 – Revised 27 November 2006 – Accepted 29 November 2006)

We evaluated the validity and reproducibility of the FFQ used in the Shanghai Men's Health Study (SMHS). The study included 195 randomly selected participants of the SMHS who completed one FFQ at baseline, twelve 24-hour dietary recalls (24-HDR) (once a month for twelve consecutive months) and a second FFQ at the end of the study. The FFQ accounted for 88.78% of the foods recorded in the 24-HDR surveys. The validity of the FFQ was evaluated by comparing nutrient and food group intake levels from the second FFQ and the multiple 24-HDR. Correlation coefficients ranged from 0.38 to 0.64 for macronutrients, 0.33 to 0.58 for micronutrients and 0.35 to 0.72 for food groups. Misclassification to opposite quartiles for nutrients and food groups was rare, ranging from 1.5 to 7.7%, while exact agreement rates were between 31.8 and 53.3%. The reliability of the FFQ was assessed by comparing the intake levels from the two FFQ. Correlation coefficients were 0.39 to 0.53 for macronutrients, 0.38 to 0.52 for micronutrients and 0.39 to 0.64 for food groups. Exact agreement rates for quartile distribution were between 31.8 and 49.2%, while misclassification to opposite quartiles was between 1.5 and 6.2%. These data indicate that the SMHS FFQ can reasonably categorise usual intake of nutrients and food groups among men living in urban Shanghai.

Food-frequency questionnaire: Validity: Reliability

Obtaining an accurate estimate of long-term habitual food intake remains the main challenge in diet–disease research. FFQ have been used as an epidemiological tool to assess diet for several decades (Willett, 1990). Because short-term recalls and diet records are generally expensive and unrepresentative of usual intake and not good for assessment of past diet, FFQ have been the primary method of dietary assessment for most epidemiological studies. FFQ are easy to administer and relatively inexpensive to use in large populations. However, they are very sensitive to the cultural and dietary practices of the population of concern (Sharma *et al.* 1996). Use of inappropriate food lists in the FFQ may result in underestimation of nutrients due to omission of key items (Sharma *et al.* 1996). Thus, the validity and reliability of an FFQ needs to be evaluated for the specific population of concern. In addition, the use of methods with low validity seriously attenuates the association between nutritional intake and disease in epidemiological studies, a problem known as regression dilution (Day *et al.* 2001).

Methods to assess and interpret the validity and reproducibility of FFQ have tended to rely on correlation analysis of nutrients and/or foods measured by two or more dietary assessment methods. As correlation coefficients do not measure agreement, Bland–Altman plots (Bland & Altman, 1986) have been used

in conjunction with correlation coefficients in some studies (Thompson & Margetts, 1993; Ambrosini *et al.* 2001; Flood *et al.* 2004). In many epidemiological studies, however, the main concern is to classify individuals into different groups according to exposure levels rather than to assess their absolute intake. Thus, comparisons of percentage agreement in quartile distributions are also often used in the evaluation of the validity and reliability of an FFQ.

In 2002, we launched a population-based cohort study of 61 582 men aged 40–74 years in Shanghai, China (Shanghai Men's Health Study (SMHS)) with a main study focus on diet and cancer risk. One of the objectives of the study is to evaluate associations between diet and chronic disease, mainly *via* categorising participants according to their intake of both nutrients and food groups. We used an FFQ that was developed based on a similar dietary questionnaire used in a sister study, the Shanghai Women's Health Study (SWHS). The SWHS FFQ has been validated previously (Shu *et al.* 2004). The SMHS FFQ includes 88.78% of all foods that were commonly consumed by men in Shanghai at the time of the baseline survey. We conducted a dietary calibration study to evaluate the validity and reliability of the FFQ in the SMHS population and we report the results of the study in this paper.

Subjects and methods

The Shanghai Men's Health Study

Recruitment for the SMHS started in April 2002 and was completed in June 2006. A total of 83 125 male residents of eight communities of urban Shanghai between the ages of 40 and 70 years were invited to participate by trained interviewers through in-person contact. A total of 61 582 participants were enrolled in the study with a response rate of 74.1%. Reasons for non-participation were refusal (21.1%), out of the area during enrolment (3.1%), and other miscellaneous reasons including poor health or hearing problems (1.7%).

Dietary calibration study

Before the SHMS began, a pilot, dietary validation study was conducted among ninety-six male residents of Shanghai to estimate the sources of variation in their dietary intake and to determine the number of 24-hour dietary recalls (24-HDR) and samples needed to evaluate the validity and reliability of the FFQ in the study population (Cai *et al.* 2005). We administered twenty-four 24-HDR to the ninety-six men over a 1-year period. We found intra-person variations to be the main contributor to dietary intake variation and that a validation study could be adequately carried out with 12 d of dietary recalls for a study of 100 or more participants (Cai *et al.* 2005).

The dietary calibration study was initiated on 14 November 2004. Study participants were a random sample of SMHS participants. A total of 214 SMHS participants were recruited from two SMHS communities approximately 2–3 months after they completed the baseline survey. Fifteen interviewers conducted the study, with each being responsible for the follow-up of seventeen subjects throughout the 12-month study period. The study communities were chosen based on neighbourhood proximity to the residence of the interviewers. In total, seventeen primary and sixty-eight alternative contacts were identified for possible recruitment by each interviewer.

Among participants in this study, approximately 69.3% were primary contacts. Study participants were contacted once a month during the 12-month period of the study to provide the name and amount of foods that they consumed over the preceding 24 h. The days that the 24-HDR were administered were chosen to ensure a balanced representation of weekdays and weekend days for each participant. All recalls were obtained by an unannounced in-person interview in the evening after dinner (around 19.00 h). At the end of the 12-month study, a second FFQ was administered. The second FFQ was completed by 196 subjects (with an average time interval between the administration of two FFQ of 1.2 years, range: 1.1–2.1 years).

Of the 196 participants with two FFQ and at least ten 24-HDR, there was one participant with an implausible average daily energy intake (9432 kcal/d). This subject was excluded from the analyses, resulting in 195 participants for the analysis.

Food-frequency questionnaire

The FFQ used in the SMHS was developed based on a similar dietary questionnaire that was used in a sister study, the

SWHS. The SWHS FFQ has been validated previously (Shu *et al.* 2004). A total of eighty-one food items were included in the FFQ used in the SMHS. For each food item or food group, subjects were asked how frequently (daily, weekly, monthly, yearly or never) they consumed the food or food group, which was followed by a question on the amount consumed in lians per unit of time. Lian is a unit of weight in China (1 lian = 50 g). The main purpose of this FFQ is to rank individuals along the distribution of dietary nutrient and food intake, so that individuals with low intake can be separated from those with high intake.

Statistical analysis

The Chinese Food Composition Tables (Yang *et al.* 2002) were used to estimate the intake levels of major nutrients for study participants. Food groups were formed by combining the intake levels of selected foods with similar nutrients, phytochemicals or botanic classifications.

The validity and reliability of the FFQ were assessed by comparing the median nutrient and food intakes, agreement of quartile distribution of nutrients and food groups, and calculating correlations between the intakes derived from the two different dietary survey methods (FFQ versus 24-HDR) and the two different surveys (baseline and second FFQ). The nutrients and food groups were not normally distributed, and log transformation failed to normalise the distribution. Therefore, Spearman correlation coefficients were applied for the analysis. Virtually all the correlation coefficients presented herein were statistically significant with a *P* value < 0.01, and are not individually presented.

We also analysed the data using Bland–Altman plots. The Bland–Altman plot assesses the agreement between two dietary assessment methods across a range of intakes. The difference between the two methods was plotted against the average of the two methods. Natural-log (ln) transformations were performed in order to narrow the limits of agreement (LOA), as recommended by Bland & Altman (1986). The antilogs of the LOA were calculated yielding a ratio of FFQ over 24-HDR. The ratios were multiplied by 100, with 100% representing ideal agreement.

Results

Participants in the dietary calibration study did not differ from the entire SMHS with regard to age, occupation, education, smoking status, alcohol consumption status, exercise reporting, weight, BMI, waist-to-hip ratio, waist circumference, daily energy intake or macronutrient intake (fat, protein, fibre and carbohydrates) at the baseline survey (Table 1). The dietary data shown in this table are from the information collected in the FFQ administered at study recruitment (baseline survey). Among the 195 dietary calibration study participants, three subjects completed eleven 24-HDR, while 192 subjects completed twelve 24-HDR. This resulted in a total of 2337 24-HDR. These dietary recalls represent 36 783 food entries and 655 unique food items. The FFQ accounted for 88.78% of the total food entries (some as food groups) recorded in the 24-HDR.

Table 2 presents the median nutrient intake derived from the baseline, second FFQ, average of the 24-HDR and the

Table 1. Comparison of participants in the validation study with all cohort members – Shanghai Men's Health Study, 2003–2005

	Calibration study (<i>n</i> = 195)	SMHS participants (<i>n</i> = 61 581)	<i>P</i> value
kcal/d*	1967.6	1908.6	0.09
Fat*	36.0	34.6	0.22
Protein*	81.1	78.3	0.11
Carbohydrates*	329.7	320.8	0.15
Fibre*	12.1	11.5	0.06
Age (recruitment)	54.8	54.9	0.98
Weight (kg)	69.0	68.4	0.45
BMI	23.8	23.7	0.53
Waist (cm)	85.1	85.1	0.99
WHR	0.90	0.90	0.76
Education			
None	7.7	6.8	0.84
Middle school	35.4	33.5	
High school	33.3	36.0	
College	23.6	23.6	
Occupation			
Professional	27.2	26.6	0.60
Clerical	19.0	21.9	
Farmer	53.8	51.5	
% current smokers	60.5	58.6	0.59
% current drinkers	29.7	29.3	0.89
(%) exercise	40.5	35.6	0.15

SMHS, Shanghai Men's Health Study; WHR, weight-to-height ratio.

* From the FFQ.

percentage of differences. The median nutrient intakes assessed by the two FFQ agree considerably well, especially for macronutrients, with the differences in median intakes between the two assessments being 5.2% ($P=0.13$) for fat, 3.4% ($P=0.17$) for protein and 2.5% ($P=0.42$) for carbohydrates. The median differences in consumption of micronutrients are all under 9% ($P>0.05$ for all nutrients). The nutrient intakes assessed on the second FFQ are, in general, lower than those assessed at the baseline FFQ, with the exception of fibre, vitamins B₁ and B₂, niacin and vitamin C consumption, which were 1.7, 1.8, 1.0, 0.6 and 1%, respectively, higher on the second FFQ. There is also good agreement between nutrient intake assessed by the second FFQ and by the average of the 24-HDR (both assessments cover the same time period). Differences in median intake were between 4.4 (protein, $P=0.04$) and 20.6% (fibre, $P<0.001$) for macronutrients and between 0.4 (niacin) and 31.8% (vitamin C) for micronutrients ($P<0.05$ for micronutrients with the exception of vitamin B₂ and retinol). Compared with the 24-HDR, the FFQ tends to overestimate intake of most nutrients except for the consumption of protein, fat and niacin.

The nutrient intakes assessed by the two FFQ approximately 1.2 years apart correlate reasonably well; the correlation coefficient (r) for nutrient intake ranges from 0.38 to 0.53 (Table 3). The correlations of nutrient intake between the second FFQ and the 24-HDR are between 0.33 (retinol) and 0.64 (carbohydrates).

Overall agreement between nutrient intake measured by the FFQ and the 24-HDR was assessed using Bland–Altman plots. The Bland–Altman plots showed that the difference in nutrient intake derived from the FFQ and 24-HDR did not appear to depend on the 'true' intake as assessed by the

Table 2. Median comparison of nutrient intake between two FFQ and average of 24-HDR among 195 Chinese men, Shanghai Men's Health Study

	Median (25–75th percentile)				Wilcoxon test <i>P</i> values		Percentages of median difference	
	FFQ1		FFQ2		FFQ1 versus FFQ2	24-HDR-FFQ2	FFQ1 versus FFQ2	24-HDR-FFQ2
	Median	25–75th percentile	Median	25–75th percentile	FFQ2	24-HDR-FFQ2	(%)*	(%)†
kcal/d	1929.5	1607.1–2269.7	1852.9	1557.9–2153.3	0.25	0.10	4.1	-3.1
Protein	76.3	63.1–95.8	73.7	62.0–87.2	0.17	0.04	3.4	-4.4
Fat	32.5	26.5–43.7	30.9	25.0–38.4	0.13	<0.001	5.2	-20.3
Carbohydrates	318.5	263.4–386.3	310.6	260.3–377.7	0.42	0.11	2.5	3.7
Fibre	11.5	8.9–14.5	11.7	8.9–13.8	0.57	<0.001	-1.7	20.6
Retinol	146.8	87.0–205.6	140.6	95.3–195.5	0.87	0.99	4.4	1.4
Carotene	3168.8	2311.9–4399.0	3017.6	2166.0–4159.6	0.49	<0.001	5.0	27.3
Vitamin B ₁	1.1	0.9–1.3	1.07	0.88–1.29	0.52	0.01	-1.8	11.2
Vitamin B ₂	1.0	0.8–1.2	0.97	0.78–1.17	0.68	0.84	-1.0	2.1
Niacin	15.7	13.1–19.6	15.8	12.8–18.5	0.28	0.03	-0.6	-0.4
Vitamin C	95.7	69.6–137.9	96.6	69.2–136.4	0.81	<0.001	-1.0	31.8
Vitamin E	14.9	11.0–18.9	13.7	10.6–16.8	0.07	<0.01	8.7	31.1

24-HDR, 24-hour dietary recall.

* FFQ1 – FFQ2/FFQ2.

† FFQ2 – 24-HDR/24-HDR.

Table 3. Spearman correlation and percentage agreement in quartile distribution of nutrient intake between the two FFQs and second FFQ with average 24-HDR, Shanghai Men's Health Study

	Percentage agreement									
	Correlation		Same quartile		Adjacent quartile		One quartile apart		Opposite quartile	
	FFQ1 versus FFQ2	24-HDR-FFQ2	FFQ1 Versus FFQ2	24-HDR-FFQ2	FFQ1 versus FFQ2	24-HDR-FFQ2	FFQ1 versus FFQ2	24-HDR-FFQ2	FFQ1 versus FFQ2	24-HDR-FFQ2
Kcal/d	0.49	0.58	43.1	44.1	41.1	42.0	12.3	11.3	3.5	2.6
Protein	0.47	0.49	40.0	35.3	37.5	45.8	19.0	14.3	3.5	4.6
Fat	0.39	0.38	33.3	35.9	44.7	40.6	16.4	17.4	5.6	6.1
Carbohydrates	0.53	0.64	41.0	46.7	42.6	40.5	13.3	10.8	3.1	2.0
Fiber	0.49	0.48	40.1	37.9	38.9	40.6	17.4	17.9	3.6	3.6
Retinol	0.45	0.33	43.6	33.3	32.8	42.1	19.5	16.9	4.1	7.7
Carotene	0.40	0.38	37.9	38.5	38.0	35.9	20.5	21.0	3.6	4.6
Vitamin B ₁	0.52	0.58	39.5	42.0	43.1	41.6	15.4	14.9	2.0	1.5
Vitamin B ₂	0.47	0.46	37.4	39.0	40.0	42.1	20.0	13.8	2.6	5.1
Niacin	0.44	0.52	31.8	38.5	45.6	38.4	19.5	20.5	3.1	2.6
Vitamin C	0.38	0.42	35.9	37.4	38.0	39.0	21.5	18.5	4.6	5.1
Vitamin E	0.48	0.38	36.9	31.8	40.6	48.7	21.0	12.8	1.5	6.7

24-HDR, 24-hour dietary recall.

24-HDR for either macronutrient or micronutrients. The Bland–Altman plots for total energy intake, fat, protein and carbohydrates are presented in Figs 1–4. Anti-logging rendered mean agreement and the LOA of 96.3% (95% CI 64–143), 82% (95% CI 39–172), 95.6% (95% CI 61–149) and 103.9% (95% CI 70–155), respectively, for kcal/d, fat, protein and carbohydrates. This suggest that, on average, the FFQ underestimates intake of energy, fat and protein by 3.7, 18 and 4.4%, respectively, and overestimates carbohydrate intake by 3.9%, as compared with that derived from multiple 24-HDR.

When the nutrient intakes were categorised into quartiles, the range of agreement rates for same quartile classifications were 31.8–43.6% for nutrients derived from the two FFQ and 31.8–46.7% for nutrients derived from the second FFQ and the 24-HDR. The agreement rates for classifying nutrient intakes into the same or adjacent quartiles were between 73.9 and 84.2% for nutrients derived from the two FFQ and between 74.4 and 87.2% for nutrients derived from the second FFQ and the 24-HDR. Misclassification of nutrient intake into extreme quartiles was rare (1.5–7.7%).

Similar sets of analyses were conducted for selected food groups (Tables 4 and 5). With the exception of fish intake, the differences in consumption of food groups between the two FFQ were within 10%. Intake of poultry, red meat, fish and eggs was lower according to the FFQ as compared with the 24-HDR, while the intake of other food groups was higher according to the FFQ as compared with the 24-HDR ($P > 0.05$ for all food groups). Intake estimates were 28.9% higher for vegetables, 77.7% for soy and 57.4% for fruits. Consumption was lower on the FFQ for poultry by 51.1%, red meat by 35.4%, fish by 21.0% and eggs by 4.9%. Despite the difference in median intake, the correlations for food group intake ranged from 0.39 to 0.64 when comparing the two FFQs and from 0.35 to 0.72 when comparing the second FFQ and the 24-HDR. Correlation coefficients between the two FFQ were 0.50 for soy, 0.43 for vegetables, 0.64 for fruits, 0.48 for poultry, 0.40 for red meat and 0.39 for eggs (Table 5). Correlation coefficients comparing the second FFQ and the 24-HDR were 0.54 for soy, 0.42 for vegetables, 0.72 for fruits, 0.35 for poultry, 0.45 for red meat and 0.41 for eggs. The agreement rates for food group consumption

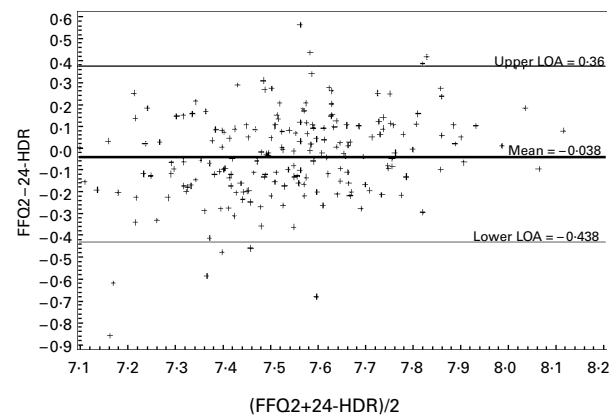


Fig. 1. Bland–Altman plot for total energy intake. LOA, limits of agreement; 24-HDR, 24-hour dietary recall.

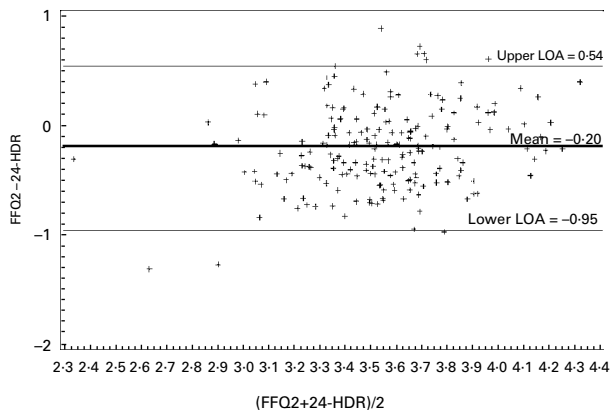


Fig. 2. Bland–Altman plot for fat intake. LOA, limits of agreement; 24-HDR, 24-hour dietary recall.

classified in the same quartile were 33.3–49.2% for the two FFQ and 32.8–53.3% for the second FFQ and the 24-HDR. The agreement rates were between 73.8 and 91.8% for food group consumption classified into the same or adjacent quartiles. Only 1.5–6.2% of subjects were misclassified into extreme quartiles of food group intake.

Discussion

This report describes the validity and reproducibility of an FFQ designed to capture the usual intake of nutrients and major foods consumed by men in urban Shanghai. The reference method was repeated monthly 24-HDR that were conducted over a 12-month period. We evaluated the performance of the FFQ by comparing intake of nutrients and selected food groups obtained from this instrument with those derived from the 24-HDR. The results suggest that the SMHS FFQ has reasonable comparative validity and reproducibility, and can categorise major nutrient and food group intakes with relative accuracy among men in urban Shanghai. The performance of the SMHS FFQ is similar to that of many FFQ that have been used in other epidemiological studies (Willett *et al.* 1985; Kaaks *et al.* 1997; Margetts & Pietinen,

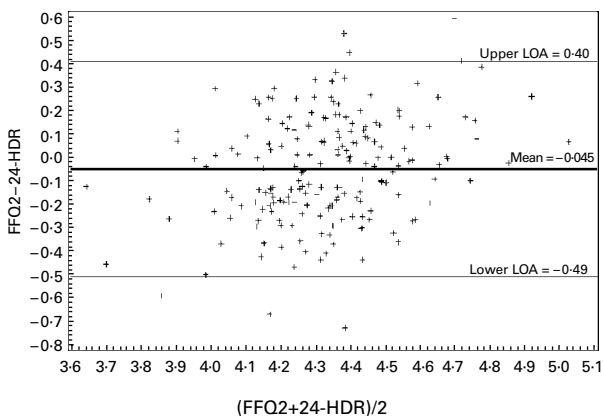


Fig. 3. Bland–Altman plot for protein intake. LOA, limits of agreement; 24-HDR, 24-hour dietary recall.

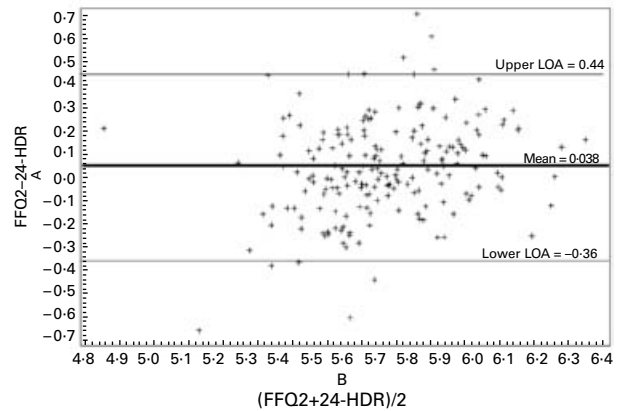


Fig. 4. Bland–Altman plot for carbohydrate intake. LOA, limits of agreement; 24-HDR, 24-hour dietary recall.

1997; Martinez *et al.* 1999; Mayer-Davis *et al.* 1999; Stram *et al.* 2000).

The validity and reproducibility of the FFQ used by the Italian arm of the European Prospective Investigation into Cancer and Nutrition (EPIC) were evaluated in a way similar to our study (Pisani *et al.* 1997). Two FFQ were administered 1 year apart and between eight and fourteen 24-HDR interviews were administered over a 1-year period. Pearson correlation coefficients of the relationship between questionnaire measurements and the individual average 24-HDR ranged from 0.28 for fat to 0.58 for carbohydrates for men. In the German arm of the EPIC study, correlation coefficients between the FFQ and monthly 24-HDR over 1 year were 0.37 for fat, 0.41 for protein, 0.41 for carbohydrates and 0.46 for fibre (Bohlscheid-Thomas *et al.* 1997). Our results are very similar to those of the two studies cited above, with correlation coefficients between the FFQ and 24-HDR being 0.38, 0.49, 0.48 and 0.64 for fat, protein, fibre and carbohydrates, respectively. The reproducibility correlation coefficients for two FFQ administered 6 months apart were between 0.59 and 0.69 for macronutrients in the German arm of the EPIC study (Bohlscheid-Thomas *et al.* 1997), while in our study the reproducibility for macronutrients derived from two FFQ administered 1–2 years apart is between 0.39 and 0.53.

In the Male Professionals Health Study, the validity and reproducibility of the FFQ was evaluated in 127 men by two 1-week dietary records and two FFQ given 1 year apart (Rimm *et al.* 1992). Correlation coefficients for log-transformed nutrient intake derived from the two FFQ administered 1 year apart ranged from 0.47 to 0.72 for micronutrients (without supplements) and from 0.59 to 0.69 for macronutrients, while in our study correlation coefficients from the two FFQ ranged between 0.39 and 0.53 for macronutrients and between 0.38 and 0.53 for micronutrients. The correlation coefficients for nutrients derived from the second FFQ and the mean of the two 1-week dietary records were between 0.28 and 0.64 for micronutrients (without supplements) and between 0.25 and 0.63 for macronutrients. Our results compare well with those of the Male Professionals Health Study, with correlation coefficients between the FFQ and 24-HDR being between 0.38 and 0.52 for micronutrients and between 0.38 and 0.64 for macronutrients.

Table 4. Median comparison of food group daily intake between the two FFQ and average of the 24-HDR, Shanghai Men's Health Study

	Median (25–75th percentile)						Wilcoxon test <i>P</i> values		Percentages of median difference	
	FFQ1		FFQ2		24-HDR		FFQ1 versus FFQ2	24-HDR-FFQ2	FFQ1 versus FFQ2 (%) [*]	24-HDR-FFQ2 (%) [†]
	Median	25–75th percentile	Median	25–75th percentile	Median	25–75th percentile				
Rice (g)	364.3	300–447.1	357.1	300.0–452.0	333.5	286.4–387.2	0.53	<0.01	2.0	7.1
Poultry (g)	10.8	4.3–21.7	10.8	4.5–17.3	22.1	11.3–35.2	0.41	<0.01	0.0	–51.1
Red meat (g)	55.4	37.7–82.9	52.9	37.8–77.3	81.9	59.1–104.2	0.46	<0.01	4.7	–35.4
Fish (g)	43.3	21.8–70.0	34.9	20.7–55.1	44.2	30.9–75.2	0.08	<0.01	24.0	–21.0
Eggs (g)	25.1	12.6–44.0	25.1	12.6–44.0	26.4	15.3–38.5	0.67	0.28	0.0	–4.9
Vegetables (g)	348.8	248.0–508.8	335.3	241.4–509.1	260.0	186.8–334.8	0.69	<0.01	4.0	28.9
Soy (g)	139.5	78.8–218.4	127.6	77.5–221.4	71.8	45.2–115.6	0.40	<0.01	9.3	77.7
Fruits (g)	132.3	43.4–232.7	126.9	47.0–200.4	80.6	39.6–138.8	0.28	<0.01	4.2	57.4

24-HDR, 24-hour dietary recall.

^{*} FFQ1 – FFQ2/FFQ2.[†] FFQ2 – 24-HDR/24-HDR.**Table 5.** Spearman correlations and percentage agreement in quartile distribution of food group intake between the two FFQs and the second FFQ with average 24-HDR, Shanghai Men's Health Study

	Correlation		Percentage agreement							
	FFQ1 versus FFQ2	24-HDR-FFQ2	Same quartile		Adjacent quartile		One quartile apart		Opposite quartile	
			FFQ1 versus FFQ2	24-HDR-FFQ2	FFQ1 versus FFQ2	24-HDR-FFQ2	FFQ1 versus FFQ2	24-HDR-FFQ2	FFQ1 versus FFQ2	24-HDR-FFQ2
Rice (g)	0.52	0.63	38.5	44.1	43.1	41.6	16.4	12.8	2.0	1.5
Poultry (g)	0.48	0.35	40.5	32.8	37.4	42.7	18.5	18.9	3.6	5.6
Red meat (g)	0.40	0.45	33.8	35.9	43.1	42.6	18.5	16.9	4.6	4.6
Fish (g)	0.41	0.49	41.2	37.9	36.2	44.5	16.4	12.3	6.2	5.3
Eggs (g)	0.39	0.41	44.1	35.3	31.3	38.5	20.5	22.6	4.1	3.6
Vegetables (g)	0.43	0.42	33.3	33.8	43.6	47.7	19.5	14.9	3.6	3.6
Soy (g)	0.50	0.54	43.1	43.1	36.9	39.4	17.4	15.4	2.6	2.1
Fruits (g)	0.64	0.72	49.2	53.3	39.0	38.5	9.2	6.7	2.6	1.5

24-HDR, 24-hour dietary recall.

The validity and reproducibility of FFQs in multiethnic populations have also been investigated (Mayer-Davis *et al.* 1999; Stram *et al.* 2000). In a multiethnic cohort study in Hawaii and Los Angeles, dietary information reported from a questionnaire was compared with three 24-HDR in a calibration substudy. Subjects from each of eight subgroups defined by sex and ethnic group (African-American, Japanese-American, Latino and white) were chosen randomly from among the cohort members. In males, estimates of the correlation between the questionnaire and 24-HDR for nutrients ranged from 0.17 to 0.64 for absolute nutrient intake. For absolute nutrient intakes, the correlations were greatest for whites, somewhat lower for Japanese-Americans and Latinos, and lowest for African-Americans (Kaaks *et al.* 1997; Stram *et al.* 2000).

The reproducibility of the SMHS FFQ compares well with that of the SWHS FFQ (Shu *et al.* 2004). The validation of the SWHS FFQ is slightly better than in the present study. For example, correlation coefficients between the FFQ and the 24-HDR were 0.60 for protein, 0.59 for fat, 0.66 for carbohydrates and 0.55 for fibre in the SWHS, while in the present study the correlation coefficients were 0.49, 0.38, 0.64 and 0.48 for protein, fat, carbohydrates and fibre, respectively. One possible reason for this could be that in Shanghai, women are mainly responsible for purchasing and preparing foods for the family and, thus, are more likely to estimate their dietary intake accurately than men.

A wide LOA indicates that the potential for large differences between methods and agreement is considered poor, even if the bias and dependency are small. The LOA of the SMHS FFQ as compared with multiple 24-HDR are narrower than those reported in two previous studies (Ambrosini *et al.* 2001; Flood *et al.* 2004). Direct comparison with other studies was not possible because the data were not presented as log-transformed (MacIntyre *et al.* 2001; Bakker *et al.* 2003).

In our study, the Bland–Altman plots did not show that the over- or underestimation of dietary intake of the SMHS FFQ depends on the 'true' nutrient intake levels. Thus, misclassification in absolute nutrient intake amount is less likely to cause systematic biases.

Finding a gold standard for measuring long-term dietary intake is the most challenging obstacle in assessing the validity of a dietary instrument. In our study, we chose monthly 24-HDR over a 1-year period as the reference method for assessing usual dietary intake. For the 214 subjects who participated in the calibration study, 93% completed at least ten 24-HDR. The multiple recalls meant we were able to minimise the effect of daily and seasonal variation in dietary intake on the dietary assessment. In order to minimise the possibility that study participants might change their dietary intake to facilitate dietary recall, none of the dietary recall interviews was scheduled in advance (i.e. all were unannounced). Thus, we believe that the dietary recall information obtained from this study is a fairly accurate measurement of the true usual intake for this study population over a 1-year period.

However, the fact remains that multiple dietary recalls may sensitise study participants regarding their dietary intake, and thus participants may answer the FFQ more accurately, resulting in an overestimation of the true validity of the FFQ. When we calculated the correlation between the first FFQ and the mean of the 24-HDR, we found that the correlations were

slightly lower than those observed for the second FFQ. On the other hand, changes in dietary intake during the year could reduce the correlation and increase the mean differences between both instruments. Nutrient and food intake derived from the second FFQ was, in general, lower than that derived from the first FFQ, which might be due to changes in diet. To address whether a possible change in diet might have resulted in a low estimation of the correlation coefficients, we restricted the analysis to those participants that reported 'no change' to the question 'Compared with 5 years ago, have you changed your dietary habits in the past year?' No major changes in correlation coefficients were found.

Assuming the 24-HDR are close to 'true' intake, we found that the SMHS FFQ overestimated soy, fruit and vegetable intake, and slightly underestimated poultry, red meat, fish and egg intake. Some of the measurement error may reflect a bias of study participants seeking social approval (Hebert *et al.* 1995). The substantial overestimation of fruit and vegetable intake is not likely to be accounted for by the seasonal variation of supply of these foods, since our pilot calibration study of this FFQ found little variance (within 5%) for day of the week or season of the year (Cai *et al.* 2005). However, correlation coefficients for these food groups are quite good. Thus, although the SMHS does not appear to do well in estimating the absolute amount of intake of selected groups, it can classify subjects reasonably well in terms of their relative intake levels. The latter is particularly important for epidemiological studies, since it is categorised dietary intake rather than the absolute amount of intake that has been more commonly used in epidemiological studies of diet and chronic disease.

As the recall time interval and format of the questionnaire used by the two methods are different, the errors in these two instruments are less likely to be correlated. The FFQ asked about frequency and amount of consumption for a list of foods (close-ended questions) during the preceding 12-month period, while the 24-HDR employed an open-ended question format and inquired about intake of foods that the study participants ate during the preceding 24 h period. However, we acknowledge that both methods suffer from the social approval bias (Hebert *et al.* 1995) as they are both self-reported.

As an alternative to self-reported methods of dietary intake, biomarkers have been used for validation purposes. Advantages of biomarkers as opposed to self-reported methods are that they are objective, unbiased and their errors are uncorrelated to the FFQ. However, biomarkers are not always available, are expensive and may not necessarily reflect long-term dietary intake habits (Willett, 1990). In the UK EPIC validation studies, the accuracy of several methods was assessed by comparison with another self-report instrument (weighed records) and the biomarkers, 24 h urine nitrogen and potassium, plasma carotenoids and plasma vitamin C (Bingham *et al.* 1997). The correlations between nitrogen from weighed records and estimated food diaries and urinary nitrogen were better than those from other methods, while the results for urinary potassium and serum carotenoids were similar among all methods. We collected four spot urine samples and quarterly blood samples during the 1-year study period from most participants of the study. These samples will be used for further validation of the FFQ when funding becomes available.

Random and uncorrelated measurement errors cause attenuation of relative risk estimates and decrease the statistical power of epidemiological studies. Willett estimated that if a true relative risk is 2.0, the observed relative risk will be attenuated to 1.62 if the correlation between the estimated and true dietary exposure is 0.7 or to 1.32 if the correlation is 0.40 (Willett, 1998). Methods for correction of attenuation and regression diluted bias have become available (Johansson *et al.* 2002). However, calibration studies require certain assumptions about the independence of measurement errors, and biomarkers of dietary intake may better fulfil this requirement (Kaaks *et al.* 2002). We will apply results obtained from this analysis in conjunction with future biomarker studies to calibrate the risk assessment for diet and disease associations in our future research.

In summary, this study suggests that the SMHS FFQ can reasonably categorise usual intake of major nutrients and food groups among men in Shanghai. The SMHS FFQ, however, may under- or overestimate the absolute amount of intake of some nutrients or foods.

Acknowledgements

This study was supported by research grant RO1 CA 82 729 from the National Cancer Institute. R. V., G. Y., D. K. L., Y. B. X., W. Z. and X. O. S. drafted and provided critical revision of the manuscript; R. V., G. Y., H. C. and X. O. S. analysed and interpreted the data; G. Y., D. K. L., B. Y. X. and X. O. S. were responsible for implementation of the study and acquisition of the data; and W. Z. and X. O. S. conceived of and designed the study. The authors have no conflicts of interest to declare.

References

- Ambrosini GL, de Klerk NH, Musk AW & Mackerras D (2001) Agreement between a brief food frequency questionnaire and diet records using two statistical methods. *Public Health Nutr* **4**, 255–264.
- Bakker I, Twisk JW, van Mechelen W, Mensink GB & Kemper HC (2003) Computerization of a dietary history interview in a running cohort; evaluation within the Amsterdam Growth and Health Longitudinal Study. *Eur J Clin Nutr* **57**, 394–404.
- Bingham SA, Gill C, Welch A, *et al.* (1997) Validation of dietary assessment methods in the UK arm of EPIC using weighed records, and 24-hour urinary nitrogen and potassium and serum vitamin C and carotenoids as biomarkers. *Int J Epidemiol* **26**, 137–151.
- Bland JM & Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* **1**, 307–310.
- Bohlscheid-Thomas S, Hoting I, Boeing H & Wahrendorf J (1997) Reproducibility and relative validity of energy and macronutrient intake of a food frequency questionnaire developed for the German part of the EPIC project. European Prospective Investigation into Cancer and Nutrition. *Int J Epidemiol* **26**, Suppl 1, S71–S81.
- Cai H, Yang G, Xiang YB, Hebert JR, Liu DK, Zheng W & Shu XO (2005) Sources of variation in nutrient intakes among men in Shanghai, China. *Public Health Nutr* **8**, 1293–1299.
- Day N, McKeown N, Wong M, Welch A & Bingham S (2001) Epidemiological assessment of diet: a comparison of a 7-day diary with a food frequency questionnaire using urinary markers of nitrogen, potassium and sodium. *Int J Epidemiol* **30**, 309–317.
- Flood VM, Smith WT, Webb KL & Mitchell P (2004) Issues in assessing the validity of nutrient data obtained from a food-frequency questionnaire: folate and vitamin B12 examples. *Public Health Nutr* **7**, 751–756.
- Hebert JR, Clemow L, Pbert L, Ockene IS & Ockene JK (1995) Social desirability bias in dietary self-report may compromise the validity of dietary intake measures. *Int J Epidemiol* **24**, 389–398.
- Johansson I, Hallmans G, Wikman A, Biessy C, Riboli E & Kaaks R (2002) Validation and calibration of food-frequency questionnaire measurements in the Northern Sweden Health and Disease cohort. *Public Health Nutr* **5**, 487–496.
- Kaaks R, Ferrari P, Ciampi A, Plummer M & Riboli E (2002) Uses and limitations of statistical accounting for random error correlations, in the validation of dietary questionnaire assessments. *Public Health Nutr* **5**, 969–976.
- Kaaks R, Slimani N & Riboli E (1997) Pilot phase studies on the accuracy of dietary intake measurements in the EPIC project: overall evaluation of results. European Prospective Investigation into Cancer and Nutrition. *Int J Epidemiol* **26**, Suppl 1, S26–S36.
- MacIntyre UE, Venter CS & Vorster HH (2001) A culture-sensitive quantitative food frequency questionnaire used in an African population: 1. Development and reproducibility. *Public Health Nutr* **4**, 53–62.
- Margetts BM & Pietinen P (1997) European Prospective Investigation into Cancer and Nutrition: validity studies on dietary assessment methods. *Int J Epidemiol* **26**, Suppl 1, S1–S5.
- Martinez ME, Marshall JR, Graver E, Whitacre RC, Woolf K, Ritenbaugh C & Alberts DS (1999) Reliability and validity of a self-administered food frequency questionnaire in a chemoprevention trial of adenoma recurrence. *Cancer Epidemiol Biomarkers Prev* **8**, 941–946.
- Mayer-Davis EJ, Vitolins MZ, Carmichael SL, Hemphill S, Tsaroucha G, Rushing J & Levin S (1999) Validity and reproducibility of a food frequency interview in a Multi-Cultural Epidemiology Study. *Ann Epidemiol* **9**, 314–324.
- Pisani P, Faggiano F, Krogh V, Palli D, Vineis P & Berrino F (1997) Relative validity and reproducibility of a food frequency dietary questionnaire for use in the Italian EPIC centres. *Int J Epidemiol* **26**, Suppl 1, S152–S160.
- Rimm EB, Giovannucci EL, Stampfer MJ, Colditz GA, Litin LB & Willett WC (1992) Reproducibility and validity of an expanded self-administered semiquantitative food frequency questionnaire among male health professionals. *Am J Epidemiol* **135**, 1114–1126.
- Sharma S, Cade J, Jackson M, Mbanya JC, Chungong S, Forrester T, Bennett F, Wilks R, Balkau B & Cruickshank JK (1996) Development of food frequency questionnaires in three population samples of African origin from Cameroon, Jamaica and Caribbean migrants to the UK. *Eur J Clin Nutr* **50**, 479–486.
- Shu XO, Yang G, Jin F, Liu D, Kushi L, Wen W, Gao YT & Zheng W (2004) Validity and reproducibility of the food frequency questionnaire used in the Shanghai Women's Health Study. *Eur J Clin Nutr* **58**, 17–23.
- Stram DO, Hankin JH, Wilkens LR, *et al.* (2000) Calibration of the dietary questionnaire for a multiethnic cohort in Hawaii and Los Angeles. *Am J Epidemiol* **151**, 358–370.
- Thompson RL & Margetts BM (1993) Comparison of a food frequency questionnaire with a 10-day weighed record in cigarette smokers. *Int J Epidemiol* **22**, 824–833.
- Willett W (1990) *Nutritional Epidemiology*, Vol. 15, New York: Oxford University Press, Inc.
- Willett W (1998) *Nutritional Epidemiology*, 2nd ed. New York: Oxford University Press.
- Willett WC, Sampson L, Stampfer MJ, Rosner B, Bain C, Witschi J, Hennekens CH & Speizer FE (1985) Reproducibility and validity of a semiquantitative food frequency questionnaire. *Am J Epidemiol* **122**, 51–65.
- Yang G, Wang G & Pan X (2002) *China Food Composition*. Beijing: Peking University Medical Press.