# Validity of a food-frequency questionnaire for a large prospective cohort study in Bangladesh

Yu Chen[1], Habibul Ahsan[1,2]*, Faruque Parvez[1,3] and Geoffrey R. Howe[1]

[1]*Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, USA*
[2]*Herbert Irving Comprehensive Cancer Center, Columbia University, New York, USA*
[3]*Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University, New York, USA*

We have developed a thirty-nine-item semi-quantitative food-frequency questionnaire (FFQ) to assess the dietary consumption of 11 746 men and women in a prospective cohort study that evaluates the health effects of As from drinking water in Bangladesh. In order to validate the FFQ, two 7 d food diaries (FD) were completed for 189 randomly selected cohort participants in two different seasons of the year. Nutrient values were converted based on both the United States Department of Agriculture's National Nutrient Database and a food composition table for the Indian subcontinent. Pearson product-moment and Spearman non-parametric rank correlation coefficients comparing food and nutrient consumptions estimated from FFQ and 7 d FD were calculated based on log-transformed consumption values with or without adjustment for total energy and correction for within-individual variation. Correlations of macronutrients and common micronutrients including total fat, monounsaturated fat, polyunsaturated fat, saturated fat, protein, carbohydrate, dietary fibre, Na, K, vitamin $B_6$, vitamin $B_{12}$, riboflavin, Mn, thiamin and Fe were moderately good, ranging from 0·30 to 0·76. However, correlations of other micronutrients were weak ($< 0·30$). Large seasonal variations in intakes of retinol equivalents and vitamin C were observed. This analysis documents the degree of validity of the FFQ in measuring specific nutrient intakes in the study population. To our knowledge, the present study is the first to document the validity of a FFQ with the use of 7 d FD in a Bangladeshi population.

**Arsenic: Validation study: Food-frequency questionnaires: Food diaries: Nutritional epidemiology**

Nutritional problems are at the root of many major chronic diseases in both developed and developing countries. The complex nature of diet has posed an unusually difficult challenge to its measurements in nutritional epidemiological studies (Willett, 1987). Inadequate dietary assessment can produce false negative or inconsistent associations between dietary exposures and diseases. Food-frequency questionnaires (FFQ) based on self-reported intakes have been shown to validly measure long-term dietary patterns in epidemiological studies conducted in Westernised countries (Salvini *et al.* 1989; Feskanich *et al.* 1993; Erkkola *et al.* 2001). While estimates of absolute values of dietary intakes from FFQ may not be accurate, they can be used to categorise subjects into broad ranges of intakes in order to compare subjects' diets relative to one another. However, attempts of formally validating FFQ have rarely been made for studies in developing countries. The validity of the FFQ depends on its specific structure and the population to which it is applied (Willett & Lenart, 1998). For instance, while self-administered and interviewer-administered FFQ have been demonstrated to have comparable validity in Westernised societies (Rimm *et al.* 1992), the same may not be applied to populations where the formal education level is low. It is therefore particularly important to validate FFQ administered in the population of developing countries.

It has been estimated that up to 57 million individuals in Bangladesh have been chronically exposed to high levels of As from drinking water (British Geological Survey, 1999). Nutritional factors may modify the detrimental health effects of chronic exposure to inorganic As from drinking water (Hsueh *et al.* 1995, 1998, 2003). However, systematic epidemiological research with respect to the potential modifying effects of nutrition on the risk of As-related diseases is lacking. We developed a thirty-nine-item semi-quantitative FFQ to assess the diet of 11 746 participants in the Health Effects of Arsenic Longitudinal Study (HEALS) in Araihazar, Bangladesh. To evaluate the validity of the thirty-nine-item FFQ, we conducted a validation study among 189 randomly selected HEALS participants using two 7 d food diaries (FD). We chose two 7 d FD administered in two different seasons within 1 year as the reference to assess

the validity of the FFQ. Although there is no perfect measure of dietary intakes, multiple weekly food records or FD are judged to be superior to FFQ and have been frequently used to assess the validity of FFQ (Willett *et al.* 1985; Feskanich *et al.* 1993; Erkkola *et al.* 2001). It is crucial that errors of the two methods being compared in a validation study are as independent as possible. Unlike multiple 24 h dietary recalls and FFQ, FD are not dependent on memory, conceptions of serving portions and the interpretation of questions (Willett & Lenart, 1998).

## Material and methods

### Parent study

The parent HEALS is a large ongoing prospective cohort study to investigate the health effects of As exposure from drinking water in Bangladesh. Briefly, water samples and household and geographic data were collected and analysed for 5966 contiguous wells serving nearly 66 000 individuals in a well-defined 25 km$^2$ geographic area in Araihazar, Bangladesh to create a sampling frame for subject recruitment in the cohort study (van Geen *et al.* 2002). From 22 October 2000 to 19 May 2002, 11 746 men and women were recruited at baseline from the users of 5966 contiguous wells. The extensive baseline interview included inquiries on water-drinking patterns and history, demographics and lifestyle characteristics. Information on diet was assessed using a newly developed FFQ. In addition, trained physicians completed comprehensive physical examinations with special emphasis on As-related signs and symptoms such as As-induced skin lesions and skin cancers. Biological specimens, including whole blood and spot urine samples, were collected from 97 % of study participants.

### Study design for the validation study

Stratified sampling was conducted for the FFQ validation study. Parallel with the baseline interview of the cohort study, 300 wells were randomly selected from the 4236 wells that enumerated the overall cohort study participants. From cohort study participants who were users of these 300 wells, 200 subjects were randomly chosen and invited to the FFQ validation study.

### The food-frequency questionnaire

The semi-quantitative thirty-nine-item FFQ was designed to assess the long-term daily diet of subjects in the cohort study. Since the average education level of residents in the present study area was low (44 % without any formal education), trained interviewers completed the FFQ through in-person interviews. HEALS investigators, with help from local nutrition experts, first identified all the food items available at the village market in the study area. Then, a preliminary version of the FFQ was developed following extensive discussions with ten focus groups, each consisting of six to ten participants. The FFQ was finalised after pilot testing among 120 local individuals who were not part of the cohort study. Only common food

items were included in the FFQ (Appendix 1), and food items with intake frequencies less than once per month during the past 1 year were deemed to be insignificant. About 10 % of the food items were removed from the original food list because of infrequent consumption. Although a close-ended format may theoretically help subjects describe their dietary patterns, its fixed categorisations are not appropriate for foods that have large differences in seasonal availability and its numerous categories may be cumbersome in an interview setting. Therefore, to simplify the FFQ, open-ended questions for the amount per meal, frequency per d, month and year were used. For instance, for mango, subjects were asked about the number of months per year they consumed mango, the number of days per week they ate it during those months, the number of times in a typical day they ate it and, last, the amount they consumed on average each time. Different locally used plates and utensils were shown to the subjects to define the portion sizes during the interviews.

### Seven-day food diaries

The two 7 d FD were completed in two major separate seasons in order to capture the seasonal variation of food availability. The first FD took place between April 2001 and July 2001 and the second between January 2002 and March 2002. Because of the large amount of work involved, twenty village interviewers were recruited. They received a 3 d training that included discussion sessions on the methods of the FD, detailed explanations of the study purpose, demonstrations of interviews and interview practice through role-playing. During data collection, each of the twenty trained interviewers visited six to ten subjects at home three times per d either during or immediately after the meal (no later than 2 h) to directly measure and record consumptions of the thirty-nine food items listed in the FFQ. Two study coordinators and two nutrition experts supervised the interviewers, checked the completeness of the FD and held discussion sessions on a daily basis for quality control. Among the 200 subjects invited, 198 completed the first 7 d FD, of which 193 completed the second. Among these 193 subjects with both 7 d FD completed, four were excluded because of incomplete FFQ. Therefore, the final sample for the validation study included 189 subjects. Baseline interviews and FFQ were conducted for these subjects between October 2000 and June 2001. FD were conducted after the FFQ for all study subjects. On average, the duration from the FFQ to the first 7 d FD was 76 d, and the duration from the first to the second FD was 190 d.

### Statistical analysis

Average daily intakes of food items and nutrients were calculated separately based on the FFQ and FD. Since nutrient values specifically for foods in Bangladesh were not available, we used both the United States Department of Agriculture (USDA) Nutrient Database for Standard Reference (abbreviated version) (United States Department of Agriculture, 2002) and an Indian food nutrient database (Gopalan *et al.* 1989). Given the geographical

and socio-cultural similarities between India and Bangladesh, we were able to identify exact food items listed in the FFQ from the Indian food composition table. While food items listed in the FFQ were closely matched to those in the Indian food composition tables, the USDA data may have more accurate assessments of nutrients. Several studies have documented As contents of foodstuffs collected in As-affected areas (Roychowdhury *et al.* 2002; Misbahuddin, 2003). The content of As in foodstuffs depends on the As concentration in the soil, in the water used for washing and cooking, in the water used for irrigation purpose and in the pesticides (Roychowdhury *et al.* 2002). Therefore, the evaluation of the As content in the diet is complex and will be the subject of a separate future analysis.

All dietary intakes were log-transformed to improve the normality of the distribution with the formula log $(x + 1)$ since the consumption of a given item could be zero. Pearson product-moment and Spearman non-parametric rank correlations between daily intakes estimated by FFQ and those calculated based on FD were computed. Only Pearson correlations were presented because results were similar. Since one of our major goals is to ultimately use the FFQ data to assess the effects of nutrient composition, rather than absolute individual nutrient intakes, correlations based on energy-adjusted nutrient intakes were calculated. Energy adjustment is based on the *a priori* biological considerations that a larger, more physically active individual will require a higher energy intake, which is also associated with higher absolute intakes of all nutrients. We elected to use the residual method to adjust the effects of total energy intakes on individual nutrient intakes (Willett & Stampfer, 1986). The residuals, derived from models regressing intakes of each nutrient on total energy intakes, were standardised to the predicted log-transformed nutrient intake of a subject with the average total energy intake of all study subjects (Willett & Stampfer, 1986). In addition, since random within-individual error in the measurement of any of the variables being compared tends to reduce correlation coefficients toward zero (Liu *et al.* 1978; Beaton *et al.* 1979), correlations with corrections for the attenuating effects of such measurement error in the two 7 d FD were computed. Correction of correlations for random within-individual error was based on the formula:

$$\gamma_t = \gamma_o \ (1 + \lambda/n)^{1/2},$$

where $\gamma_t$ is the true correlation, $\gamma_o$ is the observed correlation of energy-adjusted intakes derived from the FFQ and the FD, $\lambda$ is the ratio of the within- to between-individual variances of weekly intakes and $n$ is the number of replicates per individual for weekly intakes. For the present study, $n$ is 2, denoting the two 7 d FD. The degree of misclassification in FFQ was evaluated by the contingency table of quintiles defined based on the FFQ and the FD. When the relationship between two normally distributed variables is linear, correlation coefficients provide the same information as that contained in such a contingency table (Walker & Blettner, 1985). However, even after log-transformation, the distributions of dietary variables are sometimes still skewed. A contingency table is informative in assessing the validity of FFQ when values of intakes are not entirely linearly related with one another or are not normally distributed. Last, to examine the seasonal variation in nutrient intakes, we computed intra-class correlation coefficients (Shrout & Fleiss, 1979) between the average daily intakes derived from the two 7 d FD. These methods have been used by FFQ validation studies of large epidemiological studies in Westernised countries (Salvini *et al.* 1989; Feskanich *et al.* 1993; Erkkola *et al.* 2001). All statistical analyses were conducted using the Statistical Analysis System (SAS) program (version 8.02; Cary, NC, USA).

## Results

Table 1 shows the distributions of socio-demographic characteristics among the FFQ validation study participants and the rest of the cohort study participants. The mean age of the participants in the FFQ validation study is slightly greater than that of the rest of the cohort members. However, distributions of age and other attributes including sex, BMI, education, occupation and religion are similar between the two groups.

### Validity

Table 2 presents the average daily intakes and correlations between values derived from the FFQ and the FD for twelve food groups. Food items that contribute to each food group are listed in Appendix 1. Compared with the FD, food intakes are more likely to be overestimated by the FFQ, especially for more desirable food groups such as tea and fruits. Large underestimation is observed for food groups such as bread, poultry and eggs that are less frequently consumed among study subjects in Bangladesh. Correlations based on energy-adjusted values are similar to those based on unadjusted intakes for most food groups, except rice for which energy adjustment greatly improved the correlation. Correlations based on unadjusted and energy-adjusted values for liquids, rice, bread and meat are better than those for vegetables, eggs, fruits and fish, which could partly be due to large seasonal variation in the availability of the latter items. Corrections for within-individual variation in the FD improved correlation estimates, especially for estimates associated with eggs and vegetables, indicating greater week-to-week within-individual variation as compared with between-individual variation.

Average daily intakes of nutrients and correlations of nutrient intakes derived from the FFQ and FD are presented in Table 3. Estimates of average daily intakes obtained from the FFQ and FD are fairly similar to one another, although there are differences in values derived using the USDA and Indian food composition tables. Consistent with the finding of overestimation of fruit consumption, the average daily intake of vitamin C measured by the FFQ is more than 1·25 times greater than that of FD. Compared with nutrient intakes estimated by FD, variation in intakes derived from the FFQ is greater, as indicated by the larger standard deviations of nutrient intakes. Energy adjustments in general improved the correlations. Large

**Table 1.** Comparison between participants in the validation study and the overall cohort study participants

| | Validation study partici-pants (n 189) | | Rest of cohort participants (n 11 561) | | |
|---|---|---|---|---|---|
| | n | % | n | % | P value* |
| **Sex** | | | | | |
| Male | 85 | 45.0 | 4958 | 42.9 | 0.57 |
| Female | 104 | 55.0 | 6603 | 57.1 | |
| **Age (years)** | | | | | |
| 17–29 | 36 | 19.1 | 2921 | 25.3 | 0.12 |
| 30–39 | 67 | 35.5 | 4079 | 35.3 | |
| 40–49 | 52 | 27.5 | 2954 | 25.6 | |
| 50–59 | 26 | 13.8 | 1352 | 11.7 | |
| 60–75 | 8 | 4.2 | 253 | 2.2 | |
| Missing | 0 | | 2 | | |
| Mean | 38.9 | | 37.0 | | 0.01 |
| SD | 10.7 | | 10.1 | | |
| **BMI (kg/m²)** | | | | | |
| < 18.1 | 50 | 26.5 | 3736 | 33.1 | 0.14 |
| 18.1–20.5 | 71 | 37.6 | 3712 | 32.9 | |
| > 20.5 | 68 | 35.9 | 3831 | 34.0 | |
| Missing | 0 | | 282 | | |
| Mean | 19.8 | | 19.8 | | 0.76 |
| SD | 3.6 | | 3.2 | | |
| **Education (years)** | | | | | |
| 0 | 67 | 35.5 | 5173 | 44.8 | 0.08 |
| 1–5 | 66 | 34.9 | 3405 | 29.4 | |
| 6–10 | 34 | 18.0 | 1716 | 14.9 | |
| > 10 | 22 | 11.6 | 1261 | 10.9 | |
| Missing | 0 | | 6 | | |
| Mean | 3.9 | | 3.5 | | 0.08 |
| SD | 3.8 | | 3.8 | | |
| **Religion** | | | | | |
| Muslim | 182 | 96.3 | 11 003 | 95.2 | 0.90 |
| Hindu | 7 | 3.7 | 552 | 4.8 | |
| Christian | 0 | 0.0 | 2 | 0.0 | |
| Other | 0 | 0.0 | 4 | 0.0 | |
| **Occupation** | | | | | |
| Daily labourer | 12 | 6.4 | 422 | 3.7 | 0.48 |
| Farmer | 14 | 7.4 | 696 | 6.0 | |
| Factory | 17 | 9.0 | 1168 | 10.1 | |
| Other paid job | 7 | 3.7 | 434 | 3.8 | |
| Business | 31 | 16.4 | 1824 | 15.8 | |
| Unemployment | 6 | 3.2 | 239 | 2.1 | |
| Homemaker | 95 | 50.3 | 6216 | 53.8 | |
| Others | 7 | 3.7 | 561 | 4.8 | |
| Missing | 0 | | 1 | | |

\* P value for *t* test (comparing means) or $\chi^2$ test (comparing proportions).

random within-individual error is observed for most of the nutrients in the two 7 d FD and therefore corrections for such error improved the correlations. Among the correlations based on the Indian food composition table, correlations of macronutrients including total fat, protein, carbohydrate and dietary fibre are moderately good, ranging from 0.40 to 0.56, according to estimates based on energy-adjusted intakes with corrections for random within-subject error. Correlations of common micronutrients such as Na, K, vitamin $B_6$, riboflavin, Mg, thiamin and Fe, ranging from 0.30 to 0.45, are better than those of other vitamins and trace elements. Interestingly, correlations derived based on the USDA food composition table demonstrate a similar pattern, except that the correlations for vitamin $B_6$ and thiamin are weak. Among the nutrients that are not included in the Indian

food composition table, correlations for monounsaturated fat, polyunsaturated fat, saturated fat and vitamin $B_{12}$ are strong, ranging from 0.49 to 0.76.

*Classification*

Table 4 presents the joint classification of average daily intakes in quintiles defined by the FFQ and FD. The proportions of correctly categorised subjects by the FFQ in the same extreme and adjacent quintiles are calculated. Subjects that are assigned by the FFQ to the opposed extreme quintiles of their intakes based on FD can be considered grossly misclassified. Compared with nutrients with poor correlations, nutrients with better correlations, such as lipid fats, protein, Na, carbohydrate and vitamin $B_{12}$, in general are more likely to have better joint classification, with more than 50 % of subjects in the lowest and highest FD quintile being categorised to the lowest and highest two FFQ quintiles, respectively. Riboflavin, Fe, retinol equivalents, vitamin C, niacin and folate are observed to have greater misclassification, with 15 % or more of subjects in the lowest and highest FD quintile being assigned to the highest and lowest FFQ quintile, respectively. Importantly, the FFQ and FD are more likely to agree with one another in higher ranges of intakes, indicating that those with high intakes may be more likely to be identified as so by the FFQ. The proportions of subjects that are correctly classified in the highest quintiles are greater than the proportions of subjects that are correctly classified in the lowest two quintiles by FFQ for nineteen of the twenty-six nutrients evaluated.

*Seasonal variation*

Intra-class correlations between intakes of the two 7 d diaries are presented in Table 5. Reproducibility of the FD can serve as an indicator to the degree of week-to-week and seasonal variations in nutrient intakes. Intakes of vitamin B, total energy, vitamin E, carbohydrate and polyunsaturated fat are observed to have the least week-to-week variation. Interestingly, nutrients such as retinol equivalents and vitamin C that had the lowest correlations of intakes based on the two FD are those with poor correlations of the FFQ and FD.

**Discussion**

We evaluated the performance of a semi-quantitative thirty-nine-item FFQ administered in a rural population of Bangladesh by comparing average daily dietary intakes derived from this instrument with those recorded in two 7 d FD. Correlations of macronutrients and common micronutrients including lipid fats, protein, carbohydrate, dietary fibre, Na, K, vitamin $B_6$, vitamin $B_{12}$, Fe, Mn, thiamin and riboflavin were moderately good ($\geq 0.30$). However, the validity of the FFQ in assessing intakes of other micronutrients such as retinol equivalents, vitamin E, vitamin C, Zn, folate and niacin is questionable, as indicated by the weak correlations between the FFQ and FD. Little is known about the validity of self-reported dietary intakes in populations of developing countries. To our knowledge,

**Table 2.** Average daily intake of food groups estimated by food diaries (FD) and the food-frequency questionnaire (FFQ)

(Mean values, standard deviations and correlation coefficients)

| Food (g)† | FD | | FFQ | | FFQ (% of FD) | Pearson correlation coefficients | | |
|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | | Unadjusted‡ | Energy-adjusted§ | Corrected‖ |
| Milk (ml) | 36·4 | 49·8 | 33·6 | 50·8 | 92 | 0·43 | 0·39 | 0·78 |
| Rice | 1574·3 | 472·9 | 1532·2 | 507·7 | 97 | 0·15 | 0·39 | 0·59 |
| Tea (ml) | 17·7 | 50·8 | 55·0 | 145·4 | 311 | 0·37 | 0·37 | 0·59 |
| Bread | 29·6 | 32·2 | 12·6 | 35·5 | 43 | 0·43 | 0·37 | 0·56 |
| Beef and lamb | 24·6 | 25·8 | 17·4 | 36·9 | 71 | 0·23 | 0·27 | 0·38 |
| Potatoes | 105·5 | 67·2 | 120·9 | 79·8 | 115 | 0·23 | 0·22 | 0·35 |
| Vegetables | 257·1 | 117·2 | 255·6 | 151·1 | 99 | 0·17 | 0·12** | 0·34** |
| Eggs | 14·1 | 13·0 | 9·0 | 19·4 | 64 | 0·11* | 0·10* | 0·32* |
| Beans | 125·1 | 75·5 | 126·0 | 95·4 | 101 | 0·22 | 0·24 | 0·30 |
| Poultry | 11·3 | 17·5 | 4·6 | 16·3 | 41 | 0·23 | 0·19 | 0·25 |
| Fish | 44·9 | 24·2 | 62·5 | 44·3 | 139 | 0·08* | 0·09* | 0·19* |
| Fruits | 39·0 | 41·4 | 133·8 | 122·6 | 343 | 0·16** | 0·10* | –¶ |

\* $P > 0.1$, \*\* $0.05 \leq P < 0.1$; for all other correlations $P < 0.05$.

† Food items were listed in the descending order of correlations based on energy-adjusted intake values with correction for random within-individual error in the two 7 d FD.

‡ Correlations were calculated based on log-transformed values of intakes.

§ Correlations were calculated based on log-transformed values of intakes with adjustment for total energy intake using the residual method.

‖ Correlation coefficients were calculated based on energy-adjusted intake values with correction for random within-individual error in the two 7 d FD.

¶ Correlation coefficient was not estimated due to very large ratio of within-individual to between-individual variances.

the present study is the first to document the validity of a FFQ with the use of 7 d food records in a Bangladeshi population.

Trained interviewers conducted in-person, face-to-face dietary assessments for both the FFQ and the FD, and therefore potential errors of self-administration were avoided. Although dietary interviews were time-consuming and required more manpower, unlike self-administered dietary measures, the FFQ and FD developed in the present study do not require literacy. Other potential advantages of interviewer-administered FFQ included increased subjects' interest, reduced misunderstanding and better quantification of consumption (Jain et al. 1996). The moderate validity for measuring intakes of common foods and nutrients demonstrated in the present study provides encouragement to the conducting of nutritional epidemiological studies using FFQ in low-income developing countries.

Similar to studies conducted in Westernised countries (Salvini et al. 1989; Feskanich et al. 1993), we observed under-reporting of meats and over-reporting of fruits in the FFQ in the present study. The levels of overestimation and underestimation of food consumptions by the FFQ as compared with those recorded in the FD are somehow greater than what have been reported in FFQ validation studies conducted in Westernised countries (Salvini et al. 1989; Feskanich et al. 1993; Erkkola et al. 2001). It cannot be ruled out that subjects enrolled in the validation study might have changed their dietary behaviours because of the fact that their diet was being observed. To the extent that their diet was altered, correlations between the FFQ and FD would be underestimated. In addition, the joint classification of the FFQ and FD indicates heavy consumers were more likely to be correctly classified as such by the FFQ. Greater fluctuations in seasonal availabilities of foods in Bangladesh might have made it more difficult for subjects to recall their dietary intakes in the FFQ, and heavy consumers may have recalled their diet behaviours better. Indeed, intra-class correlations between the two

7 d FD that empirically assessed diets indicate that the week-to-week variation for micronutrients such as retinol equivalents, vitamin C and cholesterol was large. The literature on diets in Bangladesh evidently indicates a striking seasonal variation in dietary consumption, especially for vitamin A (Abdullah & Wheeler, 1985; Brown et al. 1985; Abdullah & Ahmed, 1993). The reliance on green leafy vegetables as the predominant source of vitamin A can result in substantial fluctuations in intake, depending on the seasonal availability of specific vegetables. Although the two 7 d FD were selected in two different seasons, the large seasonal variation in foods such as fruits and vegetables may be well beyond what was ascertained in the FD. Hence, we were limited in assessing the validity of the FFQ in measuring long-term intakes of micronutrients with wide seasonal variation in their availability. As opposed to conventional recommendations for FFQ validation studies (Willett et al. 1985; Rimm et al. 1992), two 7 d FD in the setting of a rural area in Bangladesh may not be sufficient to provide satisfying estimates for long-term consumption for micronutrients. Due to logistic reasons, it was not possible to administer more 7 d FD to capture more week-to-week variations. This is a limitation of the present study. As blood and urine samples were collected for most of the subjects in the cohort study, intercorrelations of nutrient intakes from the FFQ, FD and their biochemical measures would be of interest in the future.

Unlike conventional open-ended food records, subjects in the two 7 d FD were assessed for intakes of the same thirty-nine food items listed in the FFQ. If sources of essential nutrients were not covered by the food list, the use of the same food list in both the FFQ and FD may lead to some extent of overestimation of the validity of the FFQ in assessing the long-term diet. We are confident that the food list covers more than 90 % of the typical daily diet in the study population, provided that the development of the food list was based on extensive pilot work and that

**Table 3.** Average daily intake of nutrients estimated by food diaries (FD) and the food-frequency questionnaire (FFQ)

(Mean values, standard deviations and correlation coefficients)

| | USDA nutrient database | | | | | | | | Indian nutrient database | | | | | | | |
| | FD | | FFQ | | FFQ/FD (%) | Pearson correlation coefficients | | | FD | | FFQ | | FFQ/FD (%) | Pearson correlation coefficients | | |
| Nutrients† | Mean | SD | Mean | SD | | Unadjusted‡ | Energy-adjusted§ | Corrected‖ | Mean | SD | Mean | SD | | Unadjusted‡ | Energy-adjusted§ | Corrected‖ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Total fat (g) | 28.0 | 8.2 | 25.2 | 12.7 | 90 | 0.24 | 0.35 | 0.70 | 10.7 | 3.6 | 9.8 | 5.1 | 92 | 0.27 | 0.29 | 0.56 |
| Protein (g) | 76.6 | 18.9 | 75.1 | 24.4 | 98 | 0.16 | 0.30 | 0.53 | 93.2 | 24.3 | 88.6 | 31.8 | 95 | 0.17 | 0.29 | 0.52 |
| Na (mg) | 537.4 | 242.6 | 456.1 | 263.9 | 85 | 0.25 | 0.25 | 0.42 | 158.8 | 63.8 | 182.8 | 100.4 | 115 | 0.26 | 0.25 | 0.45 |
| Carbohydrate (g) | 431.8 | 116.3 | 439.6 | 125.9 | 102 | 0.13 | 0.26 | 0.35 | 534.5 | 134.8 | 544.3 | 152.3 | 102 | 0.11* | 0.32 | 0.43 |
| Dietary fibre (g) | 42.5 | 11.0 | 43.8 | 12.6 | 103 | 0.13 | 0.17 | 0.24 | 5.1 | 1.6 | 6.2 | 2.9 | 122 | 0.19 | 0.18 | 0.40 |
| Vitamin B$_6$ (mg) | 3.4 | 0.9 | 3.5 | 1.0 | 103 | 0.15 | 0.04* | 0.05* | 0.6 | 0.2 | 0.6 | 0.2 | 100 | 0.13 | 0.29 | 0.39 |
| Riboflavin (mg) | 1.1 | 0.3 | 1.0 | 0.4 | 91 | 0.25 | 0.25 | 0.46 | 1.1 | 0.3 | 1.1 | 0.4 | 100 | 0.21 | 0.22 | 0.37 |
| Se (mg) | 192.7 | 48.4 | 184.3 | 53.3 | 96 | 0.13** | 0.23 | 0.34 | 225.3 | 83.2 | 278.4 | 134.6 | 124 | 0.15 | 0.20 | 0.37 |
| K (mg) | 2711.9 | 718.7 | 2976.8 | 1048.8 | 110 | 0.15 | 0.14** | 0.20** | 1379.7 | 497.4 | 1482.5 | 794.5 | 107 | 0.16 | 0.21 | 0.36 |
| Fe (mg) | 14.1 | 3.4 | 13.1 | 4.0 | 93 | 0.14** | 0.15* | 0.28** | 25.3 | 6.8 | 26.0 | 9.9 | 103 | 0.13 | 0.18 | 0.34 |
| Mn (mg) | 16.4 | 4.4 | 15.9 | 4.6 | 97 | 0.13** | 0.26 | 0.36 | 6.5 | 1.7 | 6.3 | 1.9 | 97 | 0.10 | 0.22 | 0.30 |
| Thiamin (mg) | 2.1 | 0.5 | 2.1 | 0.6 | 100 | 0.12* | 0.04* | 0.05* | 2.2 | 0.6 | 2.2 | 0.7 | 100 | 0.10* | 0.20 | 0.30 |
| Cu (mg) | 2.4 | 0.6 | 2.4 | 0.7 | 100 | 0.11* | 0.03* | 0.04* | 3.4 | 1.3 | 3.5 | 1.8 | 103 | 0.14** | 0.14 | 0.29 |
| Zn (mg) | 14.5 | 3.5 | 13.7 | 4.3 | 94 | 0.10* | 0.19 | 0.30 | 11.0 | 3.0 | 10.9 | 3.5 | 99 | 0.09* | 0.13** | 0.19** |
| Mg (mg) | 854.9 | 221.1 | 839.9 | 235.4 | 98 | 0.15 | 0.24 | 0.33 | 834.7 | 217.1 | 958.8 | 334.5 | 115 | 0.04* | 0.13** | 0.18* |
| Ca (mg) | 407.3 | 111.9 | 393.0 | 138.0 | 96 | 0.19 | 0.14 | 0.23 | 635.1 | 217.1 | 704.9 | 323.1 | 111 | 0.12* | 0.10* | 0.17* |
| Folate (µg) | 315.4 | 103.0 | 310.2 | 126.0 | 98 | 0.09* | 0.08* | 0.30* | 166.7 | 57.6 | 137.9 | 64.5 | 83 | 0.11 | 0.07* | 0.14* |
| Energy (kJ) | 9506.0 | 2406.6 | 9492.2 | 2699.5 | 100 | 0.12* | – | 0.17* | 10939.9 | 2674.0 | 10959.6 | 3065.6 | 100 | 0.11* | – | 0.14* |
| Retinol equivalents (mg) | 439.0 | 254.2 | 390.4 | 323.0 | 89 | 0.06* | 0.06* | –¶ | 734.5 | 354.6 | 865.0 | 685.5 | 118 | 0.01* | 0.01* | 0.11* |
| Vitamin C (mg) | 91.0 | 40.0 | 114.8 | 69.2 | 126 | 0.08* | 0.05* | 0.12* | 104.9 | 42.4 | 137.9 | 70.6 | 131 | 0.06* | 0.02* | –¶ |
| Niacin (mg) | 31.0 | 8.1 | 30.8 | 8.7 | 99 | 0.11* | 0.28 | 0.39 | 27.6 | 6.8 | 27.6 | 7.7 | 100 | 0.08* | 0.06* | 0.08* |
| Saturated fat (g)†† | 7.4 | 2.9 | 6.2 | 4.6 | 84 | 0.32 | 0.38 | 0.76 | – | – | – | – | – | – | – | – |
| Monounsaturated fat (g)†† | 9.8 | 3.2 | 8.4 | 5.0 | 86 | 0.23 | 0.33 | 0.66 | – | – | – | – | – | – | – | – |
| Vitamin B$_{12}$ (mg)†† | 2.0 | 0.9 | 2.1 | 1.6 | 105 | 0.24 | 0.27 | 0.57 | – | – | – | – | – | – | – | – |
| Polyunsaturated fat (g)†† | 7.2 | 1.8 | 6.7 | 2.0 | 93 | 0.16 | 0.26 | 0.49 | – | – | – | – | – | – | – | – |
| Cholesterol (mg)†† | 135.6 | 70.2 | 119.2 | 111.7 | 88 | 0.14 | 0.14 | 0.32 | – | – | – | – | – | – | – | – |
| Vitamin E (mg)†† | 5.6 | 1.4 | 5.7 | 2.1 | 102 | 0.13** | 0.09* | 0.16* | – | – | – | – | – | – | – | – |

*$P>0.1$, **$0.05 \leq P<0.1$; for all other correlations $P<0.05$.

† Nutrients were listed in the descending order of correlations calculated based on Indian food composition table with energy adjustment and correction for within-individual error.

‡ Correlations were calculated based on log-transformed values of intakes.

§ Correlations were calculated based on log-transformed values of intakes with adjustment for total energy intake using the residual method.

‖ Correlation coefficients were calculated based on energy-adjusted intake values (except for energy intake which was based on unadjusted intake values) with correction for random within-individual error in the two 7 d FD.

¶ Correlation coefficient was not estimated due to very large ratio of within-individual to between-individual variances.

†† Nutrient values were not available for the Indian food composition table.

**Table 4.** Comparison of energy-adjusted nutrient intakes estimated by food diaries (FD) and the food-frequency questionnaire (FFQ) based on joint classification by quintiles

| Nutrients* | Lowest quintile on FD† | | | Highest quintile on FD† | | |
|---|---|---|---|---|---|---|
| | Lowest quintile on FFQ† (%) | Lowest two quintiles on FFQ† (%) | Highest quintile on FFQ† (%) | Highest quintile on FFQ† (%) | Highest two quintiles on FFQ† (%) | Lowest quintile on FFQ† (%) |
| Polyunsaturated fat (g) | 23 | 51 | 10 | 36 | 64 | 10 |
| Saturated fat (g) | 44 | 57 | 7 | 41 | 64 | 3 |
| Monounsaturated fat (g) | 23 | 50 | 7 | 38 | 63 | 3 |
| Total fat (g) | 26 | 52 | 15 | 36 | 63 | 12 |
| Protein (g) | 38 | 56 | 18 | 36 | 62 | 5 |
| Vitamin $B_6$ (mg) | 31 | 46 | 20 | 33 | 62 | 8 |
| Na (mg) | 38 | 57 | 10 | 31 | 59 | 10 |
| Mn (mg) | 28 | 49 | 10 | 28 | 59 | 13 |
| Carbohydrate (g) | 38 | 56 | 7 | 44 | 57 | 15 |
| Zn (mg) | 10 | 28 | 10 | 33 | 56 | 10 |
| Cu (mg) | 18 | 49 | 10 | 36 | 54 | 8 |
| Vitamin $B_{12}$ (mg) | 36 | 54 | 7 | 26 | 52 | 5 |
| Riboflavin (mg) | 33 | 44 | 15 | 28 | 51 | 18 |
| Fe (mg) | 38 | 46 | 15 | 36 | 51 | 18 |
| Ca (mg) | 33 | 38 | 12 | 23 | 51 | 18 |
| Dietary fibre (g) | 33 | 49 | 7 | 26 | 50 | 15 |
| Thiamin (mg) | 23 | 43 | 8 | 31 | 50 | 13 |
| Cholesterol (mg) | 26 | 59 | 10 | 31 | 49 | 7 |
| Vitamin E (mg) | 21 | 42 | 18 | 28 | 46 | 13 |
| Mg (mg) | 15 | 41 | 18 | 23 | 46 | 5 |
| Niacin (mg) | 23 | 41 | 21 | 23 | 44 | 21 |
| K (mg) | 41 | 62 | 7 | 21 | 44 | 13 |
| Vitamin C (mg) | 21 | 49 | 15 | 26 | 41 | 18 |
| Se (mg) | 38 | 63 | 13 | 21 | 39 | 21 |
| Folate (μg) | 23 | 46 | 15 | 25 | 38 | 21 |
| Retinol equivalents (mg) | 15 | 44 | 21 | 13 | 32 | 18 |

* Nutrients were listed in descending order of percentages of those in highest quintiles of FD that are correctly classified to FFQ highest two quintiles.
† Categorisation was determined based on log-transformed, energy-adjusted intake values.

the diet in rural Bangladesh is relatively homogeneous and simple. Food diversity is limited in scope by the low availability of tillable land. The average diet is almost entirely cereal-based, comprising very few highly processed food products. The reasonable average daily intakes of energy computed from both FFQ and FD gave reassurance on this aspect. However, it is possible that including more infrequent food items or defining each food item more specifically would have improved correlations for micronutrients. For example, different species of small freshwater fish may contain different amounts of micronutrients. While the majority of food items listed on the FFQ are distinct from one another, categories including beef and lamb, poultry, big fish and small fish are associated with multiple different food items. A more detailed FFQ may be needed for future nutritional epidemiological studies focusing on micronutrients in Bangladesh.

In the present study, FD were administered after the FFQ, and therefore the process of FD would not have affected the accuracy of the FFQ. Importantly, unlike the FFQ, which was entirely based on recall, food consumption in FD was ascertained on a meal-to-meal basis, and therefore it is implausible that the 15 min FFQ could have appreciably affected the consumption in the two 7 d FD that took place on average 76 and 190 d subsequent to the FFQ, respectively. On the other hand, the detailed process of the FD is likely to artificially improve the accuracy of the

FFQ if the FFQ were administered shortly afterwards (Willett *et al.* 1985; Willett & Lenart, 1998). Due to difficulties in controlling the timing and limited resources, we did not administer a second FFQ, and therefore reproducibility of the FFQ was not assessed. It is difficult to judge the impact of this on the study findings. While changes in dietary behaviours could have taken place during the period of time between the FFQ and FD, the high correlations between the two 7 d FD conducted in two different seasons for macronutrients suggest that diet remained relatively unchanged for common foods over the two seasons in a year. In the follow-up visits of the HEALS, a full FFQ was not administered, and rather only changes in diet since baseline interview were queried. Less than 5 % of subjects reported changes in intakes of any food items at the first follow-up visit, indicating relatively stable dietary behaviours in this population.

Other strengths of the present study include the representativeness of the study subjects to the underlying study population, the use of FD and the use of multiple food composition tables. Regardless of the differences in absolute values of dietary intakes derived from the USDA and Indian composition tables, the patterns of correlations computed based on the different food composition tables were similar. Since nutrient intakes derived from the Indian food composition table should be more representative of food items consumed in Bangladesh, future analysis

**Table 5.** Correlations of average daily nutrient intakes obtained from the two food diaries (Intra-class correlation coefficients (ICC) and 95% confidence intervals)

| Nutrients* | ICC† | 95% CI |
|---|---|---|
| Vitamin $B_6$ (mg) | 0.70 | 0.63, 0.76 |
| Energy (kJ) | 0.64 | 0.51, 0.72 |
| Polyunsaturated fat (g) | 0.61 | 0.53, 0.68 |
| Carbohydrate (g) | 0.61 | 0.52, 0.68 |
| Vitamin E (mg) | 0.61 | 0.53, 0.69 |
| Mn (mg) | 0.59 | 0.51, 0.67 |
| Niacin (mg) | 0.58 | 0.49, 0.66 |
| Riboflavin (mg) | 0.56 | 0.47, 0.65 |
| Mg (mg) | 0.56 | 0.46, 0.64 |
| Thiamin (mg) | 0.55 | 0.46, 0.63 |
| Zn (mg) | 0.54 | 0.44, 0.63 |
| Na (mg) | 0.54 | 0.44, 0.63 |
| K (mg) | 0.53 | 0.43, 0.62 |
| Total fat (g) | 0.52 | 0.41, 0.61 |
| Protein (g) | 0.51 | 0.41, 0.60 |
| Cu (mg) | 0.49 | 0.39, 0.59 |
| Ca (mg) | 0.43 | 0.32, 0.53 |
| Saturated fat (g) | 0.41 | 0.30, 0.52 |
| Monounsaturated fat (g) | 0.40 | 0.28, 0.51 |
| Dietary fibre (g) | 0.40 | 0.30, 0.50 |
| Vitamin $B_{12}$ (mg) | 0.38 | 0.25, 0.50 |
| Fe (mg) | 0.37 | 0.24, 0.49 |
| Se (mg) | 0.33 | 0.20, 0.46 |
| Cholesterol (mg) | 0.27 | 0.11, 0.40 |
| Retinol equivalents (mg) | 0.21 | 0.07, 0.35 |
| Vitamin C (mg) | 0.17 | 0.06, 0.29 |

\* Food items were listed in the descending order of the ICC.
† Calculated based on log-transformed intake values.

of the cohort study will be based on these values. Nutrient values based on USDA data could be of interest if these nutrients are not available in the Indian food composition table. For example, values of different fat components including monounsaturated fat, polyunsaturated fat and saturated fat derived based on USDA data may be considered, provided that correlations of total fat were strong for values computed based on both USDA and Indian nutrient data. Although the two 7 d FD required labour-intensive work, FD are considered superior to 24 h dietary recalls for assessing the validity of FFQ. The care of the directly measured 14 d of recording of the FD as well as the large sample size in the parent study made the validation study all the more worthwhile. Distributions of socio-demographic characteristics of participants in the validation study were very similar to those of the overall cohort study participants. Therefore, the results demonstrated in the present study should be generalisable to the overall study population of the parent study.

In summary, the semi-quantitative thirty-nine-item FFQ used in this population of rural Bangladesh provided reasonably valid measurements for long-term dietary intakes of common foods and nutrients. However, we are uncertain about the validity of the FFQ in measuring micronutrient intakes. Given that subjects in the present study are representative of the overall study population of the parent cohort study and that the format of the FFQ is simple, we conclude that the FFQ is a useful tool for assessing long-term intakes of common foods and nutrients in the cohort study.

## References

Abdullah M & Ahmed L (1993) Validating a simplified approach to the dietary assessment of vitamin A intake in preschool children. *Eur J Clin Nutr* **47**, 115–122.

Abdullah M & Wheeler EF (1985) Seasonal variations, and the intra-household distribution of food in a Bangladeshi village. *Am J Clin Nutr* **41**, 1305–1313.

Beaton GH, Milner J, Corey P, *et al.* (1979) Sources of variance in 24-hour dietary recall data: implications for nutrition study design and interpretation. *Am J Clin Nutr* **32**, 2546–2549.

British Geological Survey (1999) Arsenic contamination of groundwater. Bangladesh phase 1. http://www.bgs.ac.uk/arsenic/

Brown KH, Black RE, Robertson AD & Becker S (1985) Effects of season and illness on the dietary intake of weanlings during longitudinal studies in rural Bangladesh. *Am J Clin Nutr* **41**, 343–355.

Erkkola M, Karppinen M, Javanainen J, Rasanen L, Knip M & Virtanen SM (2001) Validity and reproducibility of a food frequency questionnaire for pregnant Finnish women. *Am J Epidemiol* **154**, 466–476.

Feskanich D, Rimm EB, Giovannucci EL, Colditz GA, Stampfer MJ, Litin LB & Willett WC (1993) Reproducibility and validity of food intake measurements from a semiquantitative food frequency questionnaire. *J Am Diet Assoc* **93**, 790–796.

Gopalan C, Rama Sastri BV & Balasubramanian SC (1989) *Nutritive Value of Indian Foods*. Hyderabad, India: Indian Council of Medical Research, National Institute of Nutrition.

Hsueh YM, Cheng GS, Wu MM, Yu HS, Kuo TL & Chen CJ (1995) Multiple risk factors associated with arsenic-induced skin cancer: effects of chronic liver disease and malnutritional status. *Br J Cancer* **71**, 109–114.

Hsueh YM, Ko YF, Huang YK, Chen HW, Chiou HY, Huang YL, Yang MH & Chen CJ (2003) Determinants of inorganic arsenic methylation capability among residents of the Lanyang Basin, Taiwan: arsenic and selenium exposure and alcohol consumption. *Toxicol Lett* **137**, 49–63.

Hsueh YM, Wu WL, Huang YL, Chiou HY, Tseng CH & Chen CJ (1998) Low serum carotene level and increased risk of ischemic heart disease related to long-term arsenic exposure. *Atherosclerosis* **141**, 249–257.

Jain M, Howe GR & Rohan T (1996) Dietary assessment in epidemiology: comparison on food frequency and a diet history questionnaire with a 7-day food record. *Am J Epidemiol* **143**, 953–960.

Liu K, Stamler J, Dyer A, McKeever J & McKeever P (1978) Statistical methods to assess and minimize the role of intra-individual variability in obscuring the relationship between dietary lipids and serum cholesterol. *J Chronic Dis* **31**, 399–418.

Misbahuddin M (2003) Consumption of arsenic through cooked rice. *Lancet* **361**, 435–436.

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.

Roychowdhury T, Uchino T, Tokunaga H & Ando M (2002) Survey

of arsenic in food composites from an arsenic-affected area of West Bengal, India. *Food Chem Toxicol* **40**, 1611–1621.

Salvini S, Hunter DJ, Sampson L, Stampfer MJ, Colditz GA, Rosner B & Willett WC (1989) Food-based validation of a dietary questionnaire: the effects of week-to-week variation in food consumption. *Int J Epidemiol* **18**, 858–867.

Shrout PE & Fleiss JL (1979) Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* **86**, 420–428.

United States Department of Agriculture (2002) Agricultural Research Service Nutrient Data Laboratory Home Page. USDA Nutrient Database for Standard Reference, Release 15. http://www.nal.usda.gov/fnic/foodcomp/Data/SR15/sr15.html

van Geen A, Ahsan H, Horneman AH, *et al.* (2002) Promotion of well-switching to mitigate the current arsenic crisis in Bangladesh. *Bull World Health Organ* **80**, 732–737.

Walker AM & Blettner M (1985) Comparing imperfect measures of exposure. *Am J Epidemiol* **121**, 783–790.

Willett W (1987) Nutritional epidemiology: issues and challenges. *Int J Epidemiol* **16**, 312–317.

Willett W & Lenart E (1998) Reproducibility and validity of food-frequency questionnaires. In *Nutritional Epidemiology*, pp. 101–147 [WC Willett, editor]. New York: Oxford University Press.

Willett W & Stampfer MJ (1986) Total energy intake: implications for epidemiologic analyses. *Am J Epidemiol* **124**, 17–27.

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.

**Appendix 1.** Food items listed in the food-frequency questionnaire

| Food groups | Food item name (description) | Local name |
| --- | --- | --- |
| Milk | Milk | Doodh |
| Rice | Steamed rice (hand-pounded) | Bhat |
| | Puffed rice (hand-pounded) | Muri |
| | Water rice (hand-pounded) | Panta bhat |
| Bread | Wheat bread (brown) | Roti |
| Tea | Tea | Cha |
| Beef and lamb | Red meat (Beef or mutton) | Mangso or gosto |
| Beans | Lentil | Daal |
| | Beans (scarlet runner) | Sim |
| Potato | Potato | Aloo |
| | Sweet potato | Mistialoo |
| Poultry | Poultry (duck or fowl) | Murgi or hansh |
| Vegetables | Yam | Gherkochoo |
| | Pumpkin | Kumra |
| | Spinach | Palang sagh |
| | Spinach stalks | Data |
| | Brinjal (a kind of egg-plant) | Begun |
| | Okra (lady's fingers) | Dherash |
| | Bitter gourd (a kind of squash) | Karela |
| | Bottle gourd (a kind of squash) | Lau or koddu |
| | Ridge gourd (a kind of squash) | Jhinga |
| | Snake gourd (a kind of squash) | Chachinga |
| | Cauliflower | Fool kopee |
| | Cabbage | Bandha kopee |
| | Tomato | Tomato |
| | Parwar (a kind of squash) | Patol |
| | Ghosala (a kind of squash) | Dhundal |
| | Green papaya | Kancha pepe |
| | Radish | Mula |
| Fruits | Watermelon | Tarmuj |
| | Mango | Aam |
| | Banana | Kala |
| | Jack fruit | Kanthal |
| | Guava | Payra |
| Eggs | Egg (hen eggs) | Dim |
| Fish | Dried fish | Shootky |
| | Small fish (fresh-water) | Bacha, bele or bata |
| | Big fish (fresh-water) | Ruhi |
| | Salted fish | Elish |