

# Refinement and validation of an FFQ developed to estimate macro- and micronutrient intakes in a south Indian population

Romaina Iqbal<sup>1</sup>, Kamalasanan Ajayan<sup>2</sup>, Ankalmadagu V Bharathi<sup>3</sup>, Xiaohe Zhang<sup>4</sup>, Shofiqul Islam<sup>4</sup>, Chithhakkudam R Soman<sup>2</sup> and Anwar T Merchant<sup>4,\*</sup>

<sup>1</sup>Departments of Community Health Sciences and Medicine, Aga Khan University, Karachi, Pakistan:

<sup>2</sup>Health Action by People, Trivandrum, India: <sup>3</sup>St. John's Research Institute, Bangalore, India:

<sup>4</sup>Population Health Research Institute, Department of Clinical Epidemiology and Biostatistics, Hamilton, Ontario, Canada

Submitted 13 February 2007: Accepted 14 January 2008: First published online 7 March 2008

## Abstract

**Objective:** Potential error sources in nutrient estimation with the FFQ include inaccurate or biased recall and overestimation or underestimation of intake due to too many or too few items on the FFQ, respectively. Here we report the refinement of an FFQ that overestimated nutrient intake and its validation against multiple 24 h recalls.

**Study design:** Data on 2527 participants in south India (Trivandrum) were available for the original FFQ (OFFQ) that overestimated nutrient intake (132 food items). After excluding participants with implausible energy intake estimates ( $<2.72$  MJ/d ( $<650$  kcal/d),  $>15.69$  MJ/d ( $>3750$  kcal/d)) we ran stepwise regression analyses with selected nutrients as the outcomes and food intake (servings/d) as predictor variables ( $n$  1867). From these results and expert consultation we refined the FFQ (RFFQ), and validated it by comparing intakes obtained with it and the mean of two 24 h recalls among 100 participants.

**Results:** The OFFQ overestimated usual daily nutrient intake before and after exclusions [for energy:  $13.39$  (SD  $5.46$ ) MJ ( $3201$  (SD  $1305$ ) kcal) and  $10.96$  (SD  $2.65$ ) MJ ( $2619$  (SD  $634$ ) kcal), respectively]. In stepwise analyses, fifty-seven food items explained 90% of the variance in nutrients; we retained thirteen food items because participants consumed them at least twice monthly and twelve food items that local nutritionists recommended. Mean energy intake estimated from the RFFQ (eighty-two food items) was  $7.94$  (SD  $2.05$ ) MJ ( $1897$  (SD  $489$ ) kcal). The de-attenuated correlations between mean 24 h recall and RFFQ intakes ranged from 0.25 (vitamin A) to 0.82 (fat).

**Conclusion:** We refined an FFQ that overestimated nutrient intake by shortening and redesigning, and validated it by comparisons with 24 h dietary recall data.

## Keywords

FFQ

Developing country

Over-reporters

India

Overestimated nutrients

Culture- or region-specific FFQ are developed to assess dietary intake because foods vary by culture and region<sup>(1–3)</sup>. The FFQ consists of a list of foods eaten commonly in a particular region or by a particular population, each food's commonly eaten portion size and the reported intake frequency. The FFQ food list typically explains 80–90% of the variability in the nutrients of interest. The ideal method to derive such a food list is to run several backward stepwise regression models with the nutrients of interest as the outcome and a long list of candidate foods as predictors. However, such an approach requires dietary data from about 1000 to 2000 persons in the study population assessed by some alternative method, such as 24 h dietary recall or food records<sup>(4)</sup>. Because this approach is expensive it is infrequently used. The more

common method to derive a food list for the FFQ is based on smaller dietary surveys, existing instruments and expert opinion<sup>(2)</sup>. This approach is pragmatic but may be prone to error. Nutrient intake is underestimated if the foods predicting a particular nutrient of interest are under-represented on that FFQ, or overestimated owing to inflated total frequency of intake if a large list of foods representing one food group or nutrient has been included (double counting). For example, if the FFQ contains questions on intakes of rice and biryani (a rice dish that includes meat or vegetables) the respondent may report eating the same food for both questions. In the present paper we describe methods to refine an FFQ used in an ongoing epidemiological study that overestimated nutrient intake due to double counting, and we also compare nutrient intakes

\*Corresponding author: Email anwar.merchant@post.harvard.edu

assessed by the refined (RFFQ) with nutrient intakes assessed with two 24 h recalls.

## Materials and methods

### Study population

The Prospective Urban Rural Epidemiological Study (PURE) is a large, ongoing, prospective cohort study being conducted worldwide to investigate societal and individual determinants, including diet, of chronic conditions such as obesity and CVD. Briefly, data are being collected in fourteen countries (total of  $\approx 140\,000$  adults) in urban and rural areas. In India, there are five data collection sites with a target to enrol approximately 30 000 participants. Trivandrum is one of the five Indian sites where data are being gathered in urban and rural areas on 4000 participants. Ethics review boards at McMaster University, Canada as well as appropriate institutional ethics committees in India have approved the study.

There are two parts in the present analysis: first, refinement of the FFQ and, second, its validation. Data were available for 2527 participants in the refinement part of the study (1351 urban and 1169 rural, seven missing data). Data were collected using the original FFQ (OFFQ) from September 2004 to January 2005. From this we excluded participants with implausible total energy intake values ( $<2.72$  MJ/d ( $<650$  kcal/d),  $>15.69$  MJ/d ( $>3750$  kcal/d)) and had a total data set of 1867 participants (OFFQe). We then conducted a pilot study, consisting of a convenience sample of the PURE study participants ( $n\ 100$ ), to validate the RFFQ between September 2005 and January 2006.

### Original FFQ

In PURE, diet is assessed with FFQ that are specific for each population<sup>(5)</sup>. The FFQ at the Trivandrum site overestimated nutrient intake. This was a 132-item quantitative FFQ that was developed from 24 h recalls in India. Participants reported the usual portion size of each food in the questionnaire and how often on average they consumed it in the previous year. Participants saw different sized cups, plates, bowls and other utensils to help them estimate portion size. Intake frequency consisted of four categories: never, monthly, weekly and daily. As fruits and vegetables availability is seasonal, this was considered when estimating their intake. Fruit and vegetable season duration was determined by interviewing vendors; the median reported value in months was considered as the duration of the season.

To estimate nutrient intake we multiplied the reported intake frequency of each food on the FFQ by the reported portion size and its respective nutrient composition, summing over all foods. The composition of raw food items was determined from the Indian food composition table<sup>(6)</sup>. In certain cases where this information was not available in the Indian food composition table, McCance

and Widdowson's *The Composition of Foods*<sup>(7)</sup> and the US Department of Agriculture's National Nutrient Database for Standard Reference release 19 (USDA, Washington, DC, USA) were consulted. For prepared foods, we collected recipes and verified them by preparing the foods in a metabolic kitchen or in the participants' homes. We used the reference food composition table to estimate nutrient content, accounting for preparation method.

As this FFQ provided implausible values of nutrient intake we explored for its possible reasons. We excluded participants with implausible nutrient estimates, i.e. daily energy intake of  $<2.72$  MJ ( $<650$  kcal) or  $>15.69$  MJ ( $>3750$  kcal). We named the FFQ data after excluding over- and under-reporters for energy intake as 'OFFQe'. We also systematically explored for other potential sources of error, such as errors in data entry, the food composition table, recipe analysis and interviewing techniques, and identified the likely cause to be double counting of foods. For instance, an individual may have reported eating biryani (a rice dish) and then counted it again when reporting rice intake. To reduce the likelihood of this error we shortened this FFQ.

### Shortened FFQ

To shorten the FFQ we used stepwise regression analyses with all items on the OFFQe (132 items) as independent variables and the nutrients of interest as dependent variables as suggested by Willett<sup>(8)</sup> and described below under 'Statistical analyses'. We also considered how often the food item was eaten and knowledge of local food items while refining the food list.

### Design for validation of the refined FFQ

A subset of participants from the main Trivandrum study population was invited to participate in the FFQ validation study. These participants completed two FFQ and two 24 h recall forms over 4 months. The first refined FFQ (RFFQ1) and 24 h recall were administered in September 2005, and the second refined FFQ (RFFQ2) and 24 h recall in November 2005. Trained field staff interviewed the participants at their homes for dietary 24 h recalls. It took 30 minutes to conduct a 24 h recall. Various aids were used during the interview to assist in portion size estimation for the 24 h recalls.

### Statistical analyses

#### Shortening and refinement of the FFQ

To shorten the FFQ we ran a series of stepwise regression analyses with all items on the OFFQ (132 items) as independent variables. The dependent variables included energy, protein, carbohydrate, fat, SFA, fibre, vitamin A, vitamin C, Ca, folate and Zn. The  $P$  value for a variable to enter into the model was 0.10, and that for it to remain was 0.05. We included all the foods that predicted any of the nutrients in these models. The SAS statistical software

package version 9.1 (SAS Institute, Cary, NC, USA) was used for all analyses.

#### Validation and reliability analysis

Mean nutrient intakes with their standard deviations were computed for the OFFQ, OFFQe, the two RFFQ and the mean of the two 24 h recalls. Nutrient estimates were log-transformed as they tended to be skewed positively. Pearson product-moment correlations between intakes estimated by the FFQ and those calculated from the recalls were computed. We corrected for errors in nutrient comparisons arising from within-person variation as described by others<sup>(8–10)</sup>. We assessed the crude as well de-attenuated correlations for nutrient estimates. We also assessed the reliability of the RFFQ by calculating intra-class correlation coefficients between energy-adjusted nutrient estimates for RFFQ1 and RFFQ2. To estimate the degree of bias in nutrient estimates obtained from the RFFQ, the regression was performed of energy-adjusted nutrient intakes estimated from the mean of the two 24 h recalls as the outcome *v.* those from the RFFQ2 as the predictor.

## Results

The mean age of the participants (after excluding participants who over- and under-reported total energy intake, *n* 662) included in the main study was 50.7 (SD 10.1) years, with 73.9% being women. The mean BMI of this population was 24.1 (SD 4.1) kg/m<sup>2</sup>; the population was overweight on average according to WHO guidelines for assessing overweight in South Asians ( $\geq 23$  kg/m<sup>2</sup>). Only 10.8% had received university education and 53.7% of the participants resided in urban areas while the rest lived in rural areas (Table 1).

The numbers of food items explaining 90% and 99% of the variance in nutrient intake are presented in Fig. 1. Between five and twelve food items largely explained intakes of some nutrients such as vitamin A, vitamin C and Ca (cumulative  $R^2 = 90\%$ ), while for other nutrients such as energy, Zn and SFA between nineteen and twenty-two food items were needed. Fifty-seven food items explained 90% of the variation in eleven nutrients. Likewise, fewer food items were needed to obtain a cumulative  $R^2$  of 99% for vitamin A, vitamin C and Ca, while more food items were needed to explain a similar level of variation for total energy intake and other nutrients. We shortened our food items list based on the 90% variance models for total energy, protein, fat, SFA, fibre, carbohydrate, vitamin A, vitamin C, Ca, folate and Zn. In addition to this, thirteen food items that were consumed at least twice monthly were also retained. We then expanded this list based on input from experts in the field and obtained a shortened FFQ (a list of food items in the shortened FFQ appears in the Appendix). Intake responses of this

**Table 1** Sociodemographic characteristics of the study participants (over- and under-reporters of energy intake excluded, *n* 1867): Prospective Urban Rural Epidemiological Study, Trivandrum, south India

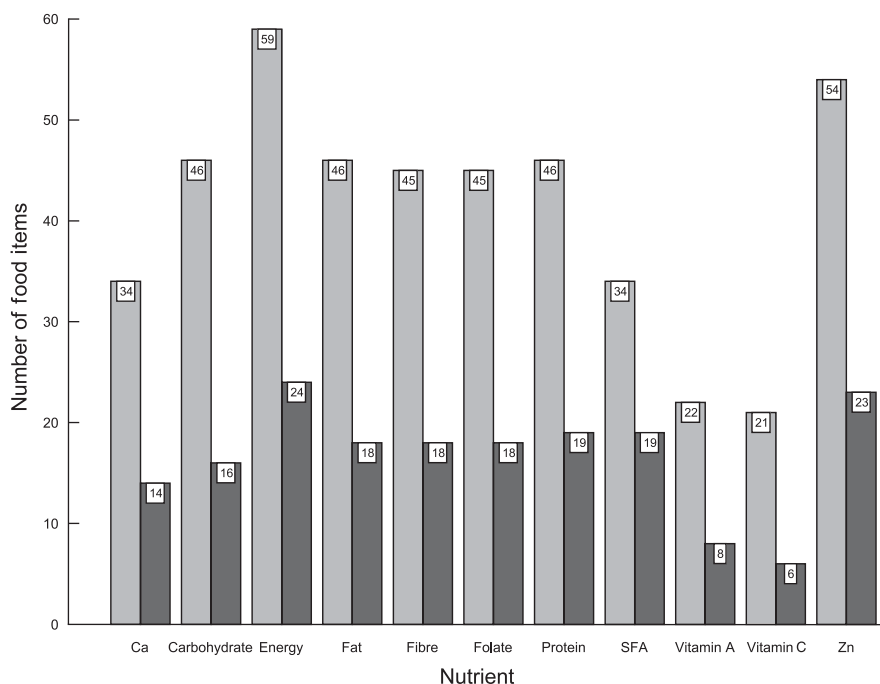
Characteristic	Mean	SD
Age (years)	50.7	10.1
BMI (kg/m <sup>2</sup> )	24.1	4.1
Waist:hip ratio	0.86	0.10
	<i>n</i>	%
Sex		
Female	1374	73.9
Marital status		
Unmarried	29	1.6
Married	1486	79.9
Widowed	287	15.4
Divorced	9	0.5
Education		
Graduate & higher	201	10.8
Location		
Urban	999	53.7
Rural	861	46.3

questionnaire were modified so that they were one of nine categories ranging from never or <1 time/month to  $\geq 6$  times/d. The portion sizes were fixed based on the median reported serving size in this population as opposed to having an open response category.

Intakes of energy and macronutrients were similar using the RFFQ1, RFFQ2 and 24 h recalls, but higher for the OFFQ and OFFQe (Table 2). Mean usual daily energy intake estimated from the OFFQ was 13.39 (SD 5.46) MJ (3201 (SD 1305) kcal), daily protein intake was 96.1 (SD 44.1) g and fat 120.8 (SD 57.7) g. Despite excluding the over- and under-reporters, the nutrient estimates were still higher from the OFFQe relative to the RFFQ and 24 h recalls. Mean usual daily energy intake estimated from OFFQe was 10.96 (SD 2.65) MJ (2619 (SD 634) kcal), protein was 77.9 (SD 22.2) g, fat 96.0 (SD 33.5) g. In contrast to the OFFQ, mean usual daily intakes estimated from the RFFQ1 were 8.31 (2.20) MJ (1985 (SD 527) kcal) for energy, 58.8 (SD 17.9) g for protein and 64.8 (SD 24.5) g for fat; while the corresponding values estimated from RFFQ2 were 7.94 (SD 2.05) MJ (1897 (SD 489) kcal), 54.5 (SD 16.0) g and 62.0 (SD 23.8) g.

Comparing RFFQ1 and 24 h recalls, the correlation coefficients ranged from 0.11 for vitamin A to 0.44 for protein intake, while the correlations for RFFQ2 ranged from 0.09 for vitamin A and SFA to 0.35 for protein intake. The de-attenuated correlations between RFFQ1 and 24 h recalls ranged from 0.25 for vitamin A to 0.82 for total fat intake. The de-attenuated correlations between RFFQ2 and 24 h recalls ranged from 0.12 for fibre to 0.49 for protein intake (Table 3).

In the analyses in which the mean of the two 24 h recall intakes was the outcome and nutrient intake from RFFQ2 was the predictor, most food items such as carbohydrate, Zn and the vitamins had very small coefficients indicating



**Fig. 1** Results of regression analyses of FFQ-derived data showing the number of food items required to explain different levels of between-person variation (□, explained 99% of the variation in nutrient intake; ■, explained 90% of the variation in nutrient intake) for selected nutrients: Prospective Urban Rural Epidemiological Study, Trivandrum, south India

**Table 2** Mean daily nutrient intakes estimated by the FFQ as well the 24 h recalls: Prospective Urban Rural Epidemiological Study, Trivandrum, south India

Variable	OFFQ ( <i>n</i> 2527)		OFFQe ( <i>n</i> 1867)		Mean of two 24 h recalls ( <i>n</i> 100)		RFFQ1 ( <i>n</i> 100)		RFFQ2 ( <i>n</i> 100)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Energy (MJ)	13.39	5.46	10.96	2.65	8.74	2.80	8.31	2.20	7.94	2.05
Energy (kcal)	3201	1305	2619	634	2088	669	1985	527	1897	489
Protein (g)	96.1	44.1	77.9	22.2	61.7	32.2	58.8	17.9	54.5	16.0
Carbohydrate (g)	426.5	163.7	358.3	89.0	321.5	108.0	289.8	64.1	277.2	60.6
Fibre (g)	18.9	8.7	15.2	5.0	13.1	6.1	14.0	5.4	13.2	5.7
Fat (g)	120.8	57.7	96.0	33.5	60.6	22.6	64.8	24.5	62.0	23.8
SFA (g)	89.5	42.8	71.2	23.0	44.6	17.1	45.2	18.4	43.7	18.1
Vitamin A (RE)	1205.6	702.9	967.3	436.9	877.4	589.3	2996.5	716.9	2895.1	105.6
Vitamin C (mg)	174.7	107.4	145.0	77.8	75.5	57.2	142.0	55.3	129.0	47.0
Ca (g)	1233.8	607.1	1009.0	344.7	529.6	308.4	864.5	289.6	827.0	53.8
Zn (mg)	12.1	5.4	9.8	2.5	8.6	3.5	9.8	3.0	8.7	2.2
Folate (mg)	361.2	167.6	299.3	109.4	194.6	100.1	268.5	86.7	248.1	70.9

OFFQ, original FFQ; OFFQe, original FFQ after excluding over- and under-reporters; RFFQ1, refined FFQ first administration; RFFQ2, refined FFQ second administration; RE, retinol equivalents.

little bias. The RFFQ underestimated total energy, protein and Ca intakes but overestimated intakes of fat and folate (Table 3). The intra-class correlations for the nutrients computed from RFFQ1 and RFFQ2, assessing the reliability of the RFFQ, ranged from 0.26 for vitamin C to 0.51 for Ca intake.

## Discussion

We refined an FFQ that overestimated nutrient intake in an ongoing study in a developing country. We systematically

addressed, to the extent possible, potential sources of error in the estimation of nutrient intakes, which included verifying that the food composition table, recipes and data entry were accurate, to eliminating double counting by systematically shortening the FFQ by regression analyses. We also reformatted the older, quantitative FFQ into a semi-quantitative instrument. Reasonable estimates were obtained when the RFFQ was validated against multiple 24 h recalls. The RFFQ on average took less time to administer (10 minutes) than the OFFQ (18 minutes). While these techniques are described in the literature on

**Table 3** Crude and de-attenuated correlations between nutrient estimates from RFFQ1 and RFFQ2 respectively and the mean of two 24 h recalls; beta coefficients ( $\beta$ ) from the regression of adjusted nutrient intakes estimated from the mean of the two 24 h recalls as the outcome *v.* those from RFFQ2 as the predictor; and intra-class correlation coefficients (ICC) between energy-adjusted nutrient estimates for RFFQ1 and RFFQ2: Prospective Urban Rural Epidemiological Study, Trivandrum, south India

Nutrient*	RFFQ1		RFFQ2		$\beta$	ICC
	Crude	De-attenuated	Crude	De-attenuated		
Energy	0.35	0.69	0.20	0.39	0.25	0.45
Protein	0.44	0.62	0.35	0.49	1.16	0.46
Carbohydrate	0.29	0.53	0.13	0.24	0.03	0.46
Fibre	0.25	0.52	0.06	0.12	-0.13	0.34
Fat	0.31	0.82	0.11	0.30	-0.32	0.40
SFA	0.28	0.80	0.09	0.28	-0.28	0.43
Vitamin A	0.11	0.25	0.09	0.20	0.02	0.43
Vitamin C	0.17	0.32	0.11	0.21	-0.07	0.26
Ca	0.28	0.48	0.21	0.36	0.26	0.51
Zn	0.35	0.49	0.21	0.30	0.03	0.37
Folate	0.28	0.50	0.18	0.33	-0.26	0.39

RFFQ1, refined FFQ first administration; RFFQ2, refined FFQ second administration.

\*Log-transformed nutrients.

FFQ development<sup>(11)</sup>, we did not find any example of them being applied to improving estimation of existing instruments. This application may be particularly important in epidemiological studies, as often data collection and FFQ validation occur in parallel<sup>(12)</sup>.

The results from the RFFQ were plausible and consistent. For example, fewer food items predicted 90% of the vitamin A and C intake (concentrated in a few food items), as opposed to a larger number of food items that predicted total energy intake (Fig. 1). Our findings are similar to those observed by others in which few food items were needed to explain some nutrients due to their limited occurrence in foods, while the more ubiquitous nutrients were explained by more food items<sup>(4,13)</sup>. The mean nutrient intakes estimated by the RFFQ were similar to those obtained from the 24 h recall, and within the range reported by others in South Asia<sup>(14,15)</sup>. In an Indian investigation, mean usual daily energy intake was observed to be 7.32 MJ (1749 kcal) and 7.99 MJ (1910 kcal) in the urban and rural population, respectively<sup>(14)</sup>. Likewise, in a study conducted in south India, the mean daily energy intake was 8.64 (SD 1.83) MJ (2066 (SD 437) kcal) for men and 7.30 (SD 1.44) MJ (1745 (SD 343) kcal) for women<sup>(15)</sup>. The de-attenuated correlation coefficients we observed between RFFQ1 and the 24 h recalls (ranging from 0.25 to 0.82) in the present study were similar to those reported in other validation studies (range of 0.32–0.61 for a study conducted in Kerala<sup>(16)</sup> and 0.55–1.00 for a study conducted in Gujarat<sup>(17)</sup>). In agreement with other reports<sup>(17)</sup>, correlations observed in the present study for vitamin A and C were lower compared with other nutrients. As vitamins A and C are concentrated in a few foods they tend to have high-within person variability and lower correlation coefficients in validation studies, as reported elsewhere<sup>(17)</sup>. Consistent with other studies, energy intake estimates from the FFQ were lower than those obtained by the 24 h recall<sup>(17)</sup>.

Fewer food items explained 90% of variation in our analysis compared with the 126 food items in an Israeli population using a similar approach<sup>(18)</sup>. This may have been because the participants in the Israeli study had varied ethnic and cultural backgrounds, but our population was homogeneous. Moreover, in poorer communities there are fewer food choices and consequently there is a smaller range of dietary variation<sup>(4)</sup>. The number of food items (132 on the OFFQ) might not seem excessive by Western standards, but was probably large for an FFQ in this ethnically homogeneous, lower-income Trivandrum community. We also changed the FFQ format so that the question asked the participants about average portion sizes of foods consumed (obtained from the 24 h recall). By doing so we eliminated the need to show participants visual aids to estimate portion size. We adopted this strategy because it has been reported that frequency of intake alone explains 84% of the variance in nutrient intake, and addition of open-ended questions on portion size may increase respondent burden and the chances of incomplete data<sup>(19)</sup>. Similar findings have also been reported before<sup>(20)</sup>. We also changed the response categories into a 9-point ordinal scale adapted from other validated FFQ<sup>(21,22)</sup>. The RFFQ was therefore semi-quantitative. This change required the interviewer to check a category rather than write a number, and asked the participant to estimate a range of usual intake rather than provide an exact number.

Some limitations of our work merit consideration. The correlations we observed in the present study were in general lower than those reported by others who compared FFQ data to several weeks of diet records<sup>(9,15,23)</sup>, but similar to estimates comparing FFQ data to multiple 24 h recalls<sup>(24)</sup> generally and to studies done in the subcontinent in particular<sup>(25,26)</sup>. A possible reason why our correlations are lower than those reported for FFQ validated against diet records may be that we had data



from only two 24h recalls as a reference method, as opposed to estimates from several days considered by others<sup>(21,26,27)</sup>. Moreover, we observed that the correlations for RFFQ2 with the mean of the two 24h recalls were lower than those from RFFQ1. Generally, the correlation coefficients improve for the second FFQ owing to a learning effect or similar reference period for the two dietary methods. The lower correlations observed in the present study may be because of participant fatigue, as the second 24h recall and RFFQ2 were administered simultaneously.

Although it has been suggested<sup>(8)</sup> that stepwise regression analysis can be performed to develop an FFQ food item list, this may not be the optimal strategy if the sample size is less than 1000 to 2000, as some food items may not enter the statistical model due to inadequate power. The strength of our work is that we had a large sample size, which reduces the chances of beta error.

Lengthy FFQ have the potential to overestimate nutrient intake and also are not feasible to administer from a logistics point of view. Shorter FFQ take less time to administer and also provide valid and reliable nutrient estimates. We have shared a strategy for refining lengthy questionnaires and arriving at reasonable estimates of nutrient intake in India. Other researchers in the field may be able to adapt this approach to obtain more valid nutrient estimates from existing FFQ.

## Acknowledgements

The authors wish to thank Dr Salim Yusuf, Director of the Population Health Research Institute, McMaster University and Principal Investigator for the PURE study, for providing us the opportunity to analyse and report these data. We also thank Joseph Michael for diligently entering the data from the FFQ as well as the 24h recalls.

*Author contributions:* Data collection – K.A., C.R.S., A.V.B.; study design – A.T.M., R.I.; analyses – X.Z., R.I., S.I.; writing – R.I., A.T.M.; critical revisions – R.I., K.A., A.V.B., X.Z., S.I., C.R.S., A.T.M.

None of the authors reported any conflict of interest.

*Sources of funding:* Funding for this study was provided by the Population Health Research Institute, Hamilton, Canada.

## References

1. Feskanich D, Rimm EB, Giovannucci EL, Colditz GA, Stampfer MJ, Litin LB & Willett WC (1993) Reproducibility and validity of food-intake measurements from a semi-quantitative food frequency questionnaire. *J Am Diet Assoc* **93**, 790–796.
2. Kelemen LE, Anand SS, Vuksan V, Yi QL, Teo KK, Devanesen S & Yusuf S (2003) Development and evaluation of cultural food frequency questionnaires for South Asians, Chinese, and Europeans in North America. *J Am Diet Assoc* **103**, 1178–1184.
3. Jain M (1999) Culture specific food frequency questionnaires: development for use in a cardiovascular study. *Can J Diet Pract Res* **60**, 27–36.
4. Shai I, Shahar DR, Vardi H & Fraser D (2004) Selection of food items for inclusion in a newly developed food-frequency questionnaire. *Public Health Nutr* **7**, 745–749.
5. Dehghan M, Al Hamad N, Yusufali A, Nusrath F, Yusuf S & Merchant AT (2005) Development of a semi-quantitative food frequency questionnaire for use in United Arab Emirates and Kuwait based on local foods. *Nutr J* **27**, issue 4, 18.
6. Gopalan C, Sastari R & Balasubramanian S (1991) Food composition tables. In *Nutritive Value of Indian Foods*, revised ed., pp. 45–80 [BS Narasinga Rao, YG Deothale and KC Pant, editors]. Hyderabad: National Institute of Nutrition.
7. Krebs J (2002) *McCance and Widdowson's The Composition of Foods: Summary Edition*, 6th summary ed. Cambridge/London: The Royal Society of Chemistry/Food Standards Agency.
8. Willett W (1998) *Nutritional Epidemiology*, 2nd ed. New York: Oxford University Press.
9. Fornes NS, Stringhini MLF & Elias BM (2003) Reproducibility and validity of a food-frequency questionnaire for use among low-income Brazilian workers. *Public Health Nutr* **6**, 821–827.
10. 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.
11. Merchant AT, Dehghan M, Chifamba J, Terera G & Yusuf S (2005) Nutrient estimation from an FFQ developed for a black Zimbabwean population. *Nutr J* **13**, issue 4, 37.
12. Shu XO, Yang G, Jin F, Liu D, Kushi L, Wen W, Gao Y-T & 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.
13. Stryker WS, Salvini S, Stampfer MJ, Sampson L, Colditz GA & Willett WC (1991) Contributions of specific foods to absolute intake and between-person variation of nutrient consumption. *J Am Diet Assoc* **91**, 172–178.
14. Chadha SL, Gopinath N, Katyal I & Shekhawat S (1995) Dietary profile of adults in an urban & a rural community. *Indian J Med Res* **101**, 258–267.
15. Shobana R, Snehalatha C, Latha E, Vijay V & Ramachandran A (1998) Dietary profile of urban south Indians and its relations with glycaemic status. *Diabetes Res Clin Pract* **42**, 181–186.
16. Hebert JR, Gupta PC, Bhonsle RB, Murti PR, Mehta H, Verghese F, Aghi M, Krishnaswamy K & Mehta FS (1998) Development and testing of a quantitative food frequency questionnaire for use in Kerala, India. *Public Health Nutr* **1**, 123–130.
17. Hebert JR, Gupta PC, Bhonsle RB, Sinor PN, Mehta H & Mehta FS (1999) Development and testing of a quantitative food frequency questionnaire for use in Gujarat, India. *Public Health Nutr* **2**, 39–50.
18. Shahar D, Shai I, Vardi H, Brenner-Azrad A & Fraser D (2003) Development of a semi-quantitative food frequency questionnaire (FFQ) to assess dietary intake of multiethnic populations. *Eur J Epidemiol* **18**, 855–861.
19. Noethlings U, Hoffmann K, Bergmann MM & Boeing H (2003) Portion size adds limited information on variance in food intake of participants in the EPIC–Potsdam study. *J Nutr* **133**, 510–515.
20. Tjonneland A, Haraldsdottir J, Overvad K, Stripp C, Ewertz M & Jensen OM (1992) Influence of individually estimated

- portion size data on the validity of a semi-quantitative food frequency questionnaire. *Int J Epidemiol* **21**, 770–777.
21. Martin-Moreno JM, Boyle P, Gorgojo L, Maisonneuve P, Fernandez-Rodriguez JC, Salvini S & Willett WC (1993) Development and validation of a food frequency questionnaire in Spain. *Int J Epidemiol* **22**, 512–519.
  22. Willett WC, Sampson L, Stampfer MJ, Rosner B, Bain C, Witschi J, Hennekens CH & Speizer FE (1985) Reproducibility and validity of a semi quantitative food frequency questionnaire. *Am J Epidemiol* **122**, 51–65.
  23. Toft U, Kristoffersen L, Ladelund S, Bysted A, Jakobsen J, Lau C, Jørgensen T, Borch-Johnsen K & Ovesen L (2007) Relative validity of a food frequency questionnaire used in the Inter99 study. *Eur J Clin Nutr* (Epublication ahead of print version).
  24. Shatenstein B, Nadon S, Godin C & Ferland G (2005) Development and validation of a food frequency questionnaire. *Can J Diet Pract Res* **66**, 67–75.
  25. Chen Y, Ahsan H, Parvez F & Howe GR (2004) Validity of a food-frequency questionnaire for a large prospective cohort study in Bangladesh. *Br J Nutr* **92**, 851–859.
  26. Pandey D, Bhatia V, Boddula R, Singh HK & Bhatia E (2005) Validation and reproducibility of a food frequency questionnaire to assess energy and fat intake in affluent north Indians. *Natl Med J India* **18**, 230–235.
  27. Sevak L, Mangtani P, McCormack V, Bhakta D, Kassam-Khamis T & Silva ID (2004) Validation of a food frequency questionnaire to assess macro- and micro-nutrient intake among South Asians in the United Kingdom. *Eur J Nutr* **43**, 160–168.

## Appendix – List of food items for the refined FFQ

1.	Idli	43.	Unniappam
2.	Vada	44.	Liver, brain, kidney, etc.
3.	Dosa	45.	Chicken curry
4.	Chapatti/paratha	46.	Fish fry
5.	Roti	47.	Fish curry
6.	Puri/bhatura	48.	Fish peera
7.	Puttu (rice/wheat/rava)	49.	Prawns, crab, shellfish, mussels, etc.
8.	Appam	50.	Dried fish, dried seafood curry
9.	Idiappam	51.	Dried fish, dried seafood fry
10.	Kozhikatta/ada	52.	Egg omelette
11.	Upma (all types)	53.	Plain milk
12.	Bread	54.	Flavoured milk (Bournvita, Horlicks)
13.	Corn flakes/rice flakes	55.	Curd
14.	White boiled rice	56.	Buttermilk
15.	Kanji/pazhamhkanji	57.	Ice cream
16.	Vegetable fried rice	58.	Custard apple
17.	Biryani	59.	Water melon
18.	Rasam/puligcurry	60.	Jackfruit
19.	Sambar/dhal curries	61.	Mango
20.	Thiyal	62.	Apple
21.	Erussery	63.	Orange
22.	Tapioca curry	64.	Guava
23.	Whole gram curry	65.	Grapes
24.	Jackfruit/jackfruit seed curry	66.	Pineapple
25.	Pulissery/moru curry	67.	Papaya
26.	Green leafy vegetable curry	68.	Dried fruits
27.	Tomato curry	69.	Papad fried
28.	Aviyal	70.	Mixture
29.	Coconut chutney	71.	Nuts
30.	Raw vegetable salad	72.	Chips
31.	Curd salad	73.	Vazhakyappam (banana roast)
32.	Amarakka/beans	74.	Biscuits (sweet, cream, etc.)
33.	Jackfruit/jackfruit seed	75.	Chocolate
34.	Leafy vegetables	76.	Pickle
35.	Beetroot, yam, etc.	77.	Tea
36.	Potato	78.	Coffee
37.	Snake gourd/bitter gourd/other gourds	79.	Fresh fruit juice
38.	Raw banana	80.	Soft drinks/others, etc.
39.	Plantain flower/stem preparations	81.	Tender coconut water
40.	Ladies fingers	82.	Spirits (rum, whiskey, etc.)
41.	Brinjal	83.	Arrack
42.	Drumstick	84.	Mutton, beef, pork curry