

Selection of food items for inclusion in a newly developed food-frequency questionnaire

Iris Shai^{1,2,*}, Danit R Shahar², Hillel Vardi³ and Drora Fraser^{2,3}

¹Harvard School of Public Health, Departments of Nutrition and Epidemiology, 665 Huntington Avenue, Boston, MA 02115, USA; ²The S Daniel Abraham International Center for Health and Nutrition, Ben-Gurion University of the Negev, Beer-Sheva, Israel; ³Epidemiology and Health Services Evaluation Department, Faculty of Health Sciences, Ben-Gurion University of the Negev, Beer-Sheva, Israel

Submitted 1 October 2003; Accepted 7 January 2004

Abstract

Objectives: To highlight the differences between the food list required in a food-frequency questionnaire (FFQ) to rank people by their intake and the food items that contribute to absolute intake, and to discuss possible applications.

Methods: We conducted a nutritional survey among 1173 adults using an adapted 24-hour recall questionnaire.

Statistical analysis: To develop an FFQ, we analysed the 24-hour recall survey data by performing a stepwise multiple regression after grouping conceptually similar food items into 175 food groups.

Results: In total, 126 food groups were included in the developed FFQ in order to explain at least 80% of the variance in the consumption of each of 27 nutrients. The nutrients that were explained by a few food groups were vitamin A (one food group), alcohol (two), β -carotene (two), vitamin E (three) and cholesterol (five). Nutrients that were explained by a large number of food groups were energy (37 food groups), potassium (31), magnesium (31), dietary fibre (30), phosphorus (31) and sodium (29). Using energy intake as an example, soft drinks were the best between-person energy classifiers, while providing only 2.4% of the total energy intake. Wine, seeds and nuts, which contributed highly to the variance, were minor energy contributors. In contrast, milk, sugar, fried chicken/turkey breast or whole chicken/turkey, which explained little of the variation in the population, were major energy contributors.

Conclusions: Developing an FFQ on the basis of common foods may not explain the between-person variation required for ranking individual intake in diet–disease studies. Producing lists of ‘discriminating items’ can be a useful application in developing mini-FFQs for selected nutrients.

Keywords
Food-frequency questionnaire
24-Hour recall
Between-person variation
Absolute intake

A food-frequency questionnaire (FFQ) is designed to identify foods that discriminate between individual nutrient intakes^{1,2}. One approach to developing an FFQ is to base it upon a population survey, as done in The Netherlands³, Germany⁴, Spain⁵ and the USA⁶. The development of the food list can be based on stepwise multiple regression analyses^{1,2,7}. As demonstrated previously^{7–9}, the number of foods needed to explain a given proportion of variability in nutrient intake is less than the number needed to explain the same proportion of absolute intake.

The combination of foods that contributes most to between-person variability is unique and characteristic for each specific population. Therefore, it is necessary to determine these key foods in each different country and culture. We conducted a population-based dietary survey in order to develop an FFQ in Israel. In this paper we highlight the differences between the food list required in

the FFQ to rank people by their intake and the food items that contributed to the absolute intake, as well as demonstrate several practical uses of this knowledge.

Methods

Dietary assessment and the food database

A full description of the Negev Nutritional Study (NNS) has been given elsewhere^{10,11}. To summarise briefly, we generated a random sample population aged 35 years and above in the Negev (the southern region of Israel), using voters registries. Participants were interviewed at home regarding their dietary intake using a modified version of the US Department of Agriculture’s 24-hour recall questionnaire, with additional questions added to address general health and eating habits. We then developed a food database and data-entry system for the 24-hour recall questionnaires¹² based on the logic of the US Food

*Corresponding author. Email irish@bgumail.bgu.ac.il

Information Analysis System¹³. The system was designed as bilingual (English and Hebrew). Local foods and recipes were collected during the NNS.

Development of the FFQ

The complete process of developing our FFQ has been described elsewhere¹⁴. In short, the results of the NNS were used to develop an FFQ using the following steps. First, conceptually similar foods were aggregated into groups based on their fat and caloric content per portion eaten. Fruits and vegetables were aggregated based on their vitamins and minerals content per portion. Afterwards, in stepwise multiple regression analyses¹⁵, foods important in contributing to between-person variation were identified. Items were entered into the model based on their specific nutrient content and the frequency of intake in the population. Foods in the model that explained at least 80% of the between-person variability were considered for the final questionnaire. The procedure described was undertaken for each of 27 nutrients including energy, folic acid, calcium, vitamin E, dietary fibre, etc. The contribution to the total nutrient intake was also calculated for each of the foods and recipes included in the final questionnaire.

Statistical analysis

Statistical analyses were conducted using SPSS version 10 (SPSS Inc., Chicago, IL, USA). After aggregation of all reported items into 175 groups, the means of all foods included in the aggregated group were used in the model weighted by the frequency of consumption of that item. Stepwise regression models were used to assess the contribution of each food item to the total intake and the between-person variability, with individual food items as the independent variables and total nutrient intake as the dependent variable.

Results

During the study, 1465 24-hour recall questionnaires were obtained from 1173 participants (650 women and 523 men). This group reflects those who remained after 9.9% refusals, 2.7% language problems and 3.4% deaths in the original sample. Body mass index (BMI) was $27.0 \pm 4.9 \text{ kg m}^{-2}$ for women and $26.5 \pm 4.1 \text{ kg m}^{-2}$ for men (mean \pm standard deviation (SD)). Mean \pm SD of reported energy intake was $1498.7 \pm 875.5 \text{ kcal}$ for women and $1989.4 \pm 1138.5 \text{ kcal}$ for men. The women reported 20.0 ± 9.8 food items, 2.8 ± 0.9 meals and 6.1 ± 4.6 snacks; and the men reported 21.0 ± 10.8 food items, 2.8 ± 0.9 meals and 6.2 ± 5.4 snacks (mean \pm SD).

An average of 19 foods (range 1–37) was needed to account for the consumption of nutritional components in the 27 regression models. The nutritional components explained by a few food groups were vitamin A (one food group), alcohol (two groups), β -carotene (two groups),

vitamin E (three groups) and cholesterol (five groups). On the other hand, nutritional components that required a large number of foods were energy (37 food groups), potassium (31 groups), magnesium (31 groups), dietary fibre (30 groups), phosphorus (31 groups) and sodium (29 groups). The number of items required to explain different levels of between-person variation for selected nutrients is presented in Fig. 1 along with the number of foods required for each level of between-person variability. For energy intake, two items explained 20% of the between-person variability, seven items explained 40%, 17 items explained 60% and 37 items explained a total of 80% of the between-person variation. For calcium, 20% of the between-person variability was explained by only one item, 40% by three items, 60% by eight items and 80% by 20 items.

Tables 1 and 2 demonstrate the differences between contribution of food groups to between-person variation and total intake in the population. The earlier the stage of entry into the regression, the higher is the contribution to variance in consumption (R^2) of the food group to the nutritional component.

We detected 37 food groups that accounted for 80% of the variance in energy consumption, the top 20 of which are presented in Table 1. Most variance in energy intake (20%) in the population resulted from the consumption of carbonated beverages and fruit-flavoured drinks (not including diet). Dry wine, seeds and nuts, which entered in early stages to the regression (4, 5 and 9, respectively), were minor energy contributors (rank 72, 25 and 34, respectively). However, 3% fat milk and sugar, which entered in the later stages to the regression (16 and 17, respectively), were a major energy contributors (rank 2 and 3, respectively). Similarly, fried chicken, turkey breast and whole chicken/ turkey which entered almost at the end step of the 37 food items (31 and 33, respectively) were major energy contributors (rank 4 and 5, respectively) (data not shown).

A total of 11 items were required to explain 80% of the between-person variation in folic acid intake (Table 2).

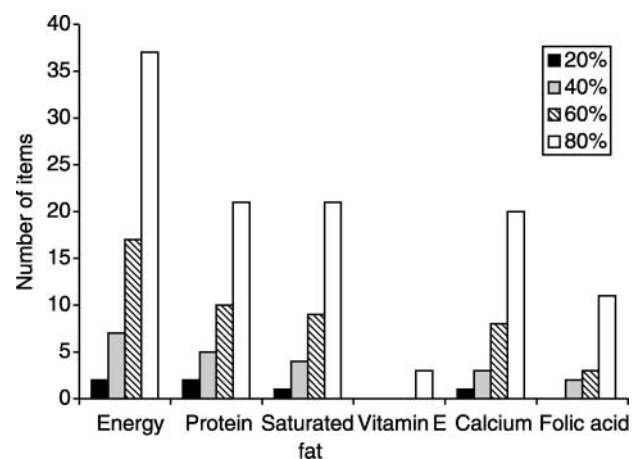


Fig. 1 Number of food items required to explain different levels of between-person variation for selected nutrients

Table 1 Contributions of the top 20 specific grouped foods to between-person variation of energy intake (using stepwise multiple regression analysis*) and to total energy intake

Stage of entry into regression	Grouped foods	Contribution to between-person variation	Contribution to total intake	
		Cumulative R^2	Percentage (rank†)	Cumulative percentage
1	Carbonated drinks, non-dietetic	0.11	2.37 (7)	2.37
2	Drinks, fruit-flavoured	0.19	2.51 (5)	4.88
3	Bread	0.24	9.33 (1)	14.21
4	Dry wine	0.29	0.34 (72)	14.55
5	Sunflower, watermelon, pumpkin seeds	0.33	1.18 (25)	15.73
6	Plain cookies	0.36	2.11 (12)	17.84
7	Dry cakes	0.38	1.46 (19)	19.30
8	Vegetable oils	0.41	2.45 (7)	21.75
9	Nuts, almonds, pistachios	0.43	0.87 (34)	22.62
10	French fries	0.46	1.39 (20)	24.01
11	Mixed meat dishes	0.48	2.32 (10)	26.33
12	Hard cheeses	0.50	1.15 (26)	27.48
13	Pita bread	0.52	1.97 (13)	29.45
14	Potatoes (cooked, baked, mashed)	0.54	2.33 (9)	31.78
15	Cooked yellow/white rice with additions	0.56	2.27 (11)	34.05
16	3% fat milk	0.57	3.04 (3)	37.09
17	Sugar	0.60	3.27 (2)	40.36
18	Individual puff pastry filled with cheese, potato	0.62	0.88 (32)	41.24
19	Cheese/cream cakes	0.63	0.66 (43)	41.90
20	Peanuts	0.64	0.27 (81)	42.17

* Dependent variable: total energy intake, independent variable: grouped foods.
 † Rank: the order of contributors to total energy intake from higher to lower.

The significant prominent group was internal organs of poultry/meat, which contributed almost half of the explained variance. Other foods contributed only between 1% and 8% of the variance. Sweetened fruit juice and breakfast cereals were good classifiers (stage 5 and 7, respectively, in the model) but minor contributors to total folic acid intake (rank 47 and 58, respectively).

Discussion

The purpose of this study was to highlight differences between the foods needed to explain the between-person variability as compared with those contributing to absolute nutrient intake. We designed an FFQ that incorporates 126

foods, regressed from 1173 survey interviewees' nutritional data. In the Israeli population, soft drinks are the best between-person energy classifiers, while providing only 2.4% of the total energy intake. About half of the energy variation (46%) was explained by soft drinks, wine, regular bread, seeds and mixed nuts, cakes, cookies, vegetable oil, French fries, beef dishes and hard cheese. However, these 10 food groups accounted for only 24% of the total energy intake.

Fewer foods are required to explain the between-person variance in energy consumption than the larger number required for evaluating total intake. Stryker *et al.*⁸ studied the correlation between the absolute contribution of foods to energy and their contribution to the variance of

Table 2 Contributions of the entire specific grouped foods to between-person variation of folic acid (using stepwise multiple regression analysis*) and to total folic acid intake

Stage of entry into regression	Grouped foods	Contribution to between-person variation	Contribution to total intake	
		Cumulative R^2	Percentage (rank†)	Cumulative percentage
1	Internal organs of poultry/meat	0.47	5.10 (4)	5.1
2	Juice, fruit-flavoured	0.55	5.29 (2)	10.39
3	Falafel in pita bread	0.61	1.89 (13)	12.28
4	Sunflower, pumpkin, watermelon seeds	0.65	2.71 (7)	14.99
5	Sweetened fruit juice (nectar)	0.69	0.54 (47)	15.53
6	Cooked dried legumes	0.72	2.15 (10)	17.68
7	Breakfast cereals	0.74	0.45 (58)	18.13
8	Vegetable soup (home-made)	0.76	3.50 (5)	21.63
9	Bread	0.78	8.66 (1)	30.29
10	Fresh cucumber	0.79	1.91 (12)	32.20
11	Orange or grapefruit	0.80	1.57 (17)	33.77

* Dependent variable: total folic acid intake, independent variable: grouped foods.
 † Rank: the order of contributors to total folic acid intake from higher to lower.

consumption. In the American population, cake was pointed out as the best classifier of energy consumption (a contribution of 19.1% to the variance between people). In order to reach a cumulative $R^2 = 0.50$, questions regarding frequency of intake of cake, butter, hard cheeses, whole milks and margarine intake were sufficient, although these foods contributed only 14% of the total energy. In the analysis during the development of an FFQ in Denmark⁷, only three foods – potatoes, butter and barley bread – were inquired in order to explain half of the variance in energy consumption. These same three jointly contributed 18.3% of the total energy. Ninety per cent of the variance in Denmark was explained by 25 foods only, a less diverse list than our own. Similar findings concerning the fewer number of foods required to identify variance were found by Byers *et al.*⁹ and Block *et al.*⁶. The longer list of 37 foods required for reaching the level of $R^2 = 0.80$ in our population is probably due to ethnic variability, characterised by different dietary patterns among people from different ethnic origins. As we have previously shown¹⁴, dietary intakes of energy and vitamin E are significantly lower among European–American born subjects and calcium intake is significantly lower among Asian–African born subjects. Differences between ethnic groups were seen in the list of foods, which explained the between-person variability for energy. For most nutrients, fewer items were needed to explain the between-person variation in the group of people born in Israel. Thus, an extended list of food items explaining energy intake variation might reflect the multi-ethnic population.

The contribution of food items to the consumption of nutritional components differs according to the specific nutritional component. Vitamins, in particular, were explained by a small number of foods. For example, 80% of the variance in the consumption of vitamin E was explained by three groups only: sunflower/pumpkin/watermelon seeds, vegetable oils and mixed nuts. On the other hand, the between-person variation of common minerals, such as potassium, nitrate, magnesium and phosphorus, was explained by the largest number of foods, as found in other populations⁷. In our survey, only 11 food groups were required to explain the between-person variation in folic acid. The food group internal organs of poultry/meat, which explained 47% of the variance, can categorise folic acid intake almost exclusively. This can be explained by the folic acid content of the foods. While it ranges from 5 to 60 μg per 100 g of food for most items, liver is a rich source of folic acid (770 μg per 100 g). If we combine this fact with the great variation of liver consumption frequency in the population, we get the primary indicator for the classification of folic acid consumption in this population, by exclusive use of liver consumption patterns.

A potential application for producing lists of ‘discriminating items’ is to develop specific mini-FFQs for selected

nutrients. For example, when designing a preventive cardiology intervention study, one could obtain a list of the foods most pertinent (e.g. cholesterol, fats, etc.) and then develop and use a mini-checklist of the best-discriminating items, thereby saving time and expense compared with the more complete traditional FFQs. The technique is also applicable when using a different method, e.g. a 24-hour recall for dietary assessment, in that one could add a list of the foods responsible for the major between-person variation following the open questionnaire. This list could include, for our own population, contributors such as soft drinks for energy, mixed meat dishes for saturated fat and internal organs for folic acid.

Some limitations of the regression analysis merit consideration¹. Because the original food list may include several hundred potential food variables, some will enter as ‘statistically significant’ variables on the basis of chance alone. Therefore, the sample should be large, between 1000 and 2000 subjects rather than a few hundred. However, even with a large sample, a few foods that are associated with the relevant items may occasionally make a moderate contribution to the variation, but not make sense in terms of containing the nutrient being predicted. Thus, nutritional experts rather than statisticians should make the final decision for inclusion of foods when developing FFQs.

Acknowledgements

We thank the S Daniel Abraham International Center for Health and Nutrition for its support of this project.

References

- 1 Willett W, ed. Food frequency methods. In: *Nutritional Epidemiology*. New York: Oxford University Press, 1998; 69–91.
- 2 Block G. A review of validation of dietary assessment methods. *American Journal of Epidemiology* 1982; **115**: 492–505.
- 3 Ocke MC, Bueno-De-Mesquita HB. The Dutch EPIC FFQ. I. Description of the questionnaire, and relative validity and reproducibility for food groups. *American Journal of Epidemiology* 1997; **26**: S37–48.
- 4 Stefanie BT, Ina H, Heiner B. Reproducibility and relative validity of food group intake in a FFQ developed for the German part of the EPIC project. *American Journal of Epidemiology* 1997; **26**: S59–70.
- 5 Martin-Moreno JM, Boyle P, Gorgojo L, Maisonneuve P, Fernandez-Rodriguez JC, Salvini S, *et al.* Development and validation of a FFQ in Spain. *International Journal of Epidemiology* 1993; **22**(3): 512–9.
- 6 Block G, Hartman AM, Naughton D. A reduced dietary questionnaire: development and validation. *Epidemiology* 1990; **1**: 58–64.
- 7 Overvad K, Tjonneland A, Haraldsdottir J, Ewertz M, Jensen OM. Development of a semiquantitative food frequency questionnaire to assess food, energy and nutrient intake in Denmark. *International Journal of Epidemiology* 1991; **20**(4): 900–5.
- 8 Stryker WS, Salvini S, Stampfer MJ, Sampson L, Colditz GA,

- Willett WC. Contributions of specific foods to absolute intake and between-person variation of nutrient consumption. *Journal of the American Dietetic Association* 1991; **91**: 172–8.
- 9 Byers T, Marshall J, Fiedler R, Zielezny M, Graham S. Assessing nutrient intake with an abbreviated dietary interview. *American Journal of Epidemiology* 1985; **122**: 41–50.
 - 10 Shahar D, Shai I, Vardi H, Fraser D. Dietary intake and eating patterns of elderly people in Israel: who is at nutritional risk? *European Journal of Clinical Nutrition* 2003; **57**(1): 18–25.
 - 11 Bilenko N, Shahar D, Shai I, Weitzman S, Fraser D. Prevalence and characteristics of myocardial infarction, diabetes and hypertension in the adult Jewish population: results from the Negev Nutritional Study. *Harefuab* 2003; **142**(1): 17–21.
 - 12 Shai I, Vardi H, Shahar RD, Azrad BA, Fraser D. Adaptation of international nutrition databases and data-entry system tools to a specific population. *Public Health Nutrition* 2003; **6**(3): 401–6.
 - 13 *Food Intake Analysis System, Version 3*. University of Texas Health Science Center at Houston School of Public Health. US Department of Agriculture, Human Nutrition Information Service, 1996.
 - 14 Shahar DR, Shai I, Vardi H, Brener-Azrad A, Fraser D. Development of a semi quantitative food frequency questionnaire (FFQ) to assess dietary intake of multiethnic populations. *European Journal of Epidemiology* 2003; **18**(9): 855–61.
 - 15 Willett W, Sampson L. Reproducibility and validity of a semiquantitative food frequency questionnaire. *American Journal of Epidemiology* 1985; **122**: 51–65.