International Journal of Epidemiology
Vol. 26, No. 1(Suppl. 1)

Printed in Great Britain

Reproducibility and Relative Validity of Food Group Intake in a Food Frequency Questionnaire Developed for the German Part of the EPIC Project

STEFANIE BOHLSCHEID-THOMAS,* INA HOTING,*,** HEINER BOEING† AND JÜRGEN WAHRENDORF*

Bohlscheid-Thomas S (Division of Epidemiology 0345, German Cancer Research Centre, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany), Hoting I, Boeing H and Wahrendorf J. Reproducibility and relative validity of food group intake in a food frequency questionnaire, developed for the German part of the EPIC project. *International Journal of Epidemiology* 1997; **26 (Suppl. 1):** S59–S70.

Background. For the German part of the European Prospective Investigation into Cancer and Nutrition (EPIC), a self-administered, optically-readable food frequency questionnaire (FFQ), including 158 food items and 87 coloured portion size photographs, was developed to assess the usual food and nutrient intake of individuals during the past year. In 1991/1992, the reproducibility and validity of the questionnaire measurements were studied according to the EPIC protocol. This article reports the results on reproducibility and relative validity of measurement of food group intake.

Methods. A total of 104 men and women aged 35–64 years, who are members of the local health insurance institution, AOK Heidelberg, participated in this study. Reproducibility of the questionnaire measurements was obtained by a repeated administration of the FFQ to the same study subjects at a 6-month interval. The mean of 12 24-hour dietary recalls applied at monthly intervals served as reference method for the estimation of the relative validity of questionnaire measurements. A second version of the FFQ that integrated questions on general food consumption patterns was also investigated.

Results. Spearman test-retest correlations of food group intake ranged from 0.49 for bread to 0.89 for alcoholic beverages (median = 0.70). Spearman correlations between food group intake values derived from the 24-hour diet recalls and the FFQ completed in the summer of 1992 varied from 0.14 for legumes to 0.90 for alcoholic beverages (median = 0.45). Correction for attenuation due to within-person error in the reference method as well as the correction for general consumption patterns improved the correlations.

Conclusion. The results indicate that our newly developed FFQ gives reproducible estimates of food group intake. Large day-to-day variation in food group intake complicated the evaluation of FFQ validity. Overall, moderate levels of relative validity were observed for estimates of food group intake.

Keywords: food frequency questionnaire, diet recalls, validity, reproducibility, food intake measurements, EPIC

A self-administered dietary questionnaire was developed and tested in order to estimate dietary food habits and nutrient intakes in the German component of a large cohort study, the European Prospective Investigation into Cancer and Nutrition (EPIC).[1] The questionnaire, however, was designed with a view to its possible use in other epidemiological studies. Reproducibility and validity of dietary intake data are the major problems in nutritional epidemiology.[2] Validity and reproducibility studies were therefore carried out in all seven European countries of the EPIC project following a standardized study protocol.[3] The purpose of the EPIC study is to assess whether dietary exposure differences between subjects are related to cancer risk. The validity of the dietary questionnaire is related to its ability to discriminate between subjects with true exposure differences. On the other hand, reproducibility measures the consistency of questionnaire (exposure) measurements made at different points in time with the same subjects. Associations between diet and cancer can be observed for individual foods,[4] food groups,[5] and nutrients

* Division of Epidemiology 0345, German Cancer Research Centre, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany.
** Institute for Medical Information Processing, Biometry and Epidemiology, University of Munich, Germany.
† Unit of Medical Epidemiology, German Institute of Human Nutrition Potsdam-Rehbrücke, Germany.

TABLE 1 *Sex and age distribution of the participants included in the analysis of the Heidelberg dietary validation study*

Age group	35–44 years	45–54 years	55–64 years	Total
Sex				
Male	15	21	13	49
Female	24	19	12	55
Total	39	40	25	104

TABLE 2 *Study design of the Heidelberg dietary validation study*

Month	FFQ[a]	QGC[b]	24HDR[c]	QLF[d]	Blood	24h-urine
October 1991			X			
November 1991			X			
December 1991			X			X
January 1992			X		X	
February 1992	X		X			
March 1992			X			X
April 1992			X			
May 1992			X			
June 1992			X		X	X
July 1992			X			
August 1992	X		X	X		
September 1992			X			X
January 1993		X				

[a] Food frequency questionnaire.
[b] Questionnaire on general consumption pattern.
[c] Dietary recalls.
[d] Questionnaire of lifestyle factors.

(macro- and micronutrient levels).[6] Consequently, studies on the validity and reproducibility of a questionnaire are important at all these levels. An analysis at food group level of the FFQ data was performed to identify problems in collecting information on food intake, and to make changes (e.g. in the FFQ formulation) to improve validity. Moreover, food consumption is the basic information from which nutrient intake will be calculated.

This article reports results of the German reproducibility and validity study. Specifically, the reproducibility of food group intakes measured by two food frequency questionnaires (FFQ) administered 6 months apart, as well as the relative validity, by using the mean of 12 24-hour dietary recalls as reference.

SUBJECTS AND METHODS

Subjects

A random sample of 528 potential EPIC cohort members from the local health insurance institution AOK Heidelberg (Allgemeine Ortskrankenkasse), aged 35–64, were invited by mail to participate in the validation study. In total 115 (22%) agreed to participate, 195 (37%) refused and 218 (41%) did not respond. An additional eight people were motivated to participate in this study by other participants. Since the eight volunteers met all criteria of recruitment they were included in this study.

Both sexes were represented almost equally in the study population (61 men, 62 women). The majority of the participants were employed in manual occupations (60%), whereas only a minority (30%) had secondary education. Nineteen of the 123 participants did not complete the study. The calculations are therefore based on 104 individuals (49 men, 55 women). Table 1 presents the sex and age distribution of the final study participants.

Study Design

The data collection was carried out from October 1991 to October 1992. During this period, the subjects completed: 1) a 24-hour dietary recall once a month; 2) two dietary FFQ with a 6-month interval (February 1992 and August 1992); 3) a self-administered questionnaire on lifestyle factors (QLF) that included questions on education, occupation, history of tobacco smoking and physical activity; and 4) a short questionnaire on general food consumption patterns (QGC). Two blood samples and four 24-hour urines were collected additionally (for a detailed description see Boeing *et al.*[7]). Table 2 describes the design of this study.

Food Frequency Questionnaire (FFQ)

The dietary questionnaire was designed to measure individuals' usual food and nutrient intakes over the past year. Moreover, for use in the EPIC project an instrument was required which facilitated data collection and data processing. Therefore, we developed a self-administered and precoded questionnaire suitable for optical reading.

The food list was defined on the basis of the dietary intake data obtained from the German National Nutrition Survey (NVS)[8] in which 7-day records were collected from approximately 22 000 subjects. Foods and dishes recorded by a subsample of 1000 participants (5000 different food codes) were grouped into 340 single food items and 25 food groups. Analysis using regression techniques identified which of the 340 food items contributed substantially to the consumption of their respective food group. Using this procedure, 158 single foods or mixed dishes were selected.

We prepared coloured pictures of portion sizes for those food items which are usually not consumed in fixed household measures. This situation arises, in particular, for the main dishes at lunch and dinner. Portion size information was obtained from the NVS by using the medians of each tertile (17th, 50th, and 83rd percentile). The distribution of the percentile values often follow the scheme 1/2, 1, 2. Therefore, we decided to present only one picture of the median portion size if this multiplicative scheme was applicable. In all other instances, three portion sizes per food item were photographed and presented in one picture. Altogether, 87 pictures were displayed in the FFQ. The food items were ordered in the FFQ according to meal patterns.

For each food item the subject had to mark if the food or dish was consumed or not during the previous year. For all food items consumed, the subjects had to select their typical portion size, the consumption frequency (1–6 times) and the time period (day, week, month or year), which suited them best. Questions about the type of fat for cooking, the fat content of milk products, and added sugar or milk to hot beverages were also integrated into the FFQ. Seasonal use based on the number of months fresh fruit is available on the market was asked for fresh fruit and tomatoes. For example, a fruit available 6 months a year, which was eaten ‘three times per week’ in season was converted to a frequency of 78 times yearly.

The FFQ was completed at home, and then reviewed with the respondent by an interviewer to check for incomplete information. A computer program calculated the consumption of the single food items, food groups and nutrients. The FFQ and 24-hour diet recalls were linked together by a list of concordance which was established from all 158 food items in the FFQ and the individual food codes from the 24-hour diet recalls.

Questionnaire on General Food Consumption Patterns (QGC)

When subjects were asked to report the frequencies of their intakes of a large number of single food items within a food group (e.g. meat), the sum often represents an overestimation of the true overall frequency.[9] This is due to the fact that the subjects loose sight of the overall picture. One way of dealing with this problem is to use summary questions for the calibration of the specific frequencies according to the overall frequency.[10]

For this purpose, we selected all food groups with a large number of single food items from the FFQ and asked for their general consumption frequency. Questions on general food consumption (e.g. average number of servings of meat per week) were requested for 14 food groups. These included bread, cereals, fruits, vegetables, legumes, potatoes, cheese, soft drinks, fats, sauces, meat, processed meat, desserts, and soups. Since the validation study was already finished at this time, we collected this information by an additional questionnaire distributed 4 months after the study.

24-Hour Dietary Recalls

The reference method was a series of 12 24-hour diet recalls, which were collected once a month over a period of a year. The 24-hour diet recalls were chosen as the reference method because the response rate and the quality of response were expected to be much better for recalls than weighed records.[3] In addition, multiple 24-hour diet recalls were thought to interfere less with the normal dietary habits of the subjects. The 24-hour diet recalls were carried out through face-to-face interviews by two trained interviewers at the German Cancer Research Centre in Heidelberg. The interviews lasted for an average of 20 minutes. In most instances, each participant was interviewed by the same interviewer for each of the 12 24-hour diet recalls. The interviews were spread across the week from Monday to Friday for each subject. The foods were recalled chronologically for the previous day guided by meal patterns. Coloured photographs, as described above, were used to estimate portion sizes if foods were not consumed in household measures. The interviewers used a checklist to enquire about specific details for particular foods. Meal combinations were used to identify those foods that are often not mentioned such as sauces or fats used in spreads. At the end of each 24-hour diet recall, the foods were summarized in chronological order for the respondent. The 24-hour diet recalls included the intake of supplements, drugs and other medications. It was also recorded whether the recalled day represented a normal day or whether the diet was influenced by specific circumstances, such as sickness or a special holiday.

The 24-hour diet recalls were coded by several nutritionists using the Federal Food Code.[11] They were then reviewed by one person to minimize variability in coding practices. Household measures or standard units were converted to grams using a common list of portion sizes.[12] For the analysis of food groups, foods from the 24-hour diet recalls were grouped as shown in the FFQ.

DATA ANALYSIS

Treatment of Missing Values

Despite a visual check of each FFQ at the time it was returned, and completion of any missing information, some inconsistencies still occurred. On average, we found eight mistakes (5% of 158 food items) per FFQ. For the food-based analysis we thought that it would be easier to compare results between food items, if the number of individuals per item is always the same. Therefore, missing data were replaced by values derived from different assumptions that were tested by comparison with the recall data to find those which best suited the data. We decided to proceed in the following manner: if a food item was marked as consumed, but information on portion size or consumption frequency or both were missing, the mean daily intake of other participants was used. If a food item was marked as not consumed, but portion size and/or consumption frequency were given, the intake of this food item was assumed to be zero. Food items with no marks were considered as not being consumed.

Analysis of Reproducibility of the FFQ

The distribution of food group intake was not normally distributed even after logarithmic transformation. Non-parametric methods were therefore used in the analysis. Different statistical methods were used to evaluate reproducibility. It is most unlikely that repeated dietary questionnaire measurements will agree exactly giving the identical result for each person. The lack of precision can be quantified by estimating the mean difference and the standard deviation of differences.[13] Since the differences were not normally distributed, we calculated the median difference and the Mean Absolute Deviation from the median differences[14] (MAD) to investigate the degree of precision. The formulae are given in the Appendix. We tested the null hypothesis, that the median difference in food intake is zero using the Wilcoxon signed rank sum test.[15] The MAD is a measure of scale of the individual differences, and must be interpreted in relation to the mean intake of the two measurements. Spearman rank order correlation coefficients were calculated between both FFQ administrations to test precision in ranking subjects. Low reproducibility can be caused by changes in dietary habits or by the method itself. Changes in food habits were assessed by analysis of trend for each individual.[14] We tested the hypothesis that there is no trend in food intake across the 12 24-hour diet recall. A trend test can be described as a k/sample test of the null hypothesis of identical distribution against an alternative of monotone order, i.e. if sample i has distribution function F_i, $i = 1, \ldots, k$, then the null hypothesis $H_0: F_1 = F_2 = \ldots = F_k$ is tested against the alternative $H_1: F_1 \geqslant F_2 \geqslant \ldots \geqslant F_k$ (or the reverse), where at least one of the inequalities is strict. We used a non-parametric quick sign test with good efficiency developed by Cox and Stuart.[16] The asymptotic relative efficiency (ARE) is 79% compared to the ARE of Spearman's rank correlation test and 78% compared to the best parametric test. All statistical computations were done using SAS.[17]

Analysis of Relative Validity of the FFQ

The relative validity of the FFQ was assessed by comparing the 24-hour diet recalls and the second FFQ (FFQ2), since this questionnaire covered the same time period as the 24-hour diet recall. The validity of the FFQ2 relative to the reference method (24-hour diet recalls) was assessed by the median difference, the MAD, the Wilcoxon rank sum test, and Spearman rank order correlation coefficients. The correlation coefficients were also corrected for attenuation due to random error in the reference measurements as described by Liu *et al*.[18]

To describe the distribution of food group consumption in our study population, the median for each quintile of 24-hour diet recalls intakes was calculated. In order to evaluate the effect of measurement error on ranking subjects, we estimated the median intake from the 24-hour diet recalls for the subjects in each quintile defined by FFQ2 intakes. In addition, the degree of misclassification was estimated by examining the proportion of subjects classified by the reference method that fell into the same, into the adjacent, or even into the extreme quintile when classified by the FFQ2. Misclassification into the extreme quintile comprises both misclassification from the first to the fifth quintile, and vice versa; from the fifth to the first quintile.

Correction for Consumption Frequency

The questionnaire on general consumption patterns was completed 4 months after the study. This information was used to adjust for possible over- or underestimation of the single consumption frequencies reported in the FFQ2. This adjustment was done only for the FFQ2, because the questionnaire on general consumption patterns referred to the same time span as the FFQ2. The consumption frequencies of single food items in a food group were added up and the sum compared to the frequency given in the general questionnaire. If the sum of the individual items differed from the general consumption frequency, a linear adjustment was made for the single frequencies given in the FFQ2. Weight factors were calculated as the general consumption frequency divided by the sum of the single frequencies. The adjustment was done by multiplying the single frequencies by their weights.

TABLE 3 *Comparison of measurements of median daily intake of 24 food groups estimated by 12 dietary recalls (24HDR) and by dietary questionnaires (FFQ1 and FFQ2). Data were provided by 104 individuals*

Food group	Median intake, g				Median difference			MAD[e]		
	FFQ1[a]	FFQ2[b]	FFQ2corr[c]	24HDR[d]	FFQ1 versus FFQ2	24HDR versus FFQ2	24HDR versus FFQ2corr	FFQ1 versus FFQ2	24HDR versus FFQ2	24HDR versus FFQ2corr
Bread	137.7	138.0	160.6	116.5	4.2	–15.8[f]	–25.4[f]	48.2	39.4	34.3
Cereals	72.6	63.5	53.8	65.6	2.1	3.7	11.3[f]	20.6	28.1	21.6
Salty snacks	15.1	13.8	–	3.5	0.0	–5.3[f]	–	10.7	11.4	–
Cakes, biscuits	45.8	45.0	–	54.3	0.4	3.2	–	22.8	24.2	–
Sweets	34.6	25.4	–	21.9	10.0[f]	–6.3[f]	–	12.6	11.1	–
Spreads	8.5	8.5	–	7.8	0.0	–0.1	–	4.3	5.0	–
Eggs	11.5	10.5	–	6.8	0.0	–4.2[f]	–	5.6	6.7	–
Fruits	247.3	202.9	110.0	125.9	28.1[f]	–50.6[f]	29.6[f]	133.3	97.8	46.5
Vegetables	214.8	172.4	115.8	129.3	32.3[f]	–45.8[f]	12.5[f]	62.1	54.8	31.9
Legumes	27.3	24.0	21.7	10.4	0.8	–13.0[f]	–8.2[f]	12.0	18.0	15.7
Potatoes	89.7	81.3	73.2	67.3	5.2	–19.2[f]	–2.0	25.7	30.9	22.4
Nuts, seeds	1.2	0.8	–	0.2	0.0	0.0	–	3.4	2.6	–
Soft drinks	675.3	722.7	670.5	597.9	43.4[f]	–60.9[f]	–49.0[f]	285.6	298.3	282.3
Milk, milk products	161.8	152.7	–	88.1	9.3	–54.0[f]	–	73.1	64.2	–
Cheese	28.6	28.7	19.2	25.7	2.5	–4.3	7.3[f]	14.2	9.8	10.0
Coffee, tea	605.3	614.4	–	655.4	18.7[f]	32.6	–	199.4	142.7	–
Alcoholic beverages	158.1	163.4	–	167.9	0.0	0.0	–	116.3	80.2	–
Fats	11.9	8.7	10.1	14.4	1.0	2.3	2.2	8.7	8.1	7.6
Sauces	42.9	41.7	23.9	37.0	4.6[f]	–5.7[f]	12.2[f]	17.1	18.2	11.4
Desserts	17.5	14.1	21.4	16.7	1.5	–0.5	–4.2[f]	15.4	23.2	27.7
Fish	14.9	15.8	–	5.6	0.5	–5.9[f]	–	5.8	10.2	–
Meat	88.1	75.4	55.4	78.1	2.9	–0.4	22.6[f]	31.3	32.5	18.7
Processed meat	56.3	48.1	42.5	59.9	2.8	6.3	11.6[f]	24.6	24.7	22.0
Soups	44.6	37.0	57.3	41.3	1.5	2.8	–14.9[f]	26.0	33.0	28.7

[a] First food frequency questionnaire.
[b] Second food freqency questionnaire.
[c] Second food frequency questionnaire corrected for consumption frequency (correction was done for 14 food groups).
[d] Dietary recalls.
[e] Mean absolute deviation from median of individual differences.
[f] Median difference significantly different from zero ($P < 0.05$).

The assumption that the general consumption frequency is more valid than the sum of the single frequencies needed to be validated.[9] For the evaluation of the effect of frequency adjustment, we did the same analyses for the corrected (FFQ2corr) as described above.

RESULTS

Reproducibility of Questionnaire Measurements

Table 3 shows the median intake, the median difference, and the MAD calculated from both FFQ. Comparing the median intake of the FFQ1 relative to the FFQ2, half of the 24 evaluated food groups showed differences within a range of ± 10%, indicating a high consistency in population estimates. For the remaining 12 food groups, the FFQ1 gave 12–33% higher values than the FFQ2. This result was true, in particular, for cereals, sweets, fruits, vegetables, legumes, nuts and seeds, fats, desserts, meat, processed meat, and soups.

The median difference and the MAD were used to describe how well both FFQ agreed at the individual level. The median difference is an estimate of the average random error of FFQ1 relative to FFQ2, while the MAD gave information about the average distance of the individual differences from the median difference.

A positive median difference that was significantly different from zero was observed for sweets, fruits, vegetables, soft drinks, coffee and tea, and sauces. For those food groups, the FFQ2 tended to give lower estimates than the FFQ1. For example, the fruit intake of an individual was, on average, 28.1 g lower in FFQ2 than estimated by FFQ1. For the remaining 16 food

TABLE 4 *Spearman rank order correlation coefficients of food group intake between the first and the second dietary questionnaire and between both dietary assessment methods for 104 individuals*

Food group	Correlation coefficients			Variance ratio[e]	Correlation coefficients (corr)[f]	
	FFQ1[a] versus FFQ2[b]	24HDR[c] versus FFQ2	24HDR versus FFQ2deatt[d]	24HDR	24HDR versus FFQ2	24HDR versus FFW2deatt
Bread	0.49	0.51	0.54	1.3	0.73	0.77
Cereals	0.73	0.42	0.55	8.4	0.54	0.70
Salty snacks	0.72	0.32	0.57	25.5	–	–
Cakes, biscuits	0.69	0.56	0.74	9.2	–	–
Sweets	0.71	0.53	0.58	2.5	–	–
Spreads	0.75	0.63	0.64	0.6	–	–
Eggs	0.73	0.41	0.61	14.0	–	–
Fruits	0.61	0.50	0.54	2.0	0.47	0.51
Vegetables	0.54	0.34	0.42	6.2	0.39	0.48
Legumes	0.70	0.14	0.35	65.1	0.20	0.51
Potatoes	0.71	0.37	0.49	8.8	0.47	0.62
Nuts, seeds	0.66	0.18	0.25	11.7	–	–
Soft drinks	0.65	0.67	0.70	1.0	0.68	0.71
Milk, milk products	0.55	0.56	0.58	0.8	–	–
Cheese	0.61	0.47	0.58	6.6	0.49	0.61
Coffee, tea	0.71	0.70	0.72	0.8	–	–
Alcoholic beverages	0.89	0.90	0.94	0.9	–	–
Fats	0.57	0.43	0.46	1.8	0.45	0.48
Sauces	0.57	0.32	0.45	11.8	0.39	0.55
Desserts	0.71	0.33	0.46	11.4	0.38	0.53
Fish	0.77	0.21	0.37	25.7	–	–
Meat	0.77	0.53	0.67	7.3	0.51	0.65
Processed meat	0.73	0.57	0.65	3.7	0.61	0.70
Soups	0.60	0.27	0.39	13.6	0.42	0.61

[a] First food frequency questionnaire.
[b] Second food frequency questionnaire.
[c] Dietary recalls.
[d] Correlation coefficients deattenuated (corrected for intra-individual variation in dietary recall data).
[e] Variance ratio (intra-individual variation/inter-individual variation) computed from 12 recalling days.
[f] Second food frequency questionnaire corrected for consumption frequency (correction was done for 14 food groups).

groups there were no significant median differences observed.

The MAD was large with regard to the average food intake of the two measurements for most groups, indicating a high variability among individual differences. For example, the mean intake of the two questionnaire measurements for salty snacks was 22.9 g (data not shown), while the MAD was 10.7 g. The absolute value of the MAD is small compared to the other food groups, but with respect to the food intake, a MAD of 10.7 g represents a high variability. Especially high variability with respect to average food intake was also found for fruits, legumes, nuts and seeds, alcoholic beverages, fats, desserts, and soups.

Spearman correlations between both FFQ are presented in Table 4. A low correlation coefficient (<0.5) was only obtained for bread. Eleven food groups showed moderate correlations (0.5–0.7). Food groups in that range were cakes and biscuits, fruits, vegetables, nuts and seeds, soft drinks, milk products, cheese, fats, sauces and soups. The remaining 12 food groups gave correlations >0.7. There were no big gender differences in reproducibility correlations (data not shown), except for soft drinks, where the correlation was much higher in women (0.80) than in men (0.50).

To test the assumption of stable dietary habits, a trend analysis of the 24-hour diet recalls was done. Only a few of the subjects (less than 5% per food group) showed changes in dietary habits over the 12 24-hour diet recalls (data not shown). Therefore, change in dietary habits could not be the main cause of low or moderate reproducibility.

Relative Validity of Questionnaire Measurements

For most food groups the 24-hour diet recalls and FFQ2 agreement in estimated median daily intakes was moderate (Table 3). There was an overall tendency for the questionnaire measurements to give higher medians than the reference method. Only six food groups gave differences within ± 10%; nine food groups showed values which were 11–30% higher, while nine food groups gave an overestimation of more than 30%. The maximum observed was for nuts and seeds with a value four times higher than the reference value.

The median difference and the MAD between both dietary methods gave information about the presence of subject-specific reporting bias (Table 3). A median difference significantly different from zero was obtained for 12 food groups, in particular, for bread, salty snacks, sweets, eggs, fruits, vegetables, legumes, potatoes, soft drinks, milk products, sauces, and fish. All median differences were negative, indicating that the FFQ2 gives higher individual intake estimates compared to the reference measurements. For nearly all food groups the MAD were large with respect to the average food intake. Especially high values were found for salty snacks, spreads, eggs, fruits, legumes, nuts and seeds, milk products, desserts, fish, and soups.

Spearman correlations between FFQ2 and 24-hour diet recalls are presented in Table 4. Nine food groups yielded correlations <0.4, 11 groups showed values between 0.4 and 0.6, and for four groups values >0.6 were obtained. Food groups with correlations <0.4 were salty snacks, vegetables, legumes, potatoes, nuts and seeds, sauces, desserts, fish, and soups. In general, slightly higher correlations were found for men than for women, but meaningful gender differences were not observed (data not shown).

The ratio of within- to between-subject variation measured in the 12 24-hour diet recalls was larger than 1.0 for nearly all food groups, and ranged from 0.6 for spreads to 65.1 for legumes (Table 4). The highest values were found for salty snacks, eggs, legumes, nuts and seeds, sauces, desserts, fish, and soups. The high variance ratio was due to the large day-to-day variability in food consumption. This lack of precision in the reference measurements can be taken into account by computing deattenuated correlation coefficients (Table 4). This correction improved the correlations between methods for all food groups. This was also true for deattenuated correlations stratified by sex. After correction, low correlations were only found for four groups, namely for legumes, nuts and seeds, fish, and soups. Twelve groups had values ranging from 0.4 to 0.6, and eight groups gave estimates >0.6. Gender differences were found for sweets, eggs, sauces, fish, and soups, i.e. for sweets we observed a correlation of 0.47 for females instead of 0.74 for males. The deattenuated correlations were mostly better for men than for women, except for sauces and fish.

The median for each recall quintile was calculated to describe the distribution of food group intake in our study population (Table 5). The effect of measurement error on the food intake distribution was evaluated by the median intake from the 24-hour diet recalls by quintiles of the FFQ2 (Table 5). The observed difference between the lowest and the highest quintile was less than for quintiles defined by the recalls themselves. This was due to overestimation of low food consumption and underestimation of high food consumption. For some food groups, we could not find an increase in median intake from the first to the fifth quintile when classified by the FFQ2. An example of this misclassification from the FFQ2 is that the median intakes for soup from the first to the fifth quintile were 25.1 g, 55.0 g, 16.7 g, 25.0 g, and 87.5 g. These values corresponded to the median soup intakes estimated from the 24-hour diet recalls of 0.0 g, 12.9 g, 40.8 g, 76.3 g, and 132.1 g, respectively.

The degree of misclassification associated with categorized intakes assessed by the FFQ2 was examined as the proportion of subjects similarly classified into the same, the adjacent, or opposite quintile (Table 6). The proportion of subjects similarly classified by both instruments ranged from 21.2% for legumes to 59.6% for alcoholic drinks, with most values lying between 30% and 40%. The proportion of subjects classified into the adjacent category was similar, ranging from 25.0% for salty snacks to 47.1% for cereals. Extreme misclassification into the opposite quintile was rare; the highest value was seen for vegetables with 4.8%.

Adjustment for Consumption Frequency

For the FFQ2, particular algorithm for calculating food consumption was developed considering the frequencies given in the questionnaire on general consumption patterns. Table 3 presents the effects of adjustment for consumption frequency on median intake, on the median difference, and on the MAD. Comparison of median intake indicated a smaller systematic bias by the corrected version for six food groups (fruits, vegetables, legumes, potatoes, soft drinks, and fats). An increase in population bias was seen for the remaining eight food groups, namely for bread, cereals, cheese, sauces, desserts, meat, processed meat, and soups.

For all corrected food groups, except for potatoes and fats, we found a significant median difference. In most instances the FFQ2corr gave higher estimates than the

TABLE 5 *Median daily food group intake (g) derived from 12 24-hour dietary recalls for quintiles of intake defined by the diet recalls and by quintiles defined by the second food frequency questionnaire (N = 104 individuals)*

Food group	Median 24HDR[a], 24HDR-cutpoints					Median 24HDR, FFQ2[b]-cutpoints					Median 24HDR, FFQ2corr[c]-cutpoints				
Quintile	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Bread	70.3	95.5	115.9	153.4	209.6	79.2	114.2	112.1	122.1	171.3	74.6	95.4	148.3	148.7	192.1
Cereals	17.6	37.9	64.6	104.1	139.8	43.5	36.9	72.5	118.3	74.6	37.2	60.4	53.3	106.3	109.6
Salty snacks	0.0	0.0	3.3	15.4	41.7	0.0	8.3	4.2	1.7	27.5	–	–	–	–	–
Cakes, biscuits	12.9	39.2	54.0	78.3	129.3	29.0	47.5	48.0	72.1	92.4	–	–	–	–	–
Sweets	1.7	10.7	21.7	39.1	67.5	5.6	15.3	24.9	24.5	49.3	–	–	–	–	–
Spreads	0.8	4.2	7.5	12.3	20.4	2.9	5.0	8.3	10.0	17.5	–	–	–	–	–
Eggs	0.0	1.8	6.7	14.9	27.5	0.0	9.2	5.5	9.2	22.9	–	–	–	–	–
Fruits	26.9	95.8	124.7	190.3	305.5	65.5	137.4	118.1	182.2	245.6	74.5	120.1	156.2	133.7	242.8
Vegetables	62.2	99.9	128.3	160.3	239.8	97.0	120.9	125.3	164.9	140.1	102.0	103.7	140.9	139.7	166.8
Legumes	0.0	1.7	10.4	19.7	64.2	9.2	4.2	13.4	16.3	12.5	7.6	4.2	13.4	14.6	17.5
Potatoes	16.7	47.5	67.1	84.2	144.2	46.4	50.0	67.5	76.7	79.3	41.6	63.3	75.8	67.1	117.1
Nuts, seeds	0.0	0.0	0.0	2.1	8.9	0.0	1.7	0.0	0.8	1.7	–	–	–	–	–
Soft drinks	222.8	444.6	587.5	804.2	1211.7	304.8	538.3	694.0	804.2	940.6	314.2	440.8	681.7	804.2	972.3
Milk, milk products	25.7	56.7	85.9	151.4	295.4	57.1	59.6	80.0	90.3	229.0	–	–	–	–	–
Cheese	8.6	17.5	25.4	36.3	59.6	20.1	19.8	23.0	27.5	48.8	19.7	20.5	26.7	38.3	34.8
Coffee, tea	292.9	493.8	650.0	787.8	1095.8	304.6	660.9	664.2	666.7	1000.9	–	–	–	–	–
Alcoholic beverages	0.0	69.2	167.5	323.5	640.4	0.0	78.3	145.8	310.8	500.0	–	–	–	–	–
Fats	3.2	9.0	14.2	19.2	32.0	8.9	9.6	14.2	16.0	23.4	10.2	9.2	13.4	19.2	19.2
Sauces	15.9	28.1	36.9	47.2	66.1	33.7	32.7	41.3	39.3	42.1	33.3	33.5	34.5	50.5	42.1
Desserts	0.0	4.2	16.7	34.2	79.2	0.0	10.0	20.8	23.8	21.7	2.5	8.3	17.5	16.7	37.5
Fish	0.0	0.0	5.0	16.7	32.0	0.0	8.3	0.0	8.3	16.7	–	–	–	–	–
Meat	25.8	60.8	77.5	99.6	138.4	48.5	70.0	91.4	77.5	127.6	51.0	84.5	76.8	91.4	127.6
Processed meat	12.6	33.8	59.8	77.1	134.3	14.6	46.4	66.7	85.0	85.0	14.6	54.6	66.7	60.0	107.5
Soups	0.0	12.9	40.8	76.3	132.1	25.1	55.0	16.7	25.0	87.5	16.0	27.9	33.3	25.0	119.6

[a] Dietary recalls.
[b] Second food frequency questionnaire.
[c] Second food frequency questionnaire corrected for consumption frequency (correction was done for 14 food groups).

TABLE 6 *Comparison of food frequency questionnaire with mean daily intakes derived from 12 24-hour dietary recalls, based on joint classification by quintiles (N = 104 individuals)*

Food group	FFQ2[a] versus 24HDR[b]			FFQ2corr[c] versus 24HDR		
	same category (%)	adjacent category (%)	extreme category (%)	same category (%)	adjacent category (%)	extreme category (%)
Bread	31.7	32.7	0.0	43.3	38.5	0.0
Cereals	26.9	47.1	2.9	29.8	41.3	0.0
Salty snacks	35.6	25.0	0.0	–	–	–
Cakes, biscuits	30.8	41.3	0.0	–	–	–
Sweets	33.7	42.3	1.9	–	–	–
Spreads	44.2	38.5	0.0	–	–	–
Eggs	32.7	32.7	0.0	–	–	–
Fruits	31.7	35.6	1.0	34.6	34.6	1.0
Vegetables	28.8	35.6	4.8	36.5	27.9	2.9
Legumes	21.2	35.6	3.8	26.0	32.7	2.9
Potatoes	30.8	34.6	3.8	32.7	41.3	1.9
Nuts, seeds	26.9	40.4	1.0	–	–	–
Soft drinks	39.4	40.4	1.0	43.3	34.6	1.0
Milk, milk products	30.8	39.4	0.0	–	–	–
Cheese	26.9	42.3	0.0	24.0	43.3	0.0
Coffee, tea	44.2	37.5	0.0	–	–	–
Alcoholic beverages	59.6	33.7	0.0	–	–	–
Fats	36.5	37.5	1.9	35.6	36.5	1.9
Sauces	31.7	28.8	3.8	30.8	40.4	3.8
Desserts	28.8	41.3	1.0	35.6	33.7	1.9
Fish	23.1	41.3	2.9	–	–	–
Meat	34.6	36.5	2.9	38.5	32.7	1.9
Processed meat	36.5	41.3	1.9	34.6	41.3	1.0
Soups	23.1	33.7	1.0	36.5	26.0	0.0

[a] Second food frequency questionnaire.
[b] Dietary recalls.
[c] Second food frequency questionnaire corrected for consumption frequency (correction was done for 14 food groups).

24-hour diet recalls. Nearly all food groups showed a smaller MAD, indicating an improvement in the individual food intake estimates. Only for cheese and for desserts did we observe larger MAD than before.

The Spearman correlations between FFQ2corr and 24-hour diet recalls are presented in Table 4. The corrected correlations were higher for all food groups, except for fruits and meat, which showed slightly lower values. Correction for attenuation due to random error improved these correlations, as well. For example, the food group soup had a low correlation (0.27) using the original data from both methods. Adjustment for consumption frequency increased this correlation to 0.42, and after the correction for measurement error we observed a value of 0.61.

The evaluation of actual food group intake defined by categories of the FFQ2corr showed only a minor improvement compared to the original questionnaire data (Table 5). The effect of classification into quintiles is summarized in Table 6. For most food groups, the proportion of subjects equally classified was slightly improved compared to the original data, and extreme misclassification was reduced (Table 6).

DISCUSSION

Study Participants

It is obvious that a participation rate of 22% is very low. The main reasons for refusing to participate were ‘no time’ (61%), and ‘no interest’ (23%). This indicates that the participants in the study were highly motivated subjects. Therefore, the results of this study may overestimate the reproducibility and validity of the FFQ.

Reproducibility

The aim of this article was to estimate reproducibility and relative validity of the German FFQ designed for EPIC. According to Spearman rank order correlations, we found high reproducibility ($r > 0.7$) for one-half of the food groups and moderate reproducibility ($r \leqslant 0.7$) for the other half. Correlation coefficients of 0.4–0.8 appear to be typical for reproducibility of food groups and single food items among middle-aged subjects.[19–26] Reduced reproducibility can be partially explained by real changes in dietary habits. However, the trend analysis on 24-hour diet recalls did not indicate major individual changes in food consumption during the study period. Thus, the moderate reproducibility was probably due to random error of the FFQ itself.

Block and Hartman[27] noted that the complexity of a questionnaire is an important factor that will have a great impact on the reproducibility. For example, an instrument without variable portion sizes for each food item exhibits less variability, and is likely to have a higher reproducibility compared to a questionnaire including these aspects. Similarly, a questionnaire limiting consumption frequency responses to a few categories, gives more reproducible results than an instrument with unlimited choice of responses. Our questionnaire requires the selection of the typical serving size and allows a free combination of consumption frequency per indicated time interval. Consequently, our FFQ is relatively more susceptible to random error than a much simpler method.

The reproducibility correlations found in our study were better than one would have expected from the investigation of the MAD which indicated a high proportion of within-subject random error. One possible explanation for this result may be that food groups which were highly reproducible are biased by correlated errors. Subject-specific reporting error will be correlated if the questionnaire measurements were repeated on different occasions.[28,29] Beaton examined the error terms in FFQ at the nutrient level.[30] He found that subject-specific errors gave rise to the high reproducibility correlations of FFQ. Therefore, the test-retest correlations may result in a misleading interpretation of reproducibility.

Relative Validity

Comparison between FFQ2 and 24-hour diet recalls was used to estimate relative validity. We selected multiple dietary recalls as the reference method. If the analysis was restricted only to those days rated by the subjects as normal days (80% of all recalls), only slightly different results with no consistent pattern were obtained. Therefore, in the analysis all recalls were used.

Spearman correlations between FFQ2 and 24-hour diet recalls showed values between 0.14 and 0.90, with most values between 0.4 and 0.6. Validation studies of intake of food groups and single food items assessed by FFQ have observed correlations, generally, between 0.3 and 0.8.[19,21,24,31–33]

Within-person variation in food consumption measured by 12 dietary recalls attenuated the correlation coefficients. The variance ratios computed from the 24-hour diet recalls revealed high day-to-day variation in the reference method for most food groups. Especially high ratios were found for salty snacks, eggs, legumes, nuts and seeds, sauces, desserts, fish, and soups. Our ratios were higher than those calculated by Salvini *et al*.[21] Correction for attenuation improved the correlations for all food groups. Salvini *et al*.[21] demonstrated that, after correction, food items with a higher variance ratio showed a stronger association than food items with a lower ratio. Our results confirmed this finding only for some foods. But for legumes, nuts and seeds, fish, and soups, which showed high variance ratios, the corrected correlations were still <0.4. The estimates presented depend on the assumption that the reference method is unbiased. This assumption may be questionable for the four food groups mentioned above. These food groups were typically consumed on Friday and Saturday, which were not covered by the 24-hour diet recalls. Despite the fact that the dietary intakes on Fridays and Saturdays were not collected, 12 days of diet recall did not adequately represent usual intake for food groups that were consumed less frequently than once per week. For example, a food group consumed twice a month has a 44% probability of not being consumed during the 12 24-hour diet recalls. Food groups mentioned in the FFQ2 that were seldom consumed by most subjects (less than twice a week) were salty snacks, sweets, eggs, legumes, nuts and seeds, desserts, fish, and soups. Looking at the validity results, we found the largest MAD for these food groups, and also the lowest correlation coefficients. For these food groups the reference measurements are probably underestimated due to only 12 days of recall and due to the exclusion of Friday and Saturday. Thus, the reference measurements may imperfectly reflect ranking. Consequently, it is difficult to make any conclusions with respect to the validity for these food groups.

Subject bias can be caused, apart from biased reference measurements, by several factors, e.g. if a food item in the FFQ that is commonly eaten by a subject is omitted, if portion sizes between methods differ, or if there is misinterpretation of food items listed in the FFQ.[27] The actual source of subject bias in our FFQ is not yet known. To identify these sources, further analyses at the single food item level are being undertaken.

Random errors found in our FFQ resulted in ranking errors and misclassification. It is obvious that this FFQ could not discriminate satisfactorily between quintiles for nearly all food groups, and the measurable contrasts were reduced compared to the 24-hour diet recalls data. The classification of subjects into quintiles was moderate. On average, 33% of subjects were correctly classified into the same quintile, and 70% were classified within one quintile.

Dietary questionnaires often overestimate consumption frequency. This may occur when several food items of a single food group are presented in a questionnaire. Flegal and Larkin[34,35] observed that errors in frequency estimation are the most important source of error in ranking individuals by levels of macronutrient intake estimated from an FFQ. Our FFQ design did not allow similar analyses to those made by Flegal and Larkin. However, we applied questions on the general consumption frequency of particular food groups. When the general questions were set as a reference we improved our ranking of subjects. In particular, the correlations between methods increased, and the classification into quintiles was improved slightly. This finding confirmed that subjects were best at estimating relative frequencies as opposed to absolute frequencies, as was observed by Smith.[36] Our results also confirmed the results found by Flegal and Larkin. Correction of reported consumption frequency had only minor effects on the population bias, but improved the ranking of subjects.

In summary, the results indicate that our newly developed FFQ gives reproducible estimates of food group intake. Large day-to-day variation in food group intake and the omission of Fridays and Saturdays complicated the evaluation of FFQ validity. Overall, moderate levels of relative validity were observed.

The results of this study provided information on how to improve the FFQ for the main study. Questions concerning general consumption frequency were integrated in the final FFQ version. Furthermore, we modified the food list, and for some food items we made new photographs of portion sizes. The consumption frequency will also now be requested by nine response categories instead of the combination of frequency and time unit.

ACKNOWLEDGEMENTS

The authors would like to thank Bruno Krüger from the local health institution AOK Heidelberg for allowing us to recruit AOK members, Renate Birr for doing the interviews, all nutritionists who did the food coding, and especially Dr Gabriele Perez-Nelius who controlled the food coding. We are grateful to Günther Laib and Elke Bauer for their help in data management and writing the computer programs.

This work was part of the EPIC project (European Prospective Investigation into Cancer and Nutrition), supported by the Europe Against Cancer Program of the CEC (grant No 91 CVV01270-0).

REFERENCES

1 Riboli E. Nutrition and cancer: Background and rationale of the European Prospective Investigation into Cancer and Nutrition (EPIC). *Ann Onc* 1992; **3:** 783–91.

2 Van Staveren WA, Burema J, Deurenberg P, Katan M B. Weak associations in nutritional epidemiology: The importance of replication of observations on individuals. *Int J Epidemiol* 1988; **17:** 964–69.

3 Riboli E. *Prospective Studies on Diet and Cancer. Report of the Pilot Study, Phase I, November 1988–October 1989*. IARC, Lyon, October 1989, pp. 42–44.

4 Steinmetz K A, Potter J D. Vegetables, fruit, and cancer. I. Epidemiology. *Cancer Causes Control* 1991; **2:** 325–57.

5 Howe G R, Friedenreich C M, Jain M, Miller A B. A cohort study of fat intake and risk of breast cancer. *J Natl Cancer Inst* 1991; **83:** 336–40.

6 Van den Brandt P A, Goldbohm R A, Van't Veer P, Volvics A, Hermus R J J, Strumanns F. A large-scale prospective cohort study on diet and cancer in the Netherlands. *J Clin Epidemiol* 1990; **43:** 285–95.

7 Boeing H, Bohlscheid-Thomas S, Voss S, Schneeweiss S, Wahrendorf J. The relative validity of vitamin intakes derived from food frequency questionnaire compared to 24-hour recalls and biological measurements: results from the EPIC pilot study in Germany. *Int J Epidemiol* 1997; **26 (Suppl. 1):** S82–S90.

8 *Materialien zur Gesundheitsforschung, Schriftenreihe zum Programm der Bundesregierung Forschung und Entwicklung im Dienste der Gesundheit.* Die Nationale Verzehrsstudie, Ergebnisse einer Basisauswertung 1991, Band 18, 3. Auflage.

9 Haraldsdóttir J. Minimizing error in the field: quality control in dietary surveys. *Eur J Clin Nutr* 1993; **47 (Suppl. S2):** S19–S24.

10 Tjønneland A, Overvad K, Haraldsdóttir J, Bang S, Ewertz M, Møller Jensen O. Validation of a semiquantitative food frequency questionnaire developed in Denmark. *Int J Epidemiol* 1991; **4:** 906–12.

11 Polensky W: *Zusammenarbeit mit dem Bundesgesundheitsamt*. Institut für Sozialmedizin und Epidemiologie. Die Bundeslebensmittelschlüsseldokumentation. 1989, Version II.1.

12 *MONICA Mengenliste, Auswertungs- und Informationsdienst für Ernährung*. Landwirtschaft und Forsten (AID) e.V. (ed.). 1991.

13 Bland J M, Altman D G. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; **i:** 307–310.

14 Hartung J, Elpelt B, Klösener K H (eds). *Statistik: Lehr- u. Handbuch d. Angewandten Statistik*. München, Wien: Oldenburg, 1982.

15 Altman D G. *Practical Statistics for Medical Research*. London: Chapman & Hall, 1991.

16 Cox D R, and Stuart A. Some quick sign tests for trend in location and dispersion. Biometrica 1955; **42:** 80–95.

17 SAS Institute Inc. *SAS User's Guide: Statistics, Version 5 Edition*. Cary, NC: SAS Institute Inc., 1985.

[18] Liu K, Stamler J, Dyer A, McKeever J, McKeever P. Statistical methods to assess and minimize the role of intra-individual variability in obscuring the relationship between dietary lipids and serum cholesterol. *J Chron Dis* 1978; **31:** 399–418.
[19] Willett W C. *Nutritional Epidemiology*. New York: Oxford University Press, 1990.
[20] Colditz G A, Willett W C, Stampfer M J, Hulshof K F A M. The influence of age, relative weight, smoking, and alcohol intake on the reproducibility of a dietary questionnaire. *Int J Epidemiol* 1987; **16:** 392–98.
[21] Salvini S, Hunter D J, Sampson L *et al*. Food-based validation of a dietary questionnaire: The effects of week-to-week variation in food consumption. *Int J Epidemiol* 1989; **18:** 858–67.
[22] Bueno de Mesquita H B, Smeets F W M, Runia S, Hulshof K F A M. The reproducibility of a food frequency questionnaire among controls participating in a case-control study on cancer. *Nutr Cancer* 1992; **18:** 143–56.
[23] Järvinen R, Seppänen R, Knekt P. Short-term and long-term reproducibility of dietary history interview data. *Int J Epidemiol* 1993; **22:** 520–27.
[24] Feskanich D, Rimm E B, Giovannucci E L *et al*. Reproducibility and validity of food intake measurements from a semi-quantitative food frequency questionnaire. *J Am Diet Assoc* 1993; **93:** 790–96.
[25] Ajani U A, Willett W C, Seddom J M. Reproducibility of a food frequency questionnaire for use in ocular research. *Invest Ophthalmol Vis Sci* 1994; **35:** 2725–33.
[26] Franceschi S, Negri E, Salvini S *et al*. Reproducibility of an Italian food frequency questionnaire for cancer studies: results for specific food items. *Eur J Cancer* 1994; **29A:** 2298–305.
[27] Block G, Hartman A M. Issues in reproducibility and validity of dietary studies. *Am J Clin Nutr* 1989; **50:** 1133–38.
[28] Willett W. An overview of issues related to the correction of non-differential exposure measurement error in epidemiological studies. *Stat Med* 1989; **8:** 1031–40.
[29] Beaton G H. Approaches to analysis of dietary data: relationship between planned analyses and choice of methodology. *Am J Clin Nutr* 1994; **59 (Suppl.):** 253S–61S.
[30] Beaton G H. Interpretation of results from diet history studies. In: *The Dietary History Method*. Proceedings of the 2nd Berlin meeting on Nutritional Epidemiology. L Kohlmeier (ed.). London: Smith-Gordon, 1991, pp. 15–37.
[31] Jain M, Howe G R, Johnson K C, Miller A B. Evaluation of a diet history questionnaire for epidemiologic studies. *Am J Epidemiol* 1980; **111:** 212–19.
[32] Pietinen P, Hartman A M, Haapa E *et al*. Reproducibility and validity of dietary assessment instruments. I. A self-administered food use questionnaire with a portion size picture booklet. *Am J Epidemiol* 1988; **128:** 655–66.
[33] Goldbohm R A, Van den Brandt P A, Brants H A M *et al*. Validation of a dietary questionnaire used in a large-scale prospective cohort study on diet and cancer. *Eur J Clin Nutr* 1994; **48:** 253–65.
[34] Flegal K M, Larkin F A, Metzner H L *et al*. Counting calories: Partitioning energy intake estimates from a food frequency questionnaire. *Am J Epidemiol* 1988; **128:** 749–60.
[35] Flegal K M, Larkin F A. Partitioning macronutrient intake estimates from a food frequency questionnaire. *Am J Epidemiol* 1990; **131:** 1046–58.
[36] Smith A F. Cognitive processes in long-term dietary recall. National Center for Health Statistics. *Vital Health Stat* 1991; **6:** 1–34.

APPENDIX

An example is given for the comparison of both FFQ. The median difference and the mean absolute deviation from the median difference (MAD) were calculated by use of the following formulas:[14]

$$d_i = |FFQ1 - FFQ2|$$

$$\tilde{d}_{0.5} = \tfrac{1}{2}\,(d_{i(n/2)} + d_{i((n+2)/2)})$$

$$MAD = \tfrac{1}{n}\sum_{i=1}^{n} |d_i - \tilde{d}_{0.5}|$$

where d_i = the absolute difference between both questionnaire administrations for each individual,
$\tilde{d}_{0.5}$ = the median difference,
n = the number of subjects (n = 104),
MAD = the mean absolute deviation from the median difference.