Validation of a dietary questionnaire used in a large-scale prospective cohort study on diet and cancer

European Journal of Clinical Nutrition (1994) **48**, 253–265
Received 25 August 1993; accepted 19 October 1993

Validation of a dietary questionnaire used in a large-scale prospective cohort study on diet and cancer

R. A. Goldbohm[1,2], P. A. van den Brandt[2], H. A. M. Brants[1], P. van't Veer[1], M. Al[1], F. Sturmans[2] and R. J. J. Hermus[1]

[1]*TNO Toxicology and Nutrition Institute, Zeist;* [2]*University of Limburg, Department of Epidemiology, Maastricht, The Netherlands*

Objective: The study was conducted to assess the validity of a self-administered 150-item food frequency questionnaire (FFQ), used in a cohort study on diet and cancer (120 852 men and women, aged 55–69).
Design & subjects: The study was carried out in a subgroup of the cohort (59 men and 50 women) 2 years after the baseline FFQ was completed. A dietary record, kept over three 3-day periods, 4–5 months apart, served as reference method. To evaluate the representativeness of the study population for the entire cohort, a comparison was made with the baseline questionnaire of a random sample of the cohort.
Results: Pearson correlation coefficients between nutrient intakes assessed by the record and the FFQ that was completed afterwards ranged from 0.40 (95% CI: 0.22–0.54) for vitamin B_1 to 0.86 (95% CI: 0.80–0.90) for alcohol intake, with correlations for most nutrients between 0.6 and 0.8. Adjustment for energy intake and sex did not materially affect these correlations, except the correlation for fat intake, which changed from 0.72 to 0.52. Correlation coefficients were only slightly modified when the results were extrapolated to the cohort at large. Correction of correlation coefficients for attenuation by day-to-day variance in the record data improved them by 0.07 on average.
Conclusions: It is concluded that the FFQ is able to rank subjects according to intake of food groups and nutrients. Despite a better performance of validation study participants, this conclusion also applies to the cohort at large.
Sponsorship: Dutch Cancer Society.
Descriptors: diet records, epidemiologic measurements, food habits, questionnaires

Introduction

A self-administered dietary questionnaire is often the method of choice in a large-scale epidemiological study, such as a prospective cohort study, into dietary habits and disease. The validity of such a questionnaire is not self-evident, since it is limited with respect to the foods included and the degree to which portion sizes are quantified. Moreover, each questionnaire needs to be tuned to the specific dietary habits of the study population. Validation studies of a number of self-administered dietary questionnaires have been published (Jain *et al.*, 1982; Willett *et al.*, 1985, 1987; Pietinen *et al.*, 1988*a*,*b*; Salvini *et al.*, 1989;

Correspondence to: R. A. Goldbohm, TNO Toxicology and Nutrition Institute, P.O. Box 360, 3700 AJ Zeist, The Netherlands.
Present addresses: P. van't Veer, Agricultural University, Department of Epidemiology and Public Health, Wageningen, The Netherlands; M. Al, University of Limburg, Department of Human Biology, Maastricht, The Netherlands.

Block *et al.*, 1990; Tjønneland *et al.*, 1991; Rimm *et al.*, 1992).

We developed a self-administered, mailed food frequency questionnaire (FFQ)* for use in a large-scale prospective cohort study on dietary habits and cancer in The Netherlands (NLCS) (Bausch-Goldbohm *et al.*, 1988). The cohort, consisting of 120 852 men and women aged 55–69, was recruited from the general population and completed the baseline questionnaire in 1986 (Van den Brandt *et al.*, 1990). The questionnaire is repeated each year in random samples of the cohort (n = 400) to assess its reproducibility and the stability of dietary habits over time.

This paper describes the validity of the FFQ as compared to a 9-day diet record. Considering the aetiological purpose of the cohort study, validity of the questionnaire is primarily defined as its ability to rank study subjects according to nutrient intake and food (group) consumption. Since the performance of a questionnaire also depends on the actual study population, the validation study was conducted within the cohort. Assessment of selection bias, potentially introduced by incomplete participation in the validation study, was included in the study design.

Subjects and methods

Study design

The diet record method was used as reference method, since its errors are assumed to be independent of the errors in a food-frequency type questionnaire (Willett, 1990). Dietary intake was recorded over three periods (of three consecutive days each, Fig. 1), representing three seasons in The Netherlands, differing with respect to consumption patterns of (specific) vegetables, fruits and meat (Dutch Board for Vegetables and Fruit, 1986; Dutch Meat Board, 1986). The 9 recording days were balanced across the days of the week for each subject and for the study group as a whole. The diet record was compared to the FFQ that was completed approximately 3 months after the last recording period.

To investigate a possible learning effect of recording of intake, the FFQ application coincided with the repeated questionnaire completed annually in the cohort by participants of the aforementioned reproducibility study (Fig. 1). Furthermore, to assess the representativeness of the validation study group (participants as well as non-participants) with regard to the cohort, the baseline questionnaires of this group were compared to those of the 1988 reproducibility study group, which constituted a random sample of the cohort.

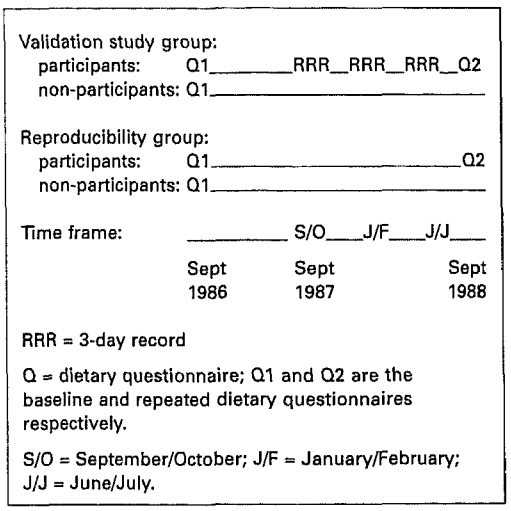

Fig. 1. Design of the validation study of the dietary questionnaire used in The Netherlands Cohort Study.

Unless specified otherwise, all results pertain to the questionnaire completed by the study subjects at the end of the year in which recording took place.

Subjects

Since the participants had to be visited at home during each recording period, recruitment was confined to 12 municipalities, located in an eastern and a western region of The Netherlands. As far as degree of urbanization was concerned, these municipalities were representative of the 204 from which the cohort was recruited.

Of a total of 212 randomly selected cohort members (107 men and 105 women), 109 subjects (59 men and 50 women) completed the validation study (51%); 92 did not participate from the start, and 11 dropped out during the study. Reasons for non-participation and drop-out could be attributed to refusal (two-thirds) and unavailability (death, no contact, absence

* An English translation of the FFQ is available on request.

during one or more recording periods, etc.). Among the non-participants, six subjects were excluded because they did not manage to keep the record or did not eat at home most of the time and were hence not expected to keep a good record.

Dietary questionnaire

The purpose of the FFQ was to assess habitual consumption of approximately 150 foods during the past year. The foods included in the questionnaire were originally selected according to their contribution to the between-person variance of the intake of energy and of the following nutrients: protein (vegetable as well as animal), fat (saturated, monounsaturated and polyunsaturated), cholesterol, carbohydrates (mono- and disaccharides, polysaccharides), dietary fibre (of cereal as well as vegetable and fruit origin), alcohol, calcium, vitamin A, ß-carotene and vitamin C (Bausch-Goldbohm *et al.*, 1988). The contribution to the variance was calculated from a data set previously collected by means of a dietary history method in a population of men and women of the same age category as the cohort.

For each item, the questionnaire asked for frequency of use on a scale of seven frequency categories: never/less than once per month, once per month, 2–3 times per month, once per week, 2–3 times per week, 4–5 times per week, 6–7 times per week. The number of servings per consumption frequency was asked in natural (e.g. apple, slice of bread) or household units (e.g. glass, spoon). For cooked vegetables and meat, the typical individual serving size in grams was asked. For several items, the frequency categories were replaced with the number of serving units taken daily (coffee, tea, bread), weekly (eggs, onions, tomatoes) or monthly (mushrooms, sweet peppers). Questions on vegetables were specified with respect to season (summer and winter). Margarine used on bread and cooking fats and oils were specified as to type and brand in open questions. An open-ended question also asked to list any foods eaten regularly (once a week or more) but not included in the questionnaire.

Questionnaires were double-keyed and automatically coded by the data-entry program. Data were checked for completeness, consistency, range, and other response errors and corrected whenever feasible by means of an SPSS computer program, which had been developed using the data from the first 3000 cohort questionnaires entered and from an earlier small validation study (Den Breeijen, 1990). The resulting program ensures identical cleaning procedures for all questionnaires.

To determine the completeness and the quality of the questionnaires, they were evaluated by means of the number of blank items and by means of an error index, which was calculated as the sum of the scores of 15 variables that indicated each the presence of a specific response error (see footnote to Table 4). Visual inspection of the cohort questionnaires had revealed that some of the subjects had consistently skipped items that they never ate, instead of checking the frequency category 'never/less than once per month'. Questionnaires were considered unacceptably incomplete when either: (a) more than 60 items (out of 150) were left blank *and* fewer than 35 items were eaten at least once a month; or (b) one or more item blocks (i.e. groupings of items in the questionnaire, e.g. beverages) were left blank. According to these criteria 6.0% of the cohort (6.5% and 5.5% for men and women respectively), among whom 1% had mistakenly skipped a page, has to be excluded from aetiological analyses, relating dietary habits to cancer. In addition, 1.0% of the cohort members were excluded because the error index of their questionnaires exceeded 10. This criterion was based on the subjective verdict of inconsistency after visual inspection of the questionnaires.

Diet records

The diet records were collected and coded by nine (student) dietitians (three for each recording period), who were trained and supervised by one experienced dietitian (H.A.M.B.), who also checked the coding of each record. The participants were asked to write down all foods and beverages taken and to specify type and brand. The amount had to be specified in their own household measures (glass, etc.) and/or weight as purchased. We did not use a complete weighed record method since, in our experience with untrained subjects, it is liable to mistakes. Moreover, it has been found that weighing could influence eating habits (Livingstone *et al.*, 1990). Instead, for each item in the diary the individual amount used was assessed in a way that was most suited for that specific food.

Standard serving sizes were only applied for foods that are sold in standard sizes (e.g. slice of bread, sugar cube).

One day before the beginning of the recording period, the participant was instructed at home by the dietitian and received the diary, including written instructions and examples. The day after the last recording day, the same dietitian checked the diary with the subject and, if necessary, with the subject's partner. During the same visit, the dietitian measured the capacity of the household utensils (glasses, cups, etc.) specified in the diary and weighed the amount of butter or margarine used on bread and the amount of sugar used in tea and coffee. For the second and third recording periods, the instruction visit, but not the check visit, was skipped and diaries were mailed to those participants who appeared to have properly understood the record-keeping procedure.

Calculation of intake of nutrients and food groups

Mean individual nutrient intake per day was calculated from the record as the average of the 9 recording days. FFQ data were converted to mean daily intake by multiplying consumption frequency, number of serving units and weight of a unit (either standard or individual). The weight of a standard serving was either derived from pilot study data or from common Dutch household measures. If the number of serving units was omitted, the median number found among other questionnaires was taken instead. Since the serving sizes of potatoes and other bulk foods, such as rice and pasta, appeared to be proportionally related within subjects, the substituted number of serving units for these bulk foods was derived from the serving size of potatoes for the same subject. Season was taken into account when applicable.

Record and FFQ data (mean daily item intake) were both converted to nutrient intake using the computerized Dutch food composition table (NEVO, 1986). Although validation of supplement use was included in the study design, nutrient intake through supplements is not taken into account in this paper. Results indicated that vitamin supplements (A, C or multivitamin supplements) were used by 3–9% of the validation study population and correctly reported by 67% of the users; calcium supplements were correctly reported by 53% of the 15 (14%) users (Dorant *et al.*, 1994).

The items in the questionnaire were also aggregated into 27 food groups according to their shared properties and origin (e.g. bread, vegetables). For each food group mean daily weight consumed was calculated. The purpose of classification was to evaluate the validity of the questionnaire with respect to food group-related properties other than the nutrients studied and to facilitate interpretation of the strengths and limitations of the questionnaire.

Data analysis

The ratio of nutrient intake assessed by FFQ to that assessed by record was calculated individually. The distribution of the ratio was represented by a 95% range, derived as mean ± 2 SD. For ratios with a skewed distribution, the 95% range was derived from interpolating the 2.5 and 97.5 percentiles.

For the calculation of Pearson correlation coefficients between record and FFQ, nutrient intakes were $\log_e$-transformed to improve their distribution towards normality. Results were, however, similar for untransformed data. An alcohol intake of 0 g per day was replaced with 0.1 g per day before transformation. Energy-adjusted nutrient intakes were calculated as residuals from regression of each $\log_e$ (nutrient) on $\log_e$ (energy) and sex (Willett, 1990).

Furthermore, men and women were divided into quintiles according to nutrient intake (unadjusted and energy-adjusted) assessed by the questionnaire. For each quintile, the corresponding mean (untransformed) nutrient intake as assessed from the record was calculated (Willett, 1990). For this procedure, energy-adjusted residuals, to which mean nutrient intake was added, were calculated from untransformed energy and nutrient intakes.

For the comparison regarding the 27 food groups, most of which had a skewed distribution, a Spearman correlation coefficient was used. The specific food items within each food group were not analysed individually, because estimation of their usual consumption frequency on the basis of a 9-day record was expected to be imprecise.

Analysis of variance was applied to both the number of blank items and the error index of the baseline questionnaire ($\log_e$-transformed),

assessing the effects of group (validation study group versus reproducibility study group) and participation status for the repeated questionnaire (participants versus non-participants). The presence of a learning effect with respect to the number of blank items and the error index (repeated versus baseline questionnaire) was investigated in both study groups with a paired *t*-test.

To account for possible differences in the error index among the validation subgroup and the cohort at large, Pearson correlation coefficients were adjusted to the distribution of the error index in the baseline questionnaires of the random reproducibility sample. Calculations were performed using the error index dichotomized at the highest tertile (scores five and over) in the cohort. Thus, subjects were classified in a group with an error index below five or in one with higher scores. Regression analyses of nutrient intake assessed by record on that assessed by questionnaire were conducted within each of the two groups. In the usual formula for a squared correlation coefficient the residual sum of squares in the numerator was replaced with the residual sum of squares within both groups together with the sum of squares of regression over both groups. Thus, correlation coefficients can be calculated according to:

$$R^2 = 1 - \left[\sum_{i=1}^{2} (df_i * RMS_i) + SS \right] / SSY.$$

In this formula, i denotes the group (high and low error-index group), df the degrees of freedom (i.e. the number of subjects in each group minus 1), RMS the residual mean square of regression within group i, SS the sum of squares of the regression over groups and SSY the variance of the dependent variable (i.e. nutrient intake assessed by record). The actual adjustment of the correlation coefficient for the error index was conducted by substituting the degrees of freedom in both groups for those derived from the distribution of the error index in the reproducibility sample. In other words, the RMS in the formula, which was derived from the validation study results, was weighted according to the distribution of the error index in the cohort at large.

Finally, because day-to-day variation will still have influenced the observed mean individual intake based on 9 days (Nelson *et al.*, 1989), correlation coefficients were adjusted for this source of variation according to Beaton *et al.* (1979) with 95% confidence intervals according to Rosner & Willett (1988). For this purpose, the ratios of within-subject to between-subject variance of nutrient intake were calculated from the 9 recording days, ignoring day-of-the-week and period effects.

Results

Out of a total of 109 questionnaires completed by the validation study subjects, two (1.8%) were incomplete according to the formal criteria, leaving 107 questionnaires (59 from men and 48 from women) for analysis. The corresponding percentage for the reproducibility study group was 4.7. No questionnaires needed to be excluded for an error index exceeding 10.

Table 1 presents the mean daily nutrient intake for both dietary methods as well as the mean and 95% range of the ratio of the two measures. Furthermore, unadjusted and adjusted (for energy and sex) Pearson correlation coefficients are presented. Data for men and women were pooled since none of the correlations differed significantly between men and women. For most nutrients mean intake according to the FFQ was lower than according to the record; only for vegetable protein, polyunsaturated fat, dietary fibre, niacin and vitamin C the questionnaire gave a higher intake. On average, the questionnaire covered 91% of the record intake. Unadjusted correlation coefficients ranged from 0.40 [95% confidence interval (CI) 0.22–0.54] for vitamin B_1 to 0.86 (CI 0.80–0.90) for alcohol, with a median of 0.69. The only substantial (though statistically non-significant) differences in correlations between men and women were found for dietary fibre (0.79 and 0.63 respectively), vitamin A (0.58 and 0.46) and vitamin B_2 (0.66 and 0.55). For fibre the sex difference was attributable to the higher range in intake of bread for men, for vitamin B_2 one women had an outlying residual that was responsible for the lower correlation. Correlation coefficients adjusted for energy intake and sex ranged from 0.33 (CI 0.15–0.49) for vitamin B_1 to 0.86 (CI 0.80–0.90) for alcohol, with a median of 0.67. Spearman correlation coefficients, calculated from untransformed nutrient intakes, are included in

Table 1. Mean daily energy and nutrient intake as assessed by 9-day record and by the questionnaire, and Pearson correlation coefficients[a] (r) between the two methods (59 men, 48 women): Netherlands Cohort Study on Diet and Cancer

Nutrient	Record		Questionnaire		Ratio questionnaire/record (%)		Record with questionnaire		Record with baseline questionnaire
	Mean	SD	Mean	SD	Mean	95% range[b]	Unadjusted r	Adjusted[c] r	Adjusted[c] r
Energy (kJ)	9284	1862	7941	1996	86	57–115	0.74 (0.70)		
Total protein (g)	77.5	15.2	68.6	13.9	89	58–120	0.61 (0.54)	0.59 (0.52)	0.61
Vegetable protein (g)	24.5	6.0	24.6	7.2	101	62–140	0.77 (0.73)	0.68 (0.67)	0.71
Animal protein (g)	53.1	14.0	44.0	10.7	85	47–124	0.61 (0.54)	0.64 (0.58)	0.69
Total fat (g)	100.2	27.1	82.5	26.5	83	52–125*	0.72 (0.69)	0.52 (0.50)	0.47
Saturated (S) fat (g)	42.2	12.5	32.4	10.9	77	41–114	0.73 (0.73)	0.58 (0.53)	0.59
Polyunsaturated (P) fat (g)	17.0	7.1	17.8	9.1	109	50–170*	0.73 (0.70)	0.75 (0.73)	0.63
P/S ratio	0.42	0.19	0.58	0.28	143	80–240*	0.76 (0.79)	0.76 (0.77)	0.66
Cholesterol (mg)	330	99	243	76	76	45–128*	0.66 (0.71)	0.62 (0.65)	0.64
Total carbohydrates (g)	227.5	51.2	200.8	57.9	88	55–121	0.77 (0.72)	0.71 (0.65)	0.71
Mono-/disaccharides (g)	111.9	32.6	92.5	37.0	83	45–133*	0.78 (0.76)	0.79 (0.77)	0.68
Polysaccharides (g)	115.6	32.4	108.3	34.1	94	58–131	0.83 (0.76)	0.79 (0.75)	0.79
Dietary fibre (g)	25.7	6.8	27.3	7.7	108	76–150*	0.74 (0.68)	0.74 (0.68)	0.70
Alcohol (g)									
All participants	13.2	14.7	10.7	12.1	93	12–236*	0.86 (0.89)	0.86 (0.88)	0.85
Alcohol users only[d]	16.3	14.8	13.3	12.1	97	12–236*	0.78 (0.85)	0.76 (0.82)	0.74
Water (g)[e]	2281	508	2140	484	95	64–125	0.73 (0.73)	0.75 (0.74)	0.67
Calcium (mg)	1076	332	908	268	87	47–145*	0.60 (0.55)	0.62 (0.54)	0.62
Phosphorus (mg)	1545	345	1402	317	92	60–124	0.66 (0.58)	0.69 (0.67)	0.60
Potassium (mg)	3654	637	3551	695	98	67–129	0.66 (0.63)	0.71 (0.67)	0.62
Vitamin A (mg eq.[f])	0.95	0.32	0.87	0.29	96	49–180*	0.52 (0.49)	0.48 (0.44)	0.25
Vitamin B_1 (mg)	1.13	0.24	1.09	0.25	99	51–146	0.40 (0.42)	0.33 (0.37)	0.57
Vitamin B_2 (mg)	1.69	0.46	1.47	0.38	89	53–126	0.62 (0.58)	0.67 (0.62)	0.72
Vitamin B_6 (μg)	1447	294	1416	322	99	63–135	0.67 (0.65)	0.67 (0.62)	0.61
Niacin (mg)	13.3	4.3	13.6	3.9	107	65–155*	0.62 (0.61)	0.61 (0.64)	0.67
Vitamin C (mg)	96.7	42.4	104.4	39.4	118	60–260*	0.58 (0.52)	0.55 (0.51)	0.42
Iron (mg)	12.8	2.7	12.4	2.7	98	69–137	0.58 (0.61)	0.53 (0.54)	0.48
Protein, % of energy intake	14.2	2.5	14.7	2.4	105	77–132	0.67 (0.62)	0.59 (0.58)	0.61
Fat, % of energy intake	40.3	5.0	38.7	5.6	97	72–121	0.57 (0.56)	0.52 (0.50)	0.47
Carbohydrates, % of energy intake	41.3	5.5	42.4	5.9	103	81–125	0.72 (0.68)	0.71 (0.68)	0.71

[a] Based on $\log_e$-transformed values. In parentheses: Spearman correlation coefficients for untransformed data.
[b] 95% range is based on mean ± 2 SD; 95% range for ratios with a skewed distribution (noted with *) are based on interpolation of 2.5 and 97.5 percentiles.
[c] Adjusted for energy intake and sex.
[d] n = 86, based on alcohol users according to questionnaire.
[e] Includes water in beverages and foods.
[f] mg equivalents vitamin A activity: retinol (mg) + ß-carotene (mg)/6.

Table 1 for the purpose of comparison. They were slightly lower than the corresponding Pearson correlation coefficients.

To assess the relation between record and FFQ data for different reference periods, the baseline questionnaire was also compared to the record (Table 1). Energy- and sex-adjusted correlation coefficients for the baseline questionnaire ranged from 0.25 (CI 0.06–0.43) for vitamin A to 0.85 (CI 0.79–0.89) for alcohol, with a median of 0.64. No systematic differences were found between the baseline and repeated questionnaires with respect to absolute intake of energy and nutrients (data not shown).

Table 2 shows the intake of food groups and the correlation between the two methods. As for the nutrients, the FFQ generally resulted in lower intakes than the record; exceptions were vegetables, citrus fruits, bread and added fats. On average, the mean of the intakes of all food groups as assessed by questionnaire accounted for 85% of the record assessment. The Spearman correlation coefficients ranged from 0.38 for vegetables to 0.89 for alcoholic beverages, with a median of 0.60.

Table 3 visualizes the actual level of and the heterogeneity in nutrient intake that could be discriminated by quintiles derived from the FFQ. Energy adjustment decreased the range for the energy-contributing nutrients. For some nutrients the results suggest non-linear relationships. For example, the questionnaire was not able to separate the two highest quintiles of vitamin C intake, but could nevertheless discriminate a twofold range. The anomaly in the two highest quintiles of vitamin C intake was likely to be attributed to subjects consuming fresh orange juice who checked both the item on (pressed) oranges and that on orange juice.

Table 4 shows the mean number of blank items and the mean error index of the FFQ according to study group and participation. The validation group had a lower number of blank items and a slightly lower mean error index at baseline than the reproducibility group. Furthermore, among both study groups, the base-

Table 2. Mean daily intake of food groups (g)[a] as assessed by 9-day record and by the questionnaire, including correlation coefficients (59 men, 48 women)

	Record		*Questionnaire*		
Food group	*Mean*	*SD*	*Mean (% of record)*	*SD*	*Spearman's r*
Potatoes	162	83	136 (84)	73	0.74
Rice	19	36	17 (88)	30	0.39
Vegetables	160	83	189 (118)	69	0.38
Fruits	207	107	189 (91)	114	0.60
Citrus fruits	67	59	72 (107)	70	0.68
Other fruits	140	81	117 (84)	91	0.60
Bread	134	54	159 (119)	70	0.80
Milk and milk products	363	220	311 (86)	192	0.60
Cheese	33	20	21 (64)	15	0.61
Eggs	20	13	15 (75)	10	0.61
Meat	99	38	97 (98)	36	0.46
Meat products	20	16	12 (57)	11	0.54
Fish	19	23	11 (58)	12	0.53
Other sandwich filling[b]	15	12	11 (71)	11	0.68
Added fats	45	21	47 (103)	25	0.57
Added sugar	19	22	16 (82)	22	0.84
Cakes, cookies	51	30	28 (56)	23	0.65
Soup	72	70	67 (94)	87	0.54
Non-alcoholic beverages	1131	420	1102 (97)	383	0.63
Alcoholic beverages	139	222	99 (71)	138	0.89

[a] Food groups with mean intake <10 g per day are not listed in this table: pulses, cereals, mixed dishes, nuts, snacks, candy and soy products.

[b] Including peanut butter, jam and other sweet fillings.

Table 3. Mean nutrient intake assessed by 9-day record according to quintile categories of nutrient intake assessed by questionnaire (59 men, 48 women)

	Unadjusted					*Adjusted for energy*[a]				
Nutrient	Q_1	Q_2	Q_3	Q_4	Q_5	Q_1	Q_2	Q_3	Q_4	Q_5
Energy (kJ)										
Men	8288	9698	10021	10506	11883					
Women	6845	8565	7640	9071	9514					
Protein (g)										
Men	78	74	83	83	99	80	75	82	88	93
Women	58	71	71	76	76	60	69	72	73	78
Total fat (g)										
Men	84	97	109	113	138	100	104	108	111	117
Women	71	87	89	93	111	85	84	91	93	99
Polyunsaturated fat (g)										
Men	12	15	20	21	26	12	17	19	22	23
Women	10	14	13	18	20	11	10	15	17	21
Cholesterol (mg)										
Men	253	284	343	414	455	272	318	343	352	465
Women	219	287	285	338	396	257	272	308	335	351
Mono-, disaccharides (g)										
Men	75	103	126	124	155	82	103	122	128	147
Women	77	103	101	115	138	88	98	102	116	130
Polysaccharides (g)										
Men	94	111	127	139	169	99	116	127	144	154
Women	77	100	99	116	112	87	95	100	102	118
Dietary fibre (g)										
Men	20	24	26	29	37	20	24	25	29	37
Women	19	23	26	24	27	20	23	24	24	29
Alcohol (g)										
Men	1	6	14	27	36	1	7	14	28	35
Women	0	3	4	11	24	1	3	4	12	24
Calcium (mg)										
Men	874	1065	1012	1084	1481	934	925	1192	1018	1463
Women	756	1026	982	1146	1297	799	945	1015	1106	1332
Vitamin A (mg eq.[b])										
Men	0.76	0.97	0.89	1.03	1.32	0.87	0.86	0.97	1.07	1.21
Women	0.76	0.81	0.90	0.87	1.16	0.75	0.87	0.93	0.93	1.02
Vitamin C (mg)										
Men	59	81	87	103	94	53	86	92	100	93
Women	75	102	101	138	141	75	93	108	139	142
Fat, % of energy intake										
Men	34	40	39	43	44	38	37	39	42	43
Women	38	39	39	43	45	38	40	40	42	44

[a] Adjusted intakes for record and questionnaire nutrients were calculated as residuals from regression analyses of nutrient on energy intake, to which mean nutrient intake was added.
[b] mg equivalents: retinol (mg) + ß-carotene (mg)/6.

line questionnaires of the participants had significantly fewer blank items and response errors than those of the non-participants; in both study groups combined, the mean number of blank items for participants was 18.6 vs 22.3 for non-participants, while the mean error index was 3.2 and 3.8 respectively. There was no evidence for an interaction effect between participation status and study group. Apparently, (future) validation study participants completed their baseline questionnaires somewhat better than the other cohort members. Compared to the baseline questionnaire, the mean error index of the repeated questionnaire was not significantly lowered, indicating the absence of a learning effect attributable to recording of intake. The number of blank items, however, was significantly lower for the

Table 4. Mean number of blank items and mean error index[a] of baseline questionnaire and (repeated) questionnaire by study group and participation status

Study group/ participation status	*Number of blank items*				*Error index*			
	Baseline		*Repeated*		*Baseline*		*Repeated*	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
Validation study group[b]								
All subjects (n = 199)	18.1	25.8			3.3	2.3		
Participants (n = 104)	16.3	25.5	14.5[d]	23.9	3.0	1.9	2.9	1.9
Non-participants (n = 95)	20.0[c]	26.2			3.5[c]	2.6		
Reproducibility study group[b]								
All subjects (n = 373)	20.8	27.5			3.4	2.4		
Participants (n = 281)	19.4	26.2	15.7[d]	24.2	3.2	2.4	3.5	2.2
Non-participants (n = 92)	25.1[c]	30.7			4.1[c]	2.4		

[a] Error index is the sum of 15 scores each representing an inconsistency or other response error in the completed questionnaire (0 = no error, 1 = moderate error, 2 = serious error). Since the value of 2 was not assigned for 7 out of 15 variables, the maximum score that could theoretically be attained was 23; actually, the highest score encountered in the cohort was 14. The error index did not include the number of blank items. Its exact composition is available on request.

[b] Data refer to accepted questionnaires.

[c] Significant difference ($P < 0.05$, $\log_e$-transformed data) between participants and non-participants, both study groups combined. There was no evidence for an interaction effect between participation status and study group.

[d] Significant difference ($P = 0.001$, paired t-test) between baseline and repeated questionnaire, both study groups combined.

repeated questionnaire, particularly in the reproducibility group.

Table 5 shows some implications of the validation study results for the cohort as a whole. The Pearson correlation coefficients for intake of the major nutrients, adjusted for energy and sex as presented in Table 1, were adjusted for the distribution of the error index in the questionnaires of the cohort, i.e. the baseline questionnaires of the random reproducibility sample. Based on the dichotomized error index, 21% of the questionnaires of the validation study participants appeared to fall within the high error group, compared to 32% of the baseline questionnaires of the random reproducibility sample. The anticipated decrease in correlation coefficients was small and appeared to be mainly restricted to cholesterol and vitamin C. There was no need to adjust for the difference in number of blank items between the two groups, since it resulted in minor (less than 1%), non-significant differences in nutrient intake, which was considered as a measure of underreporting.

The attenuation of the correlation coefficients due to the relatively low number of 9 recording days is demonstrated by the effect of correction for day-to-day variation in the record data (Table 5). Although the 95% confidence intervals became somewhat wider, correlation coefficients increased on average by 0.07. Due to their relatively large day-to-day variation, the effects of de-attenuation were most pronounced for cholesterol and vitamins A and C.

Discussion

We have validated a self-administered dietary questionnaire for use in a large-scale prospective cohort study on diet and cancer in The Netherlands. A number of parameters are available to evaluate the validity of an instrument or method relative to another method (e.g. Willett, 1990). We have chosen three of them: (a) the ratio of FFQ to record nutrient intake, (b) the (Pearson and Spearman) correlation coefficient and (c) the distribution of mean nutrient intakes assessed by the record according to quintile categories of intake assessed by the questionnaire. Although the use of a correlation coefficient for validation purposes in general is criticized by some (Bland & Altman, 1986), it has some attractive properties relevant to the

Table 5. Pearson correlation coefficients between 9-day diet record and questionnaire for nutrient intake (adjusted for energy and sex), after correction for error index of the questionnaire and day-to-day variation in the record (59 men and 48 women)

	Correlation between record and questionnaire[a]		*Adjusted for error index*[b]	*Adjusted for error index and day-to-day variation*	
Nutrient	*r*	*(95% CI)*	*r*	r^c	*(95% CI)*[d]
Protein	0.59	(0.45, 0.70)	0.58	0.64	(0.48, 0.75)
Total fat	0.52	(0.37, 0.65)	0.49	0.53	(0.36, 0.67)
Polyunsaturated fat	0.75	(0.65, 0.82)	0.76	0.80	(0.70, 0.87)
Cholesterol	0.62	(0.49, 0.72)	0.56	0.67	(0.48, 0.80)
Mono-, disaccharides	0.79	(0.71, 0.85)	0.80	0.83	(0.75, 0.89)
Polysaccharides	0.79	(0.71, 0.85)	0.80	0.84	(0.75, 0.89)
Dietary fibre	0.74	(0.64, 0.82)	0.73	0.79	(0.67, 0.87)
Alcohol	0.86	(0.80, 0.90)	0.83	0.86	(0.79, 0.91)
Calcium	0.62	(0.49, 0.72)	0.62	0.66	(0.51, 0.76)
Vitamin A	0.48	(0.32, 0.61)	0.52	0.76	(0.41, 0.91)
Vitamin C	0.55	(0.40, 0.67)	0.50	0.58	(0.39, 0.72)
Mean	0.66		0.65	0.72	

[a] Derived from Table 1.
[b] See Methods section (data analysis) for adjustment procedure.
[c] According to Beaton *et al.* (1979).
[d] According to Rosner & Willett (1988).

aetiological purpose of the cohort study: the correlation coefficient reflects the questionnaire's capacity to rank subjects according to exposure (more important than absolute agreement), taking into account the true variation in exposure in the population studied (Willett, 1990). Thus, it is an adequate measure of the performance of the questionnaire in the cohort population to which it is actually applied. Furthermore, it facilitates comparison with other validation studies of self-administered dietary questionnaires.

Compared to a number of other self-administered questionnaires developed for a similar purpose (Table 6), our questionnaire, comprising approximately 150 food items, is comprehensive. This is mainly due to our requirement to rank subjects with respect to both nutrient and energy intake. In this study, the Pearson correlation coefficients, for unadjusted as well as sex- and energy-adjusted intakes, were generally higher than for questionnaires with fewer items, but similar to the Finnish questionnaire with 276 items (Pietinen *et al.*, 1988*a*). Inspection of Table 6 may lead to the tentative conclusion that the validity of a questionnaire is proportional to its length, although not all questionnaires match this rule (Pietinen *et al.*, 1988*b*; Rimm *et al.*, 1992). Of course, other properties of the questionnaire, such as layout, data editing procedures (Block & Hartman, 1989) and characteristics of the population [dietary pattern, range in intake, motivation and ability to complete the questionnaire (Block & Hartman, 1989)] also influence the validity of the questionnaire.

The correlation coefficients for food groups appeared to be somewhat lower than those for nutrients. Differences in coding of foods between the two methods are partly responsible for this: many of the record data were coded as ingredients from recipes or mixed dishes as opposed to the questionnaire data, which were coded as food product. Consequently, the division between food groups was not always clear, resulting in lower correlations. For some food groups, such as vegetables, the relatively low correlation (0.38) was due to a lack of variation in consumption frequency combined with, in our experience, imprecise estimation of portion size. We have evidence, however, that correlations for specific vegetables will be higher due to larger variation in consumption frequency (Den Breeijen, 1990).

Table 6. Comparison of validation studies of self-administered questionnaires using the diet record as reference method with respect to the intake of some nutrients important in diet and cancer studies (Pearson correlation coefficients)

First author	Willett	Willett	Pietinen	Pietinen	Block	Tjønneland	Rimm	This study
Year	1985	1987	1988*b*	1988*a*	1990	1991	1992	
Sex of subjects	F	M+F[a]	M	M	F	M+F[a]	M	M+F[a]
Number of items	61	116	44	276	94	92	131	150
Energy	[b]	0.37	0.43	0.59	0.51	0.32	0.40	0.69
Fat								
Unadjusted	0.39	0.57	0.42	0.60	0.60	0.41	0.52	0.69
Energy-adjusted[c]	0.53	0.59	0.47	0.52		0.58	0.61	0.52
Polyunsaturated fat								
Unadjusted	0.40	0.50	0.68	0.73	0.48	0.41	0.33	0.71
Energy-adjusted[c]	0.48	0.28	0.77	0.76		0.46	0.29	0.75
Fibre								
Unadjusted	0.46	0.37	0.67	0.70		0.34	0.49	0.74
Energy-adjusted[c]	0.58	0.65	0.61	0.73		0.46	0.64	0.74
Calcium								
Unadjusted		0.42		0.62	0.56	0.38	0.52	0.60
Energy-adjusted[c]		0.57		0.66		0.55	0.53	0.62
Vitamin A								
Unadjusted	0.26	0.62	0.38	0.51	0.47	0.27	0.35	0.53
Energy-adjusted[c]	0.36	0.70	0.36	0.49		0.36	0.41	0.48
Vitamin C								
Unadjusted	0.63	0.34	0.40	0.59	0.56	0.55	0.64	0.54
Energy-adjusted[c]	0.66	0.49	0.53	0.60		0.58	0.68	0.55

[a] Studies with both sexes are adjusted for sex (study by Tjønneland, *et al.*, 1991, based on mean for men and women); study by Willet *et al.*, (1987) also adjusted for age because of the large age range of the study population (20–54 years).
[b] Empty entries: no data published.
[c] Adjusted for energy and sex, if applicable; study by Willet *et al.*, (1987) also adjusted for age.

The general underestimation by our questionnaire of absolute mean nutrient intake is more pronounced than for other questionnaires (Willett *et al.*, 1985; Pietinen *et al.*, 1988*b*; Block *et al.*, 1990; Rimm *et al.*, 1992). Underreporting, caused by an incomplete list of foods and items erroneously left blank, counteracts the effect of overreporting caused by long lists of the same sort of items (Block & Hartman, 1989). In this study, overreporting due to long enumerations was likely to have occurred for vegetables, citrus fruits and meat. The consumption frequencies for specific meat types, however, were adjusted to the reported weekly frequency of meat consumption, because the adjustment appeared to increase correlation coefficients for the meat types (Den Breeijen, 1990). The overreporting of bread may be due to occasional substitution of bread with other foods, such as crackers.

As in other studies (Willett *et al.*, 1985; Pietinen *et al.*, 1988*a*,*b*; Rimm *et al.*, 1992), comparison of the baseline questionnaire with the dietary record revealed that it performed almost as well as the repeated questionnaire, which was actually to be validated. It shows that synchronization of the period of reference for the diet record and the questionnaire was not very important, which is indicative of stable dietary habits over time. This result is reassuring when it is considered that a single measurement has to characterize a subject's long-term dietary intake to link it to cancer risk. It may also indicate the absence of a training effect of the diet record keeping which has been suggested by some.

A criticism of validation studies is that the participants are highly motivated and will do better than the population in which the method has been applied at large. This is a particular problem when response to the validation study is relatively low such as for this and other study populations that were not selected for high motivation from the very start (Hankin *et al.*, 1991; Rimm *et al.*, 1992). Indeed, the percentage of questionnaires rejected for incompleteness was 6.0 for the baseline questionnaires of the cohort and 4.7 and 1.8 for the

questionnaires repeated in the reproducibility study and the validation study respectively, whereas the number of blank items in accepted questionnaires was also lower for the participants. Similarly, subjects who were willing to participate in the validation study or the reproducibility study had fewer errors in their baseline questionnaires already. Apparently, subjects who have more problems with the questionnaire, or have completed it somewhat carelessly, are less inclined to participate for a second time, in particular in a demanding method like a diet record.

Lack of comparability of the study groups with respect to completeness of the questionnaires is largely solved by exclusion of incomplete questionnaires from all analyses according to identical criteria. Moreover, the difference regarding the number of blank items within accepted questionnaires did not result in differential underreporting. Adjustment for the impact of the difference between the two groups in the error index, which is conceptually more directly related to the questionnaire's performance than, for example, nutrient intake and level of education, has shown that selection of the validation study group did not appear to influence the generalizability of the results to the cohort at large.

The adjustment for intraindividual variation in nutrient intake as determined by the record shows that some of the observed correlation coefficients were attenuated by the relatively small number of 9 recording days. Vitamins A and C have both relatively low observed correlations. However, the low correlation of vitamin A apparently has been caused by the high day-to-day variation in the record data, whereas vitamin C assessment depends more on questionnaire performance, as was also suggested by the quintile analysis.

In conclusion, we have shown that the questionnaire is able to rank subjects adequately according to intake of the food groups and nutrients investigated. Compared to a 'perfect' method, application of the FFQ requires an increase in cohort size to account for the attenuation of associations between nutrient intake and cancer observed in the cohort study. Walker & Blettner (1985) have calculated that to detect a trend in a true relative rate of 3 between the lowest and highest quintile of intake and $\alpha = 0.05$ (two-sided) and $1\text{-}\beta = 0.95$, a threefold increase in cohort size is required when the correlation between FFQ and the perfect measure lies between 0.7 and 0.6. The observed relative rate is then reduced to between 1.5 and 2.

Although the validation study group differed from the cohort with respect to completeness and quality of their questionnaires, this appeared to be no major threat to the generalizability of the validation study's results to the cohort.

Acknowledgements—We are indebted to Gemma Snijders who was responsible for the practical organization of the study, to the students who have collected the dietary record data, to Hanneke den Breeijen and Harry van Montfort for developing part of the questionnaire processing software, to Cor Kistemaker for supplying the software for calculation of nutrient intake and to Dr Frans J. Kok for reviewing the manuscript. Supported by grants CIVO 86-1 and CIVO 90-3 from the Dutch Cancer Society.

References

Bausch-Goldbohm RA, van den Brandt PA, van't Veer P, Sturmans F & Hermus RJJ (1988): Results of the methodological study for the design of a simplified, self-administered questionnaire. In *Diet, hormones and cancer: methodological issues for prospective studies*, eds E Riboli & R Saracci, pp. 79–89. IARC Technical Report No. 4. Lyon, France: International Agency for Research on Cancer.

Beaton GH, Milner J, Corey P, McGuire V, Cousins M, Stewart E, Ramos M de, Hewitt D, Grambsch PV, Kassim N & Little JA (1979): Sources of variance in 24-hour dietary recall data: implications for nutrition study design and interpretation. *Am. J. Clin. Nutr.* **32**, 2546–2559.

Bland JM & Altman DG (1986): Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 307–310.

Block G & Hartman AM (1989): Issues in reproducibility and validity of dietary studies. *Am. J. Clin. Nutr.* **50**, 1133–1138.

Block G, Woods M, Potosky A & Clifford C (1990): Validation of a self-administered diet history questionnaire using multiple diet records. *J. Clin. Epidemiol.* **43**, 1327–1335.

Den Breeijen H (1990): *Validation of the draft version of a self-administered questionnaire used in a cohort study on diet and cancer* [in Dutch]. TNO report V 90.019. Zeist, The Netherlands: TNO Toxicology and Nutrition Institute.

Dorant E, Van den Brandt PA, Goldbohm RA, Hermus RJJ & Sturmans F (1994): Agreement between interview data

and a self-administered questionnaire on dietary supplement use. *Eur. J. Clin. Nutr.* **48**, 180–188.

Dutch Board for Vegetables and Fruit (1986): *Results of a panel survey: fresh vegetables and fruits* [in Dutch]. The Hague: Produktschap voor Groenten en Fruit, Market Research Department.

Dutch Meat Board (1986): *Results of a panel survey by on the consumption of meat and meat products* [in Dutch]. Rijswijk, The Netherlands: Voorlichtingsbureau Vlees.

Hankin JH, Wilkens LR, Kolonel LN & Yoshizawa CN (1991): Validation of a quantitative diet history method in Hawaii. *Am. J. Epidemiol.* **133**, 616–628.

Jain MG, Harrison L, Howe GR & Miller AB (1982): Evaluation of a self-administered dietary questionnaire for use in a cohort study. *Am. J. Clin. Nutr.* **36**, 931–935.

Livingstone MBE, Prentice AM, Strain JJ, Coward WA, Black AE, Barker ME, McKenna PG & Whitehead RG (1990): Accuracy of weighed dietary records in studies of diet and health. *Br. Med. J.* **300**, 708–712.

Nelson M, Black AE, Morris JA & Cole TJ (1989): Between- and within-subject variation in nutrient intake from infancy to old age: estimating the number of days required to rank dietary intakes with desired precision. *Am. J. Clin. Nutr.* **50**, 155–167.

NEVO tabel (1986): *Dutch food composition table 1986–1987*. The Hague: Voorlichtingsbureau voor de Voeding.

Pietinen P, Hartman AM, Haapa E, Räsänen L, Haapakoski J, Palmgren J, Albanes D, Virtamo J & Huttunen JK (1988*a*): Reproducibility and validity of dietary assessment instruments: a self-administered food use questionnaire with a portion size picture booklet. *Am. J. Epidemiol.* **128**, 655–666.

Pietinen P, Hartman AM, Haapa E, Räsänen L, Haapakoski J, Palmgren J, Albanes D, Virtamo J & Huttunen JK (1988*b*): Reproducibility and validity of dietary assessment instruments: a qualitative food frequency questionnaire. *Am. J. Epidemiol.* **128**, 667–676.

Rimm EB, Giovannucci EL, Stampfer MJ, Colditz GA, Litijn 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.

Rosner B & Willett WC (1988): Interval estimates for correlation coefficients corrected for within-person variation: implications for study design and hypothesis testing. *Am. J. Epidemiol.* **127**, 377–386.

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

Tjønneland A, Overvad KIM, Haraldsdóttir J, Bang S, Ewertz M & Jensen OM (1991): Validation of a semiquantitative food frequency questionnaire developed in Denmark. *Int. J. Epidemiol.* **20**, 906–912.

Van den Brandt PA, Goldbohm RA, van't Veer P, Volovics A, Hermus RJJ & Sturmans F (1990): A large-scale prospective cohort study on diet and cancer in The Netherlands. *J. Clin. Epidemiol.* **43**, 285–295.

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

Willett WC (1990): Reproducibility and validity of food-frequency questionnaires. In: *Nutritional epidemiology*, ed. WC Willett, pp. 92–126. New York/Oxford: Oxford University Press.

Willett WC, Sampson L, Stampfer MJ, Rosner B, Bain C, Witschi J, Hennekens CH & Speizer FE (1985): Reproducibility and validity of a semiquantitative food frequency questionnaire. *Am. J. Epidemiol.* **122**, 51–65.

Willett WC, Reynolds RD, Cottrell-Hoehner S, Sampson L & Browne ML (1987): Validation of a semi-quantitative food frequency questionnaire: comparison with a 1-year diet record. *J. Am. Diet. Ass.* **87**, 43–47.