Validation and calibration of food-frequency questionnaire measurements in the Northern Sweden Health and Disease cohort

Ingegerd Johansson1,*, Göran Hallmans1, Åsa Wikman1, Carine Biessy2, Elio Riboli2 and Rudolf Kaaks2

1Department of Nutritional Research, Umeå University, Sweden: 2Unit of Analytical Epidemiology, Program of Nutrition and Cancer, International Agency for Research on Cancer, Lyon, France

Submitted 14 August 2001: Accepted 29 October 2001

Abstract

Objectives: To evaluate the reproducibility of, and to compare and calibrate, diet measures by the Northern Sweden 84-item food-frequency questionnaire (FFQ) with measures from 24-hour diet recalls (24-HDR).

Design: Randomly selected respondents (n = 246) from the EPIC (diet–cancer) and MONICA (diet–cardiovascular disease) study cohort in Northern Sweden were invited to answer the FFQ twice over a one-year interval (FFQ1 and FFQ2), and to complete ten 24-hour recalls (reference method) in the months between. Plasma β-carotene concentrations were determined from a subset of 47 participants.

Setting: Västerbotten and Norrbotten, Northern Sweden.

Participants: Ninety-six men and 99 women, who completed the study.

Results: The reproducibility of the FFQ was high in terms of both mean energy and nutrient intakes and relative ranking of participants by intake levels (median Pearson correlation of 0.68). Moderately higher food intake frequencies were recorded by FFQ1 compared with 24-hour recalls for dairy products, bread/cereals, vegetables, fruits and potato/rice/pasta, whereas meat, fish, sweet snacks and alcoholic beverage intakes were lower. The median Spearman coefficient of correlation between FFQ1 and the average of ten 24-HDR measurements was 0.50. Daily energy and nutrient intakes were similar for FFQ1 and 24-HDR measurements, except for fibre, vitamin C, β-carotene and retinol (FFQ1 > 24-HDR) and sucrose and cholesterol (FFQ1 < 24-HDR). Pearson coefficients of correlation between FFQ1 and 24-HDR corrected for attenuation due to residual day-to-day variation in the 24-HDR measurements ranged from 0.36 to 0.79 (median 0.54). Adjustment for energy had only very moderate effects on the correlation estimates. Calibration coefficients estimated by linear regression of the 24-HDR on the FFQ1 measurements varied between 0.30 and 0.59 for all nutrients except alcohol, which had calibration coefficients close to 1.0. These low calibration coefficients indicate that relative risk estimates corresponding to an absolute difference in dietary intake levels measured by the FFQ will generally be biased towards 1.0. Plasma β-carotene levels had a Pearson coefficient of correlation of 0.47 with the 24-HDR measurements, and of 0.23 with FFQ1 measurements.

Conclusions: The Northern Sweden FFQ measurements have good reproducibility and an estimated level of validity similar to that of FFQ measurements in other prospective cohort studies. The results from this study will form the basis for the correction of attenuation and regression dilution biases in relative risk estimates, in future studies relating FFQ measurements to disease outcomes.

Dietary intake estimates from self-administered food-frequency questionnaires (FFQs) are commonly used in epidemiological studies1−8, but, like any other type of dietary intake measurement, suffer from systematic and random errors. As such errors generally cause bias in relative risk estimates, it is crucial to estimate the validity of the intake assessments, so as to enhance the interpretation of estimated diet–disease associations and to improve the translation of such associations into dietary recommendations. Using the results from validation or calibration sub-studies it may also be possible to adjust relative risk estimates for (at least part of) biases induced by errors in the dietary exposure assessments. Thus, the validation and calibration of FFQ measurements of habitual dietary...
intakes are an essential part of large-scale epidemiological studies, and especially for prospective cohort studies in which the dietary intake assessments will eventually be related to a variety of chronic disease outcomes.

The population of the most arctic counties of Sweden (Västerbotten and Norrbotten) is part of the global cohort for the prospective European Prospective Investigation into Cancer and Nutrition (EPIC) as well as for the WHO MONICA cardiovascular study. A self-administered FFQ was chosen for monitoring of dietary intakes in the Northern Sweden EPIC and MONICA centres (Northern Sweden Health–Diet Study, NSHDS). Data collection started in 1985 within the Västerbotten cardiovascular intervention project, in which all adults within the age span of 30–75 years in Västerbotten county are invited every 10 years for a general health examination. On each visit, dietary intake assessments are collected from all participants, and at present about 130,000 FFQ measurements of habitual dietary intake have been obtained from about 90,000 cohort participants.

The aim of the present study was to evaluate the reproducibility of the self-administered FFQ used in these northern Swedish cohorts, and to compare and calibrate FFQ measurements with 24-hour diet recalls (24-HDR) and plasma β-carotene levels. The study was conducted on a representative sub-sample of the EPIC and MONICA cohorts.

Study population and methods

Study population

A random sub-sample in individuals attending the ongoing Västerbotten county cardiovascular disease (CVD) study in 1992 was requested by mail to participate in the present study. Out of a total of 246 individuals invited, 43 gave a negative response because they could not find the free time to participate in the study. Thus, 102 men and 101 women, equally distributed over the ages 30, 40, 50 and 60 years, consented to participate. The participants were recruited in equal proportions from the coastal (more urban) and the inland/mountain (more rural) areas of Västerbotten county.

The study was approved by the Ethics Committee for Human Experiments at Umeå University, Sweden.

Food-frequency questionnaire (test method)

The participants were requested to complete one set of a self-administered FFQ immediately before initiation of the 24-HDR period (in the first quarter of 1993, FFQ1) and a second identical set at the end (in the first quarter of 1994, FFQ2). The FFQ was designed to be semi-quantitative and optically readable for data input. Frequencies of consumption of 84 food items were reported on an increasing, nine-level scale, including never, maximum once a month, 1–3 times per month, once a week, 2–3 times a week, 4–6 times a week, once a day, 2–3 times a day, and 4 or more times a day. The questionnaire included eight questions on the frequency of consumption of various types of fats used for spreading on bread or cooking, nine on milk and other dairy products, eight on bread and cereals, 10 on fruit, greens and root vegetables, and nine on soft drinks and sugar-containing snacks. Five questions on spirits, wine and beer consumption were included in a list of beverages. Twenty-nine of the remaining 35 questions recorded intake of potato, rice, pasta, meat and fish, and six were on varied items, such as salty snacks, coffee, tea and juice. The respondents indicated their average portion of (1) potato/pasta/rice, (2) vegetables and (3) meat/ground meat/sausages by comparison with four colour photos illustrating four plates with increasing portion sizes of potato, vegetables and meat. For the other food items, we assumed a standard portion size value (see below). The questionnaires were returned by mail, and the interviewer reviewed it and resolved any ambiguities by telephone.

The reported frequencies of consumption were converted to number of intakes per day, and energy and nutrient intakes were calculated by multiplying these frequencies by a portion size value and by the energy or nutrient content from a food composition database from the National Food Administration. Portion sizes used were those indicated on the photos, or were natural portion sizes such as an apple, or average portion sizes for gender and age according to a national survey. The energy and nutrient contents were calculated using the software MATs (Rudans Lättdata, Sweden).

24-hour recalls (reference method)

Because many participants lived at a large distance from the research centre in Umeå, 24-hour diet recall interviews were conducted by telephone. The interviews were done by nine trained interviewers, i.e. all interviewers had academic education in nutrition, had participated in nutritional research previously, and were trained to structure the interviews in a standardised fashion. The interviews recorded intake of food, beverages and supplements during the preceding 24 hours, at 10 unannounced occasions per subject. The ten 24-HDR were equally dispersed over the year and covered all weekdays. The estimation of food portions was facilitated by the use of a booklet with full-size portion size pictures (National Food Administration, Uppsala, Sweden), mailed in advance to the participants. These pictures included five options for portions of food on plates, five options for spread on a knife, and a set of schematic drawings to indicate thickness and sizes of various types of food items. Standard household measures were used for the estimation of portion sizes of food items not included in the booklet.

The daily energy and nutrient intake and intake frequencies of food items in 10 food groups were calculated using the software MATs (Rudans Lättdata,
mainly present for food groups with several alternative

...ores when bed-bound, was applied as cut-off for evident

underreporting.

...corresponding to a value slightly higher than the survival

...to basal metabolic rate22. A PAL value below 1.27 21,

...defined as the ratio of total energy expenditure relative

...this ratio corresponds to the physical activity level (PAL),

...to BMR. For participants in approximate energy balance,

...energy requirements

...to identify participants who might have been

...underreporting relative to daily

...energy requirements

...in order to identify participants who might have been

...underreporting relative to daily

...energy requirements

...in order to identify participants who might have been

...underreporting relative to daily

...energy requirements

...in order to identify participants who might have been

...underreporting relative to daily

...energy requirements

...theses. None of these

...took β-carotene supplementation. Samples, drawn after

...overnight fasting and two days after ceasing any vitamin

...supplementation, were protected from light and air

...exposure and stored in liquid nitrogen. Plasma concen-

...tration of β-carotene was determined by reversed-phase

...high-performance liquid chromatography (HPLC) separa-

...tion20. Briefly, after hexane extraction (0.15 g l−1 of

...butylated hydroxytoluene (BHT) added), evaporation

...under nitrogen and re-dissolving in ethanol, lipid soluble

...vitamins were separated on a Spherisorb ODS-2 column

...(C-18, 5 μm, HPLC Teknik AB, Umeå, Sweden) using a

...mobile phase of 95% methanol and 5% tetrahydrofuran.

...Tocopherylacetate (Sigma Chemicals, St Louis, USA) was

...used as an internal standard and β-carotene (Sigma

...Chemicals) as an external standard.

...Estimation of underreporting relative to daily

...energy requirements

...In order to identify participants who might have been

...underreporting relative to daily

...energy requirements

...In order to identify participants who might have been

...underreporting relative to daily

...energy requirements

...In order to identify participants who might have been

...underreporting relative to daily

...energy requirements

...In order to identify participants who might have been

...underreporting relative to daily

...energy requirements

...In order to identify participants who might have been

...underreporting relative to daily

...energy requirements

...In order to identify participants who might have been

...underreporting relative to daily

...energy requirements

...In order to identify participants who might have been

...underreporting relative to daily

...energy requirements

...In order to identify participants who might have been

...underreporting relative to daily

...energy requirements

...In order to identify participants who might have been

...underreporting relative to daily

...energy requirements

...In order to identify participants who might have been

...underreporting relative to daily

...energy requirements

...In order to identify participants who might have been

...underreporting relative to daily

...energy requirements

...In order to identify participants who might have been

...underreporting relative to daily

...energy requirements

...In order to identify participants who might have been

...underreporting relative to daily

...energy requirements

...In order to identify participants who might have been

...underreporting relative to daily

...energy requirements

...In order to identify participants who might have been

...underreporting relative to daily

...energy requirements

...In order to identify participants who might have been

...underreporting relative to daily

...energy requirements

...In order to identify participants who might have been

...underreporting relative to daily
were performed on both the log-transformed absolute intake levels and the nutrient densities. Calibrated FFQ measurements (FFQ') were calculated as:

$$\text{FFQ'} = \text{24-HDR}_{\text{mean}} + \lambda (\text{FFQ} - \text{FFQ}_{\text{mean}}),$$

where 24-HDR_{\text{mean}} and FFQ_{\text{mean}} are the population mean intake levels as measured by the 24-HDR and FFQ, respectively. As a less parametric form of 'calibration' approach, we also compared the geometric mean 24-HDR frequencies of food consumption and the mean 24-HDR intakes of nutrients for quartiles of the corresponding FFQ1 measurement.

**Results**

*Representativeness of the study population*

Prevalence of tobacco use and averages of selected medical CVD risk factors were similar among participants in the sub-sample and the global study cohort. Thus, among participants in the study sub-sample, 20 and 27% of the men and women, respectively, smoked daily compared with 19 and 27% in the global cohort; average body mass index (BMI) was 25.3 and 24.9 kg m\(^{-2}\); total cholesterol 5.6 and 5.5 mmol l\(^{-1}\) compared with 6.0 and 5.9 mmol l\(^{-1}\); systolic blood pressure 129 and 126 mmHg compared with 129 and 124 mmHg, and diastolic blood pressure 80 and 78 mmHg compared with 82 and 76 mmHg.

**Reproducibility of FFQ measurements**

Table 1 shows the reproducibility of the FFQ measurements (FFQ1 vs. FFQ2) in terms of mean intake level and Pearson correlation coefficients. Mean energy and nutrient intake levels were similar for FFQ1 and FFQ2 (geometric mean ratios between 0.93 and 1.05) except for alcohol intake in women, where FFQ1 showed lower values than FFQ2. Pearson coefficients of correlation between FFQ1 and FFQ2 measurements ranged from 0.53 to 0.99, and were mostly above or equal to 0.65.

**Comparisons of levels of consumption and underreporting between FFQ1 and 24-HDR**

In both genders FFQ1 measurements of dairy products, bread/cereals, vegetables, fruits and potato/rice/pasta were higher than intakes measured by 24-HDR (geometric mean ratio > 1.10) (Table 2), whereas meat, fish, sweet snacks and alcoholic beverage intake frequencies were lower (geometric mean ratio < 0.90). Fewer respondents (three men and nine women) reported zero intake of alcohol by FFQ1 than by the ten 24-HDR (22 men and 25 women). This can be explained by the fact that the 24-HDR covered only one weekend, during which alcohol consumption is usually concentrated. Further evaluations on alcohol consumption were restricted to participants with non-zero consumption levels, i.e. comparisons of

### Table 1 Comparison of energy and nutrient intake measured by the identical, but distributed one year apart, ‘Northern Sweden’ FFQ1 and FFQ2

| Nutrient | Gender | Mean (95% CI) | Pearson correlation
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy</td>
<td>Men</td>
<td>1.01 (0.96–1.00)</td>
<td>0.67 (0.55–0.77)</td>
</tr>
<tr>
<td></td>
<td>Women</td>
<td>0.99 (0.95–1.02)</td>
<td>0.76 (0.66–0.83)</td>
</tr>
<tr>
<td>Carbohydrates</td>
<td>Men</td>
<td>1.03 (0.97–1.08)</td>
<td>0.64 (0.57–0.71)</td>
</tr>
<tr>
<td></td>
<td>Women</td>
<td>1.01 (0.96–1.05)</td>
<td>0.76 (0.66–0.83)</td>
</tr>
<tr>
<td>Sucrose</td>
<td>Men</td>
<td>1.00 (0.93–1.07)</td>
<td>0.81 (0.73–0.87)</td>
</tr>
<tr>
<td></td>
<td>Women</td>
<td>0.94 (0.88–1.01)</td>
<td>0.75 (0.65–0.82)</td>
</tr>
<tr>
<td>Fat(_\text{total})</td>
<td>Men</td>
<td>1.01 (0.95–1.07)</td>
<td>0.62 (0.48–0.73)</td>
</tr>
<tr>
<td></td>
<td>Women</td>
<td>0.98 (0.93–1.02)</td>
<td>0.71 (0.60–0.80)</td>
</tr>
<tr>
<td>Fat(_\text{saturated})</td>
<td>Men</td>
<td>1.00 (0.93–1.07)</td>
<td>0.59 (0.44–0.71)</td>
</tr>
<tr>
<td></td>
<td>Women</td>
<td>0.96 (0.92–1.01)</td>
<td>0.70 (0.58–0.79)</td>
</tr>
<tr>
<td>Cholesterol</td>
<td>Men</td>
<td>1.00 (0.93–1.07)</td>
<td>0.53 (0.37–0.66)</td>
</tr>
<tr>
<td></td>
<td>Women</td>
<td>0.98 (0.93–1.02)</td>
<td>0.68 (0.55–0.77)</td>
</tr>
<tr>
<td>Protein</td>
<td>Men</td>
<td>0.99 (0.92–1.05)</td>
<td>0.62 (0.48–0.73)</td>
</tr>
<tr>
<td></td>
<td>Women</td>
<td>0.96 (0.94–1.01)</td>
<td>0.69 (0.58–0.79)</td>
</tr>
<tr>
<td>Alcohol</td>
<td>Men</td>
<td>0.99 (0.92–1.06)</td>
<td>0.99 (0.98–0.99)</td>
</tr>
<tr>
<td></td>
<td>Women</td>
<td>0.79 (0.59–1.05)</td>
<td>0.88 (0.82–0.92)</td>
</tr>
<tr>
<td>Fibre</td>
<td>Men</td>
<td>1.03 (0.97–1.10)</td>
<td>0.70 (0.58–0.79)</td>
</tr>
<tr>
<td></td>
<td>Women</td>
<td>1.04 (0.99–1.09)</td>
<td>0.74 (0.64–0.82)</td>
</tr>
<tr>
<td>Vitamin C</td>
<td>Men</td>
<td>1.03 (0.94–1.13)</td>
<td>0.65 (0.52–0.75)</td>
</tr>
<tr>
<td></td>
<td>Women</td>
<td>1.02 (0.94–1.11)</td>
<td>0.68 (0.56–0.76)</td>
</tr>
<tr>
<td>Vitamin E</td>
<td>Men</td>
<td>1.04 (0.98–1.10)</td>
<td>0.65 (0.52–0.75)</td>
</tr>
<tr>
<td></td>
<td>Women</td>
<td>1.02 (0.98–1.07)</td>
<td>0.69 (0.57–0.78)</td>
</tr>
<tr>
<td>β-Carotene</td>
<td>Men</td>
<td>1.02 (0.91–1.13)</td>
<td>0.71 (0.60–0.80)</td>
</tr>
<tr>
<td></td>
<td>Women</td>
<td>1.05 (0.96–1.15)</td>
<td>0.84 (0.77–0.89)</td>
</tr>
<tr>
<td>Retinol</td>
<td>Men</td>
<td>1.04 (0.95–1.13)</td>
<td>0.58 (0.43–0.70)</td>
</tr>
<tr>
<td></td>
<td>Women</td>
<td>1.02 (0.95–1.09)</td>
<td>0.73 (0.60–0.81)</td>
</tr>
<tr>
<td>Calcium</td>
<td>Men</td>
<td>0.98 (0.90–1.06)</td>
<td>0.62 (0.48–0.73)</td>
</tr>
<tr>
<td></td>
<td>Women</td>
<td>0.93 (0.86–1.00)</td>
<td>0.68 (0.55–0.77)</td>
</tr>
<tr>
<td>Iron</td>
<td>Men</td>
<td>1.01 (0.95–1.06)</td>
<td>0.63 (0.49–0.74)</td>
</tr>
<tr>
<td></td>
<td>Women</td>
<td>1.01 (0.96–1.05)</td>
<td>0.68 (0.56–0.77)</td>
</tr>
</tbody>
</table>

*Mean intake values based on FFQ1 are presented in Tables 3 and 4.† Ninety-six men and 99 women were included in all calculations. The proportions reporting zero intake of alcohol were the same by FFQ1 and FFQ2 for both genders (cf. footnote in Table 3).‡ Geometric mean (95% CI) of individual ratios between intakes as recorded by FFQ1 and FFQ2.§ Pearson correlation coefficients (log-transformed data).

FFQ1 with 24-HDR. The proportions classified as underreporters were high by FFQ1 (42.7% and 49.5%, men and women, respectively) as well as 24-HDR (55.7% and 51.5%, men and women, respectively). Spearman correlation coefficients between food intake frequencies assessed by FFQ1 and by 24-HDR were between 0.42 and 0.69 for most food groups except potato/pasta, meat and fish, which had values below 0.40 (Table 2). Cross-classification into 24-HDR quartiles of participants in the highest and lowest FFQ1 quartiles, respectively, placed more than 68% of the participants into either the two lowest or the two highest 24-HDR quartiles, and for none of the food groups were there more than four participants grossly misclassified, e.g. into the most opposite quartile (data not shown). Consumption frequencies of coffee and tea were similarly measured by FFQ1 and 24-HDR regardless of preparation method (two variants of coffee preparation are dominant in the study area – extraction by filtering or by boiling), and correlations between FFQ1 and 24-HDR were between
Comparisons of energy and nutrient intakes by FFQ1 and 24-HDR

Gender-specific intakes of energy and nutrients (mean amount per day) as measured by FFQ1 and 24-HDR, as well as the mean ratio (95% confidence interval (CI)) between these measures, are shown in Table 3 for energy and macronutrients and in Table 4 for micronutrients; Tables 5 and 6 show the corresponding results for nutrient densities. Mean intake measurements by FFQ1 and by 24-HDR were similar (geometric mean ratios FFQ1/24-HDR = 0.90–1.10) for energy and most nutrients (Tables 3 and 4), with the exception of sucrose and cholesterol (FFQ1/24-HDR < 0.90) and fibre, vitamin C, β-carotene and retinol (FFQ1/24-HDR > 1.10). The pattern was very similar for nutrient densities (Tables 5 and 6).

Pearson correlation coefficients between FFQ1 and individuals’ average 24-HDR measurements (rQR) ranged from 0.61 (total fat in men, fibre in women) to 0.28 (retinol in women) (Tables 3 and 4). Half of the correlation coefficients were equal to or greater than 0.50, 12 values were between 0.40 and 0.49, and four values (for retinol in both genders, sucrose and cholesterol in women) were below 0.35. The low correlation for retinol (vitamin A) can be explained by the fact that, to a large extent, average intake levels are determined by the irregular consumption of foods (e.g. liver) extremely rich in this nutrient, which may not have been captured accurately by only ten 24-HDR per subject27. Cross-classification of the participants in the lowest or highest FFQ1 quartiles placed more than 70% into the two lowest or two highest 24-HDR quartiles, respectively. Extreme misclassification between the top and bottom quartiles was seen in no more than one to three participants for each nutrient. The 24-HDR measurements increased in a monotonic manner over quartiles of the corresponding FFQ1 measurements (Tables 2).

The ratio of within- to between-subject variation in the ten 24-hour diet recalls ranged from 1.05 to 1.21. This elevated within-subject variation in the 24-HDR measurements compared with the between-subject variance indicated that residual random day-to-day variations in the 24-HDR attenuated the associations between FFQ1 and the mean 24-HDR measures. Correction for this attenuation effect notably improved the Pearson correlations (pqr) between FFQ1 and 24-HDR for alcohol by a factor of 1.26 and 1.36, respectively, for men and women, to 0.60 and 0.79. For the other nutrients, the attenuation correction led to much smaller improvements (Tables 3 and 4). For nutrient densities, the correlations between FFQ1 and 24-HDR measurements were similar to those for

Please note that the table in the image is not clearly visible, but the text provides the necessary context and explanations.
the absolute intake estimates, and correction for attenuation due to variations over time in the 24-HDR also led to a similar degree of improvement in these correlation coefficients.

Among those participants who had provided a blood sample, but after exclusion of one vegetarian subject with extremely elevated intakes as well as serum values of β-carotene, plasma β-carotene levels had a Pearson coefficient of correlation of 0.47 with the 24-HDR measurements, and of 0.23 with FFQ1 measurements.

**Calibration of nutrient intake levels**

For all nutrients except alcohol, the linear regression of 24-HDR on FFQ1 resulted in slope estimates (linear calibration coefficients) between 0.50 for protein (men and women) and 0.59 for vitamin C (men) (Tables 3 and 4). For most of the nutrient densities, the slopes were of similar average magnitude, except for cholesterol (men), retinol (men) and iron (women), for which the calibration coefficients were lower (Tables 5 and 6). Only for alcohol intake were the regression slopes very close to 1.0.

**Discussion**

In the present study, we examined the validity of dietary questionnaire measurements in a large prospective cohort study of nutrition and chronic disease risk in the north of Sweden, and used 24-hour diet recalls to calibrate the FFQ measurements.

The validity of food-frequency questionnaire measurements generally cannot be assumed to be independent of the population or the specific context in which the measurements were collected. For example, the correlation between FFQ and true habitual intake levels (as well as reference measurements, e.g. average 24-HDR) would generally tend to improve if the measurements are collected in a population with greater between-subject heterogeneity in true habitual intake levels and, inversely, would be lower in populations with more homogeneous dietary intake patterns. Furthermore, the magnitude of systematic and random errors may depend on the quality of response to the questionnaire, and hence on the interest and backgrounds of participants studied. Estimates of validity therefore should not be extrapolated beyond the actual populations in which measurements were collected and validated; in the present case, the Northern Sweden EPIC/MONICA cohort. A corollary of this is that validation/calibration sub-studies should be conducted on a subset of participants that is representative for the global study cohort. In the present study, such representativeness was high, as the validation study was based on a random selection of cohort participants, with a high participation rate. Representativeness was further confirmed by the great similarity between values for tobacco use, BMI, blood pressures and blood cholesterol in the validation study subgroup and the rest of the cohort.

One of the characteristics of the Northern Sweden EPIC/MONICA cohort is the repeated measurement of dietary intakes and other lifestyle variables, over 10-year intervals\(^2\). Thus, not only will it be possible to relate the

### Table 3 Comparison of daily energy and macronutrient intake as measured by the ‘Northern Sweden’ FFQ and 24-hour dietary recalls (24-HDR)

<table>
<thead>
<tr>
<th>Nutrient</th>
<th>Gender</th>
<th>Mean intake†</th>
<th>Ratio</th>
<th>Pearson correlation§</th>
<th>Means by FFQ1 quartiles¶</th>
<th>Calibration coefficient‖</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>FFQ1</td>
<td>24-HDR†</td>
<td>FFQ1</td>
<td>24-HDR</td>
</tr>
<tr>
<td>Energy (kcal)</td>
<td>Men</td>
<td>2018</td>
<td>2108</td>
<td>1.06 (0.90–1.02)</td>
<td>0.50 (0.34–0.64)</td>
<td>0.53</td>
</tr>
<tr>
<td></td>
<td>Women</td>
<td>1609</td>
<td>1588</td>
<td>1.01 (0.96–1.07)</td>
<td>0.45 (0.27–0.59)</td>
<td>0.48</td>
</tr>
<tr>
<td>Carbohydrate (g)</td>
<td>Men</td>
<td>244.2</td>
<td>240.8</td>
<td>1.01 (0.95–1.09)</td>
<td>0.44 (0.26–0.59)</td>
<td>0.46</td>
</tr>
<tr>
<td></td>
<td>Women</td>
<td>206.1</td>
<td>208.0</td>
<td>1.02 (0.97–1.08)</td>
<td>0.51 (0.35–0.64)</td>
<td>0.54</td>
</tr>
<tr>
<td>Sucrose (g)</td>
<td>Men</td>
<td>30.9</td>
<td>40.1</td>
<td>0.77 (0.69–0.86)</td>
<td>0.60 (0.45–0.72)</td>
<td>0.65</td>
</tr>
<tr>
<td></td>
<td>Women</td>
<td>23.3</td>
<td>34.0</td>
<td>0.69 (0.62–0.76)</td>
<td>0.34 (0.15–0.50)</td>
<td>0.37</td>
</tr>
<tr>
<td>Fat total (g)</td>
<td>Men</td>
<td>75.5</td>
<td>81.1</td>
<td>0.93 (0.87–0.99)</td>
<td>0.61 (0.47–0.72)</td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td>Women</td>
<td>56.0</td>
<td>61.8</td>
<td>0.91 (0.85–0.96)</td>
<td>0.54 (0.38–0.66)</td>
<td>0.59</td>
</tr>
<tr>
<td>Fat saturated (g)</td>
<td>Men</td>
<td>31.6</td>
<td>34.8</td>
<td>0.91 (0.84–0.97)</td>
<td>0.58 (0.43–0.70)</td>
<td>0.62</td>
</tr>
<tr>
<td></td>
<td>Women</td>
<td>23.4</td>
<td>26.6</td>
<td>0.88 (0.83–0.94)</td>
<td>0.54 (0.38–0.66)</td>
<td>0.59</td>
</tr>
<tr>
<td>Cholesterol (mg)</td>
<td>Men</td>
<td>0.25</td>
<td>0.32</td>
<td>0.77 (0.71–0.83)</td>
<td>0.45 (0.27–0.60)</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td>Women</td>
<td>0.19</td>
<td>0.24</td>
<td>0.77 (0.72–0.82)</td>
<td>0.31 (0.12–0.48)</td>
<td>0.37</td>
</tr>
<tr>
<td>Protein (g)</td>
<td>Men</td>
<td>74.7</td>
<td>82.2</td>
<td>0.91 (0.85–0.97)</td>
<td>0.41 (0.23–0.56)</td>
<td>0.44</td>
</tr>
<tr>
<td></td>
<td>Women</td>
<td>59.4</td>
<td>59.9</td>
<td>0.99 (0.94–1.05)</td>
<td>0.40 (0.23–0.56)</td>
<td>0.44</td>
</tr>
<tr>
<td>Alcohol (g)</td>
<td>Men</td>
<td>4.53</td>
<td>6.36</td>
<td>0.94 (0.70–1.26)</td>
<td>0.49 (0.29–0.65)</td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td>Women</td>
<td>2.51</td>
<td>3.13</td>
<td>0.92 (0.73–1.14)</td>
<td>0.58 (0.41–0.71)</td>
<td>0.79</td>
</tr>
</tbody>
</table>

* Ninety-six men and 99 women are included in calculations of nutrients. For alcohol, zero reporters (three men and nine women for FFQ1; 22 men and 25 women for 24-HDR) were excluded.
† Geometric mean intakes (amount per day) obtained by FFQ1 (mean\(_{\text{FFQ1}}\)) or 24-HDR (mean\(_{\text{24-HDR}}\)).
‡ Pearson correlation coefficients – crude coefficients (\(r_{\text{Q1}}\)) and after de-attenuation (\(r_{\text{Pot}}\)) – between intakes measured by FFQ1 and 24-HDR.
§ Geometric mean of 24-HDR measured intake (amount per day) in FFQ1 ranked quartiles.
¶ Calibration coefficient equalling the slope (\(\lambda\), with 95% CI) from the linear regression of the 24-hour recall measured intake on the corresponding intake measured by FFQ1.
dietary intake assessments to subsequent disease risk, but the replicate data can also be used to monitor changes of diet over more prolonged time periods. The reproducibility of the FFQ data over a one-year period was high, in terms of both estimated mean intake levels and relative classification of participants from low to high levels. Thus, although mean intake estimates from the FFQ may not be absolutely valid (in fact, there is clear underreporting for at least part of the participants), the differences in mean FFQ intake levels over time may be indicative of true changes in food consumption behaviour. This justifies the use of the FFQ measurements for monitoring of changes that may be induced by the local Västerbotten CVD community intervention programme. Mean intake levels are also in reasonable accordance with those of other published studies, including the European EPIC study, of which the NSHD cohort is a part.

Since perfectly error-free reference measurements for diet do not exist in practice, the validation and calibration of FFQ measurements depend crucially on statistical

### Table 4
Comparison of daily micronutrient intake as measured by the ‘Northern Sweden’ FFQ and 24-hour dietary recalls (24-HDR)

<table>
<thead>
<tr>
<th>Nutrient</th>
<th>Gender*</th>
<th>Mean FFQ Mean24-HDR†</th>
<th>Intake (95% CI)</th>
<th>Ratio FFQ1/24-HDR‡</th>
<th>Pearson correlation§</th>
<th>Means by FFQ1 quartiles¶</th>
<th>Calibration coefficient‖</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fibre (g)</td>
<td>Men</td>
<td>19.5 16.8</td>
<td>1.19 (1.10–1.29)</td>
<td>0.48 (0.31–0.62)</td>
<td>0.61 (0.47–0.72)</td>
<td>15.4 15.2 17.4 20.3</td>
<td>0.39 (0.25–0.53)</td>
</tr>
<tr>
<td>Vitamin C (mg)</td>
<td>Men</td>
<td>69.3 50.6</td>
<td>1.37 (1.22–1.53)</td>
<td>0.52 (0.36–0.66)</td>
<td>0.61 (0.47–0.72)</td>
<td>32.7 46.7 63.0 84.8</td>
<td>0.59 (0.39–0.79)</td>
</tr>
<tr>
<td>Vitamin E (mg)</td>
<td>Men</td>
<td>85.5 60.0</td>
<td>1.42 (1.28–1.58)</td>
<td>0.51 (0.35–0.64)</td>
<td>0.61 (0.47–0.72)</td>
<td>39.3 59.0 65.9 86.2</td>
<td>0.52 (0.34–0.70)</td>
</tr>
<tr>
<td>β-Carotene (mg)</td>
<td>Men</td>
<td>6.76 6.80</td>
<td>0.93 (0.90–1.00)</td>
<td>0.45 (0.45–0.71)</td>
<td>0.61 (0.47–0.72)</td>
<td>5.34 6.05 7.56 8.75</td>
<td>0.54 (0.38–0.70)</td>
</tr>
<tr>
<td>Retinol (mg)</td>
<td>Men</td>
<td>1.19 0.82</td>
<td>1.46 (1.30–1.64)</td>
<td>0.32 (0.13–0.49)</td>
<td>0.61 (0.47–0.72)</td>
<td>0.65 0.84 0.71 1.13</td>
<td>0.39 (0.15–0.63)</td>
</tr>
<tr>
<td>Vitamin C (mg)</td>
<td>Women</td>
<td>12.5 11.8</td>
<td>1.06 (1.00–1.12)</td>
<td>0.43 (0.24–0.57)</td>
<td>0.61 (0.47–0.72)</td>
<td>9.9 11.9 12.0 13.8</td>
<td>0.48 (0.23–0.56)</td>
</tr>
</tbody>
</table>

* Ninety-six men and 99 women are included in calculations of nutrients. For alcohol, zero reporters (three men and nine women for FFQ1; 22 men and 25 women for 24-HDR) were excluded.
† Geometric mean (95% CI) of individual ratios between intake (amount per day) as obtained by FFQ1 and 24-HDR.
‡ Geometric mean (95% CI) of individual ratios between intake (amount per 1000 kcal) as obtained by FFQ1 and 24-HDR.
§ Pearson correlation coefficients – crude coefficients (rQR) and after de-attenuation (pQT) – between intakes measured by FFQ1 and 24-HDR.
¶ Calibration coefficient equalling the slope (λ, with 95% CI) from the linear regression of the 24-hour recall measured intake on the corresponding intake measured by FFQ1.
‖ Geometric mean intakes (amount per day) obtained by FFQ1 (meanFFQ) or 24-HDR (mean24-HDR).

### Table 5
Comparison of macronutrient intake (density as amount per 1000 kcal) as measured by the ‘Northern Sweden’ FFQ and 24-hour dietary recalls (24-HDR)

<table>
<thead>
<tr>
<th>Nutrient</th>
<th>Gender*</th>
<th>Mean FFQ Mean24-HDR†</th>
<th>Intake (95% CI)</th>
<th>Ratio FFQ1/24-HDR‡</th>
<th>Pearson correlation§</th>
<th>Means by FFQ1 quartiles¶</th>
<th>Calibration coefficient‖</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carbohydrate (g)</td>
<td>Men</td>
<td>121 114</td>
<td>1.06 (1.04–1.08)</td>
<td>0.57 (0.41–0.69)</td>
<td>0.64 113 119 124</td>
<td>0.60 (0.42–0.78)</td>
<td></td>
</tr>
<tr>
<td>Sucrose (g)</td>
<td>Men</td>
<td>15.3 19.0</td>
<td>0.80 (0.75–0.87)</td>
<td>0.65 (0.51–0.75)</td>
<td>0.71 13.9 18.0 21.7 26.0</td>
<td>0.58 (0.44–0.72)</td>
<td></td>
</tr>
<tr>
<td>Fattotal (g)</td>
<td>Men</td>
<td>37.4 38.5</td>
<td>0.97 (0.94–1.00)</td>
<td>0.42 (0.24–0.57)</td>
<td>0.51 35.1 39.0 41.0 40.8</td>
<td>0.38 (0.21–0.55)</td>
<td></td>
</tr>
<tr>
<td>Fatunsaturated (g)</td>
<td>Men</td>
<td>34.8 38.9</td>
<td>0.89 (0.86–0.92)</td>
<td>0.57 (0.41–0.69)</td>
<td>0.70 35.8 38.2 40.5 41.6</td>
<td>0.34 (0.24–0.44)</td>
<td></td>
</tr>
<tr>
<td>Cholesterol (mg)</td>
<td>Men</td>
<td>15.6 16.5</td>
<td>0.95 (0.91–0.98)</td>
<td>0.52 (0.36–0.66)</td>
<td>0.58 14.5 16.5 17.2 18.3</td>
<td>0.47 (0.32–0.62)</td>
<td></td>
</tr>
<tr>
<td>Protein (g)</td>
<td>Men</td>
<td>37.0 39.0</td>
<td>0.95 (0.91–0.98)</td>
<td>0.35 (0.26–0.46)</td>
<td>0.50 37.6 37.9 37.4 43.7</td>
<td>0.41 (0.19–0.63)</td>
<td></td>
</tr>
</tbody>
</table>

* Ninety-six men and 99 women are included in calculations of nutrients. For alcohol, zero reporters (three men and nine women for FFQ1; 22 men and 25 women for 24-HDR) were excluded.
† Geometric mean intakes (amount per 1000 kcal) obtained by FFQ1 (meanFFQ) or 24-HDR (mean24-HDR).
‡ Geometric mean (95% CI) of individual ratios between intake (amount per 1000 kcal) as obtained by FFQ1 and 24-HDR.
§ Pearson correlation coefficients – crude coefficients (rQR) and after de-attenuation (pQT) – between intakes adjusted for energy (residuals) measured by FFQ1 and 24-HDR.
¶ Calibration coefficient equalling the slope (λ, with 95% CI) from the linear regression of the 24-HDR measured intake on the corresponding intake measured by FFQ1.
‖ Geometric mean of 24-HDR measured intake (amount per day) as obtained by FFQ1 ranked quartiles.
model assumptions\textsuperscript{15,28}. One assumption generally made is that at the population level there is a linear relationship between dietary intake assessments obtained by FFQ (Q) or by 24-HDR (R) with individuals’ true habitual intake levels\textsuperscript{29}. Then, to estimate the correlation $r_{QR}$ between individuals’ FFQ measurements and their true intake levels, it can be shown that at least two additional measurements, $X_1$ and $X_2$, are required, for which the random errors ($e_{X_1}$ and $e_{X_2}$) can be assumed to be uncorrelated both between themselves and with random errors of the FFQ measurements\textsuperscript{29}. In practice, it is generally felt that the best possible attempt to meet these requirements is to compare with replicate recordings (i.e. $X_1$, $X_2$, ... $R_1$, $R_2$, ...) of actual food consumption, obtained by food consumption diaries, weighed food consumption records or a 24-HDR. If indeed it can be assumed that the random errors $e_{R_1}$, $e_{R_2}$, ... $e_{R_{10}}$ of the ten replicate 24-HDR measurements are uncorrelated between themselves and with the random errors $e_{Q}$ in the FFQ measurements, the de-attenuated correlation coefficients in Tables 3--6 (indicated by $r_{QT}$) would provide an unbiased estimate of the correlation between the FFQ and individuals’ true habitual intake estimates. We decided to use replicate 24-HDR measurements, because the method generally has high response rates (relevant for the representativeness of the validation sub-study) and because it could be administered by telephone for participants living far from the research centre.

Unfortunately, it is questionable whether the assumption of uncorrelatedness of random errors $e_{R_1}$, $e_{R_2}$, $e_{Q}$ holds in practice. Many studies have shown that 24-HDR data, but also weighed food records or FFQ measurements, may be systematically underestimated, and that participants may vary systematically in their tendency to underestimate by each of these methods\textsuperscript{22,28}. As explained in greater detail elsewhere\textsuperscript{30}, the consequence of this is that the de-attenuated correlation coefficients of Tables 3--6 can be either an underestimation or an overestimation of the correlation of FFQ measurements with true intake levels, and it is unclear which of these two biases – upwards or downwards – would predominate in practice.

It is often assumed that correlations between random errors of FFQ and 24-HDR measurements would be mostly a consequence of individuals’ systematic differences in tendency to globally underreport intake levels, irrespective of the type of assessment method used. In this context, a crucial question is whether such underreporting is generally of an approximately equal magnitude for any type of food consumed – that is, whether or not the magnitude of underreporting is food- or nutrient-specific. If the latter can be assumed to be true, then the correction for total energy intake level, e.g. by calculating nutrient densities, would strongly diminish the error correlations and increase the validity of validation and calibration study results. This type of assumption deserves further investigation.

An approach closely related to validation is the calibration of FFQ measurements. Calibration can be defined as a scale adjustment to the FFQ measurements, such that after the adjustments estimates of relative risk corresponding to a measured absolute intake level are approximately unbiased\textsuperscript{24,25}. We estimated the calibration factor $\lambda$ by linear regression of 24-HDR on the FFQ measurements and, as shown in Tables 2--6, the estimated

### Table 6 Comparison of micronutrient intake (density as amount per 1000 kcal) as measured by the ‘Northern Sweden’ FFQ and 24-hour dietary recalls (24-HDR)

<table>
<thead>
<tr>
<th>Nutrient</th>
<th>Gender*</th>
<th>Mean intake†</th>
<th>Ratio FFQ1/24-HDR‡</th>
<th>Mean Pearson correlation§</th>
<th>Means by FFQ1 quartiles¶</th>
<th>Calibration coefficient‖</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>MeanFFQ</td>
<td>Mean24-HDR</td>
<td>Intake (95% CI)</td>
<td>$r_{QR}$</td>
<td>$p_{QT}$</td>
</tr>
<tr>
<td>Fibre (g)</td>
<td>Men</td>
<td>9.7</td>
<td>7.8</td>
<td>1.24 (1.19–1.30)</td>
<td>0.58 (0.43–0.70)</td>
<td>0.63</td>
</tr>
<tr>
<td></td>
<td>Women</td>
<td>11.2</td>
<td>9.1</td>
<td>1.23 (1.19–1.27)</td>
<td>0.70 (0.58–0.79)</td>
<td>0.79</td>
</tr>
<tr>
<td>Vitamin C (mg)</td>
<td>Men</td>
<td>34.3</td>
<td>24.0</td>
<td>1.43 (1.29–1.58)</td>
<td>0.47 (0.29–0.61)</td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td>Women</td>
<td>53.1</td>
<td>37.6</td>
<td>1.41 (1.29–1.54)</td>
<td>0.54 (0.39–0.67)</td>
<td>0.62</td>
</tr>
<tr>
<td>Vitamin E (mg)</td>
<td>Men</td>
<td>3.4</td>
<td>3.2</td>
<td>1.04 (0.99–1.09)</td>
<td>0.26 (0.06–0.44)</td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td>Women</td>
<td>3.7</td>
<td>3.8</td>
<td>0.97 (0.93–1.02)</td>
<td>0.35 (0.17–0.52)</td>
<td>0.42</td>
</tr>
<tr>
<td>$\beta$-Carotene (mg)</td>
<td>Men</td>
<td>0.9</td>
<td>9.7</td>
<td>1.29 (1.10–1.50)</td>
<td>0.43 (0.25–0.58)</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td>Women</td>
<td>2.2</td>
<td>1.4</td>
<td>1.58 (1.37–1.82)</td>
<td>0.58 (0.43–0.70)</td>
<td>0.72</td>
</tr>
<tr>
<td>Retinol (mg)</td>
<td>Men</td>
<td>0.6</td>
<td>0.4</td>
<td>1.53 (1.37–1.63)</td>
<td>0.22 (0.02–0.41)</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>Women</td>
<td>0.5</td>
<td>0.4</td>
<td>1.26 (1.15–1.39)</td>
<td>0.29 (0.10–0.46)</td>
<td>0.42</td>
</tr>
<tr>
<td>Calcium (mg)</td>
<td>Men</td>
<td>467</td>
<td>487</td>
<td>0.96 (0.91–1.01)</td>
<td>0.50 (0.34–0.64)</td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td>Women</td>
<td>473</td>
<td>502</td>
<td>0.94 (0.89–1.00)</td>
<td>0.59 (0.45–0.71)</td>
<td>0.66</td>
</tr>
<tr>
<td>Iron (mg)</td>
<td>Men</td>
<td>7.8</td>
<td>7.4</td>
<td>1.05 (1.00–1.09)</td>
<td>0.02 (0.18–0.22)</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>Women</td>
<td>7.8</td>
<td>7.4</td>
<td>1.05 (1.00–1.09)</td>
<td>0.02 (0.18–0.22)</td>
<td>0.03</td>
</tr>
</tbody>
</table>

* Ninety-six men and 99 women are included in calculations of nutrients. For alcohol, zero reporters (three men and nine women for FFQ1; 22 men and 25 women for 24-HDR) were excluded.
† Geometric mean intakes (amount per 1000 kcal) obtained by FFQ1 (meanFFQ) or 24-HDR (mean24-HDR). ‡ Geometric mean intakes (amount per 1000 kcal) obtained by FFQ1 (meanFFQ) or 24-HDR (mean24-HDR). § Pearson correlation coefficients – crude coefficients ($r_{QR}$) and after de-attenuation ($p_{QT}$) – between intakes adjusted for energy (residuals) measured by FFQ1 and 24-HDR. ¶ Geometric mean of 24-hour recall measured intake (amount per 1000 kcal) in FFQ1 ranked quartiles. \(\lambda\) Calibration coefficient equaling the slope $\lambda$ of demographic/other factors on the linear regression of the 24-HDR measured intake on the corresponding intake measured by FFQ1.
Validation and calibration of FFQ measures in Sweden

calibration factors generally have values below 1.0, reflecting a large degree of attenuation bias due to random errors in the FFQ measurements\(^{24,25}\). As discussed elsewhere\(^{24,25}\), these estimated calibration factors may be used to correct relative risks estimated for quantitative differences in food or nutrient intake levels, as measured by the FFQ. However, the validity of this correction procedure again requires the assumption that random errors of the 24-HDR (\(e_R\)) and random errors in the FFQ measurements (\(e_Q\)) are uncorrelated. In addition, it must be assumed that the 24-HDR measurements have no systematic scaling biases; i.e. they must be assumed to provide accurate mean intake measurements for any randomly chosen subgroup of the total study population. In practice, as mentioned, it is likely that random errors \(e_R\) and \(e_Q\) are positively correlated, and this would lead to an overestimation of the calibration factor \(\lambda\) and hence to an underestimation of regression dilution bias\(^{31}\). As mentioned earlier, the problem of correlated random errors might be alleviated by using energy-adjusted intake measurements such as nutrient densities.

In conclusion, The Northern Sweden FFQ measurements have good reproducibility and an estimated level of validity similar to that of FFQ measurements in other prospective cohort studies. The results from this study will form the basis for the correction of attenuation and regression dilution biases in relative risk estimates, in future studies relating FFQ measurements to disease outcomes in the Northern Sweden cohort. However, in view of possible biases in the estimates of \(\rho_{QT}\) and calibrations factors \(\lambda\), due to correlation of errors in FFQ measurements and reference measurements, such corrections must be used cautiously, and may be more easily applicable for nutrient densities than for absolute nutrient intakes.

Acknowledgements

This study was supported by the Swedish Cancer Society and the Swedish Research Council.

References


https://doi.org/10.1079/PHN2001315 Published online by Cambridge University Press


