Validation of a self-administered food-frequency questionnaire administered in the European Prospective Investigation into Cancer and Nutrition (EPIC) Study: comparison of energy, protein, and macronutrient intakes estimated with the doubly labeled water, urinary nitrogen, and repeated 24-h dietary recall methods

Anja Kroke, Kerstin Klipstein-Grobusch, Susanne Voss, Jutta Möseneder, Frank Thielecke, Rudolf Noack, and Heiner Boeing

ABSTRACT

Background: The validation of dietary assessment instruments is critical in the evaluation of diet as a chronic disease risk factor. Objective: The objective was to assess the validity of a self-administered food-frequency questionnaire by comparison with dietary recall, urinary nitrogen excretion, and total energy expenditure data.

Design: Over a 1-y period, data from twelve 24-h dietary recalls, a food-frequency questionnaire, and four 24-h urine samples were obtained from 134 study participants of the European Prospective Investigation into Cancer and Nutrition (EPIC) Study in Potsdam, Germany. In a substudy of 28 participants, total energy expenditure from doubly labeled water measurements was assessed.

Results: Energy-adjusted, deattenuated correlation coefficients between the questionnaire and the recalls ranged from 0.54 for dietary fiber to 0.86 for alcohol. Cross-classification of quintiles of nutrient intakes from the questionnaire and recalls indicated severe misclassification to be <4%. Reported protein intake correlated with estimated protein excretion (r = 0.46). Energy intake and total energy expenditure were also significantly correlated (r = 0.48); however, all but one subject underreported their energy intake. The magnitude of underreporting varied considerably, by 22% on average, and increased slightly with increasing energy intake. A similar pattern of underreporting was observed when energy intakes from the 24-h dietary recalls were compared with total energy expenditure.

Conclusions: These data indicate an acceptable relative validity of the food-frequency questionnaire in this study population. Compared with measurements of total energy expenditure and protein excretion, however, only moderate agreement with both the food-frequency questionnaire and the 24-h dietary recalls was observed. Am J Clin Nutr 1999;70:439–47.

KEY WORDS Validity, biomarkers, energy expenditure, epidemiology, dietary recalls, food-frequency questionnaire, doubly labeled water, FFQ, nutrition, EPIC Study, Germany, humans

INTRODUCTION

In epidemiologic studies, dietary assessments of the relation between diet and disease most frequently rely on self-administered, quantitative food-frequency questionnaires (FFQs) aiming at the assessment of usual long-term dietary intake and designed to rank subjects into quintiles of dietary intake. The relative validity of FFQs is usually assessed by comparing their data with those of a reference method, eg, repeated 24-h dietary recalls or weighed food records, because a gold standard as a “truth reference” is not available. Most validation studies have reported a strong correlation between estimates from the 2 methods and a generally accepted amount of misclassification when subjects were cross-classified into quintiles of the distribution of each nutrient under investigation (1–4). The underlying assumption of this validation approach requires that the measurement error of the test method be independent of the measurement error of the reference method (5). Several studies, however, have reported that both methods are affected by underreporting, a frequently observed phenomenon in dietary assessment methods that refers to the discrepancy between reported energy intakes and energy requirements (6). Therefore, biomarkers of dietary intake have become increasingly important for an unbiased ascertainment of nutrient intake in nutritional validation studies (7, 8) because their measurement error is uncorrelated with the errors of the test method (5, 9, 10).

A biomarker frequently applied in nutritional validation studies is the measurement of urinary nitrogen excretion, which allows the determination of protein intake (11, 12). The doubly labeled water (DLW) method, developed by Lifson et al (13) in the early 1950s and first used in humans by Schoeller and van Santen (14) in 1982, is commonly accepted as the only method that provides valid estimates of energy expenditure in free-living subjects.
The present validation study was conducted during the baseline recruitment phase of the European Prospective Investigation into Cancer and Nutrition Study at the Potsdam, Germany, study center (EPIC-Potsdam Study). The FFQ used in the baseline examination of the EPIC-Potsdam Study was validated previously for reproducibility and relative validity in a West German study population (15–17). Because the performance of a dietary assessment instrument depends on characteristics of the study population, additional validation studies are recommended when a previously validated instrument is used in another population (5, 10). The 2 aims of this validation study were therefore to investigate the validity of the FFQ against biomarkers as reference methods and to obtain data on relative validity by comparing FFQ data with data from 24-h dietary recalls, thereby determining the validity of the FFQ in this East German study population.

SUBJECTS AND METHODS

Background

EPIC is a multicenter cohort study conducted in 9 European countries (18, 19). The cohort from the EPIC-Potsdam Study, 1 of 2 German cohorts, will contribute ≈27,000 subjects to the overall cohort size of ≈475,000. Study participants were randomly selected from population registries in Potsdam and adjacent communities. Approval for all study procedures, including those of the validation study, was given by the Ethical Committee of the State of Brandenburg, Germany, and informed consent was obtained from all study participants.

Subjects

The 160 subjects who volunteered to participate in the validation study were recruited from participants of the baseline examination of the EPIC-Potsdam Study. Women between 35 and 66 y of age and men between 40 and 67 y of age were divided into 3 age- and sex-specific strata according to the EPIC protocol (18). Only those 134 subjects who completed at least ten 24-h dietary recalls and at least three 24-h urine collections [≥85% p-aminobenzoic acid (PABA) recovery] were included in the present analysis. In a subset (15 men and 15 women) of the validation study, total energy expenditure (TEE) was determined with the DLW method. Two persons did not complete the required measurements and were therefore excluded from this analysis.

Dietary assessment

Food-frequency questionnaire

A self-administered, scanner-readable FFQ was the basic instrument used to assess habitual food intake in this cohort study. Development, reproducibility, and relative validity of the FFQ were described in detail previously (15–17). The 146-item FFQ included questions about specific food items, such as the frequency of sauce consumption, the fat content of several food items, and seasonal consumption of fruit and vegetables. For certain food items, portion sizes were estimated on the basis of colored photographs of portion sizes; otherwise, standard portion sizes such as a cup (150 mL), jars, and pieces were used. Questions about general patterns of consumption of the main food groups (bread, luncheon meat and cheese on bread, meat, fruit, and vegetables) at the end of the FFQ were used to adjust the consumption patterns that were reported when several single food items from these food groups were requested. After the FFQ was optically read, a software program was used to check the questionnaire data for completeness and plausibility. Missing or implausible information was then corrected with the participant, resulting in complete and consistent dietary data. The FFQ was administered at the end of the validation study period, thereby covering the 1-y study phase.

Twenty-four–hour dietary recall

Three computer-assisted 24-h dietary recall interviews (EPIC-SOFT, German version; developed in collaboration with the International Agency of Research on Cancer, Lyon, France) were conducted per season during the 1-y study period by one trained interviewer. This software program was specifically developed for the standardized assessment of foods and recipes consumed on the preceding day and was adapted to the specific dietary habits of the EPIC-Potsdam Study population (20). The interviews were meal-sequence based and involved a detailed assessment and description of the food consumed. Colored photographs of different portion sizes of foods were provided to help estimate the quantity of food consumed. Interviews were conducted on all weekdays except Fridays because of organizational restrictions. Recalls of food consumed on Saturdays and Sundays were obtained on Mondays.

Nutrient database

Energy and nutrient intakes from both dietary assessments were calculated by using data from the German Food Code (21).

Lifestyle factors and anthropometry

Information on other lifestyle factors, such as smoking habits and physical activity, were obtained by conducting computer-assisted person-to-person interviews. Anthropometric measures were obtained by trained and quality-monitored personnel (22) while subjects wore no shoes and only light underwear. Body weight was measured on a digital scale to the nearest 100 g; body height was measured with a flexible anthropometer to the nearest 0.1 cm.

Relative energy intake

For all subjects in the validation study, basal metabolic rate (BMR) was calculated based on the formulas of Schofield et al (23) from weight and age. Energy intake reported on the FFQ was then divided by the estimated BMR, thereby obtaining a measure of relative energy intake that accounted for age and body weight (24).

Doubly labeled water method

The DLW method was used to assess the carbon dioxide production values needed to calculate TEE with the classic equations of Coward and Cole (25). This method is based on the differential disappearance rates of the stable isotopes 2H and 18O. We applied the multipoint approach (26) over a time period of 14 d after the oral administration of an individually calculated amount of doubly labeled water per kilogram body weight (27). The isotopes disperse throughout the body water and gradually leave the body over subsequent days. Urine samples from each day of the sampling period were stored by the participants at −20°C. When the sampling was completed, the urine samples were stored at −80°C until analyzed by mass spectrometry with an automated equilibration device. The CV for the applied analytic method was 3%; further details of the isotope analysis were described previously (28).
Twenty-four–hour urinary nitrogen

Four 24-h urine samples were collected from each participant while PABA was administered simultaneously to assess the completeness of urine collection (29). Recovery of <85% PABA indicated incomplete urine collection (30, 31) and resulted in exclusion of 93 of 640 urine samples. Urea nitrogen excretion was measured by spectrophotometry and converted to urea nitrogen excretion in g/d. Assuming that urea nitrogen excretion is a constant proportion (85%) of total urinary nitrogen (12), protein intake was derived from the formula

\[
\text{protein intake (g/d)} = \text{urinary nitrogen (g/d)} 
\]

Statistical analysis

Nutrients not normally distributed were log-transformed to achieve normal distributions, which were obtained for all nutrients except alcohol. Means and SDs of energy, protein, and macronutrient intakes were determined for the test (FFQ) and reference (24-h dietary recall, urinary nitrogen, and DLW) methods; differences and ratios between means were calculated with the FFQ and with the reference methods. A paired-difference t test and Wilcoxon’s signed-rank test for alcohol were used to determine significant differences between means. The variance ratios of the repeated 24-h dietary recalls and the urinary nitrogen measurements were determined by dividing within-person variations by between-person variations. Pearson’s and Spearman’s correlation coefficients, respectively, were obtained for crude and deattenuated data as well as for energy-adjusted crude and deattenuated data. Deattenuation to correct for intrasubject variability was accomplished by using the formula suggested by Beaton et al (32). Values were adjusted for energy by using the residual method (33).

To assess agreement between the FFQ and reference methods and to detect any bias with the test method relative to the reference method, differences between the 2 respective methods were plotted against the means, as suggested by Bland and Altman (34). To compare the FFQ with the multiple 24-h dietary recall, study participants were classified into quintiles of intake. Proportions of subjects classified into the same, adjacent, or extreme quintiles were derived for both crude and energy-adjusted cross-classifications. To compare the test method with the urinary nitrogen and DLW methods, study participants were classified into quartiles of energy and protein intakes according to the distribution of the test and the respective reference methods. Proportions of subjects classified into the same, adjacent, or extreme quartiles were determined. Results were considered statistically significant at a two-tailed α level of 0.05. Statistical analysis was performed by using SAS software (release 6.12; SAS Institute Inc, Cary, NC).

RESULTS

General characteristics of the study population are presented in Table 1. Mean ages, body weights, and energy intakes relative to estimated BMRs were significantly different between men and women. Seventeen percent of the men and 3% of the women were current smokers. Regular sporting activities were reported by 39% of the men and by 44% of the women. Mean energy and macronutrient intakes estimated with the FFQ and 24-h dietary recalls, mean differences between these 2 methods, mean ratios of the 2 methods, and variance ratios for the 24-h dietary recall data are presented in Table 2. Intakes of energy, total protein, and total carbohydrate were significantly higher with the FFQ than with the 24-h dietary recalls, whereas intakes of total fat, saturated fat, monounsaturated fat, cholesterol, and monosaccharides were lower with the FFQ than with the 24-h dietary recall. For several nutrients, there was considerably more within-person variation than between-person variation as indicated by the variance ratios for 24-h dietary recall data. Therefore, crude data as well as correlation coefficients deattenuated for within-person variation and energy-adjusted crude and deattenuated correlation coefficients are presented in Table 2. Pearson’s correlation coefficients for crude data ranged between 0.47 (polyunsaturated fat) and 0.83 (alcohol); deattenuation improved the correlation coefficients for energy intake and all nutrients, especially for those with the highest variance ratio (polyunsaturated fat and cholesterol). Energy adjustment generally improved the correlation coefficients, except for alcohol. Further deattenuation improved the energy-adjusted correlation coefficients, which ranged from 0.54 for dietary fiber to 0.86 for alcohol.

Cross-classification of energy-adjusted nutrient intakes derived from the FFQ and 24-h dietary recalls into quintiles of intake did not show extreme misclassification for total fat, saturated fat, monounsaturated fat, and alcohol (Table 3); for the other nutrients, severe misclassification ranged from 0.8% to 3.7%. Classification into the same or adjacent quintiles ranged from 70% for monosaccharides to 87% for total protein and alcohol intakes.

Comparisons of energy and protein intakes estimated with the FFQ and the biomarkers as the reference methods (urinary nitrogen and DLW) are presented in Table 4. Protein intake reported on the FFQ was 23% lower than estimates derived from urinary nitrogen measurements. The correlation was slightly improved by deattenuation for the within-person variation observed in urinary nitrogen excretion. Severe misclassification was observed in 2% of the subjects; 46% of the subjects were classified into the adjacent quintiles and 35% were correctly classified. The Bland and Altman plot (Figure 1), showing differences between methods for each subject, did not indicate that differences tended to increase as absolute energy intake

### Table 1

General characteristics of the study population: EPIC-Potsdam Study

<table>
<thead>
<tr>
<th></th>
<th>Men (n = 75)</th>
<th></th>
<th>Women (n = 59)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean ± SD</td>
<td>Minimum</td>
<td>Maximum</td>
<td>Mean ± SD</td>
</tr>
<tr>
<td>Age (y)</td>
<td>56 ± 7.6</td>
<td>40</td>
<td>67</td>
<td>52 ± 4.7</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>81 ± 10.6</td>
<td>52</td>
<td>107</td>
<td>67 ± 11.8</td>
</tr>
<tr>
<td>BMI</td>
<td>26.9 ± 3.70</td>
<td>17.6</td>
<td>35.8</td>
<td>26.1 ± 6.5</td>
</tr>
<tr>
<td>EI/BMR</td>
<td>1.41 ± 0.32</td>
<td>0.82</td>
<td>2.29</td>
<td>1.30 ± 0.33</td>
</tr>
</tbody>
</table>

1 n = 134. EPIC, European Prospective Investigation into Cancer and Nutrition; EI, energy intake; BMR, basal metabolic rate.

2–4 Significantly different from men; 2 P = 0.01, 3 P < 0.001, 4 P = 0.05.
TABLE 2
Macronutrients intakes estimated from the food-frequency questionnaire (FFQ) and twelve 24-h dietary recalls, differences between the methods, the ratio of the 2 methods, and correlation coefficients between the methods: EPIC-Potsdam Study

<table>
<thead>
<tr>
<th></th>
<th>FFQ</th>
<th>24-h Recalls</th>
<th>Difference2</th>
<th>Variance ratio</th>
<th>Pearson’s r</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sample mean ± SD</td>
<td>Sample mean ± SD</td>
<td>Sample mean ± SD</td>
<td>Sample mean ± SD</td>
<td></td>
</tr>
<tr>
<td>Energy (MJ)</td>
<td>9.05 ± 2.13</td>
<td>8.35 ± 1.90</td>
<td>0.70 ± 1.80 (0.0001)</td>
<td>1.11 ± 1.478 (0.0375)</td>
<td>0.61 ± 0.65</td>
</tr>
<tr>
<td>Total protein (g)</td>
<td>75.3 ± 19.4</td>
<td>70.1 ± 15.4</td>
<td>5.2 ± 17.6 (0.0000)</td>
<td>1.10 ± 2.511 (0.0001)</td>
<td>0.51 ± 0.56</td>
</tr>
<tr>
<td>Total fat (g)</td>
<td>75.4 ± 24.0</td>
<td>78.7 ± 22.3</td>
<td>−3.3 ± 21.1 (0.0750)</td>
<td>0.99 ± 2.014 (0.0001)</td>
<td>0.59 ± 0.64</td>
</tr>
<tr>
<td>Saturated fat (g)</td>
<td>28.1 ± 10.1</td>
<td>32.2 ± 10.1</td>
<td>−3.35 ± 8.8 (0.0001)</td>
<td>0.92 ± 1.999 (0.0001)</td>
<td>0.62 ± 0.67</td>
</tr>
<tr>
<td>Monounsaturated fat (g)</td>
<td>24.6 ± 8.3</td>
<td>29.1 ± 9.2</td>
<td>−4.5 ± 7.9 (0.0001)</td>
<td>0.88 ± 1.980 (0.0001)</td>
<td>0.59 ± 0.64</td>
</tr>
<tr>
<td>Polysaturated fat (g)</td>
<td>12.3 ± 3.9</td>
<td>12.1 ± 3.3</td>
<td>0.2 ± 3.8 (0.5407)</td>
<td>1.06 ± 4.268 (0.0001)</td>
<td>0.47 ± 0.55</td>
</tr>
<tr>
<td>Cholesterol (mg)</td>
<td>292 ± 97</td>
<td>299 ± 87</td>
<td>−7 ± 93 (0.3557)</td>
<td>0.98 ± 6.436 (0.0001)</td>
<td>0.50 ± 0.62</td>
</tr>
<tr>
<td>Total carbohydrate (g)</td>
<td>237.6 ± 57.0</td>
<td>218.7 ± 51.8</td>
<td>16.9 ± 51.3 (0.0001)</td>
<td>1.11 ± 1.658 (0.0001)</td>
<td>0.56 ± 0.60</td>
</tr>
<tr>
<td>Monosaccharides (g)</td>
<td>33.7 ± 13.3</td>
<td>38.0 ± 13.4</td>
<td>−4.3 ± 13.1 (0.0003)</td>
<td>0.94 ± 2.391 (0.0001)</td>
<td>0.52 ± 0.57</td>
</tr>
<tr>
<td>Disaccharides (g)</td>
<td>67.3 ± 31.4</td>
<td>66.5 ± 26.1</td>
<td>0.8 ± 25.6 (0.5214)</td>
<td>1.05 ± 1.670 (0.0001)</td>
<td>0.66 ± 0.71</td>
</tr>
<tr>
<td>Polysaccharides (g)</td>
<td>138.5 ± 36.9</td>
<td>105.8 ± 29.4</td>
<td>32.6 ± 31.2 (0.0001)</td>
<td>1.35 ± 1.949 (0.0001)</td>
<td>0.59 ± 0.64</td>
</tr>
<tr>
<td>Dietary fiber (g)</td>
<td>27.6 ± 7.1</td>
<td>22.2 ± 5.9</td>
<td>5.4 ± 6.8 (0.0001)</td>
<td>1.29 ± 1.747 (0.0001)</td>
<td>0.47 ± 0.50</td>
</tr>
<tr>
<td>Alcohol (g)</td>
<td>14.7 ± 16.0</td>
<td>15.6 ± 15.9</td>
<td>−0.9 ± 9.3 (0.2979)</td>
<td>1.58 ± 1.343 (0.0001)</td>
<td>0.83 ± 0.88</td>
</tr>
</tbody>
</table>

1 n = 134. EPIC, European Prospective Investigation into Cancer and Nutrition.
2 P values in parentheses (t test).
3 ± SD.
4 Wilcoxon signed-rank test.
5 Spearman’s r.

Increased. The mean difference was negative because of low reported protein intakes. The spread around the mean reflected the variation in differences, reflecting the observation that the 2 methods were only moderately correlated.

Reported energy intake on the FFQ was 22% lower than the measured energy requirement (r = 0.48). Thirty-six percent of the subjects were cross-classified into the same quartile, 39% were cross-classified into the adjacent quartile, and 3.6% of the subjects were cross-classified into the opposite quartiles of the respective distributions of energy intakes and energy expenditure. Differences in energy intakes between the FFQ and the DLW method plotted against the mean of the 2 methods are shown in Figure 2. The mean difference between the 2 methods reflected a considerable amount of underreporting of energy intake for all but one subject, and the width of the limits of agreement from 2000 to −7000 kJ (±2 SD) indicated wide discrepancies between the 2 methods for individual subjects. The differences tended to increase as the absolute energy values increased. Further analysis of this phenomenon indicated that the difference between the FFQ and the DLW method, as well as the difference between the FFQ and the urinary nitrogen method, were significantly correlated with BMI: r = 0.50 (P = 0.007) and r = 0.30 (P = 0.0044), respectively.

To evaluate the phenomenon of underreporting, we plotted the mean of the energy intake from the 24-h dietary recall and DLW data against the difference between these methods (Figure 3). Basically, we observed the same pattern as in the comparison with the FFQ: energy intake from the 24-h dietary recalls was

TABLE 3
Cross-classification of mean daily macronutrient intakes derived from the food-frequency questionnaire and from twelve 24-h dietary recalls into quintiles: EPIC-Potsdam Study

<table>
<thead>
<tr>
<th></th>
<th>Same quintile</th>
<th>Adjacent quintile</th>
<th>Extreme quintile</th>
<th>Same quintile</th>
<th>Adjacent quintile</th>
<th>Extreme quintile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy (MJ)</td>
<td>32.8</td>
<td>40.3</td>
<td>0.0</td>
<td>44.8</td>
<td>42.5</td>
<td>0.8</td>
</tr>
<tr>
<td>Total protein (g)</td>
<td>30.0</td>
<td>44.8</td>
<td>1.5</td>
<td>39.6</td>
<td>42.5</td>
<td>0.0</td>
</tr>
<tr>
<td>Total fat (g)</td>
<td>32.8</td>
<td>42.5</td>
<td>0.0</td>
<td>44.8</td>
<td>33.6</td>
<td>0.0</td>
</tr>
<tr>
<td>Saturated fat (g)</td>
<td>36.6</td>
<td>42.5</td>
<td>1.5</td>
<td>41.8</td>
<td>41.1</td>
<td>0.0</td>
</tr>
<tr>
<td>Monounsaturated fat (g)</td>
<td>36.6</td>
<td>40.3</td>
<td>0.7</td>
<td>36.6</td>
<td>44.0</td>
<td>1.5</td>
</tr>
<tr>
<td>Polysaturated fat (g)</td>
<td>29.1</td>
<td>41.0</td>
<td>0.7</td>
<td>30.6</td>
<td>41.1</td>
<td>0.8</td>
</tr>
<tr>
<td>Cholesterol (mg)</td>
<td>34.3</td>
<td>32.8</td>
<td>3.0</td>
<td>38.1</td>
<td>35.8</td>
<td>3.0</td>
</tr>
<tr>
<td>Total carbohydrates (g)</td>
<td>33.6</td>
<td>38.1</td>
<td>3.7</td>
<td>36.8</td>
<td>32.8</td>
<td>0.8</td>
</tr>
<tr>
<td>Monosaccharides (g)</td>
<td>29.9</td>
<td>40.3</td>
<td>2.2</td>
<td>40.3</td>
<td>38.8</td>
<td>1.5</td>
</tr>
<tr>
<td>Disaccharides (g)</td>
<td>36.6</td>
<td>39.6</td>
<td>0.0</td>
<td>38.8</td>
<td>35.1</td>
<td>0.8</td>
</tr>
<tr>
<td>Polysaccharides (g)</td>
<td>29.9</td>
<td>41.0</td>
<td>0.0</td>
<td>35.1</td>
<td>35.8</td>
<td>3.7</td>
</tr>
<tr>
<td>Dietary fiber (g)</td>
<td>26.1</td>
<td>40.3</td>
<td>2.2</td>
<td>51.5</td>
<td>35.8</td>
<td>0.0</td>
</tr>
<tr>
<td>Alcohol (g)</td>
<td>49.3</td>
<td>43.3</td>
<td>0.0</td>
<td>51.5</td>
<td>35.8</td>
<td>0.0</td>
</tr>
</tbody>
</table>

1 n = 134. EPIC, European Prospective Investigation into Cancer and Nutrition.
clearly underreported compared with the reference method, and
the limits of agreement, though slightly smaller, indicated
discrepancies between the 24-h dietary recall and reference
methods as well. The difference between the 24-h dietary recall
and DLW methods was strongly and highly significantly corre-
sponding difference between the FFQ and
DLW method (r = 0.74, P = 0.0001). Weaker, but also highly
significant correlations for these differences were observed for
urinary nitrogen (r = 0.58, P = 0.0001). These 2 correlation coef-
ficients clearly indicated correlation of measurement error,
which was supported by the plot of energy intakes from the FFQ
and 24-h dietary recalls (Figure 4). This comparison did not
indicate underreporting; however, a slight tendency for increas-
ing differences with increasing energy intakes was also detectable.

DISCUSSION

This validation study is among the first to use the DLW
method to evaluate the validity of a self-administered FFQ with
respect to energy intake. Compared with the quasi gold stan-
dard of TEE measurement by the DLW method, energy intake
was underreported by all but one subject, by 22% on average.
The observed differences between reported energy intake and
TEE may have been due to measurement error with both meth-
ods. The conclusion that energy intake was underreported is
based on the assumption that energy requirements are not over-
estimated by the DLW method, a method that has been suc-
cessfully validated against gas-exchange measurements in lean
and obese volunteers (27, 35–37). The discrepancy between
these methods was found to vary from 2.5% to 3% and, thus,
could not be the reason for the observed differences between
reported energy intake and TEE. Attribution of the error mainly
to the FFQ method is supported by results of other studies
investigating the relation of reported energy intake to TEE
measured by the DLW method. The only published validation
study that compared data from 2 different FFQs with DLW
measurements described underreporting of energy intake as
well (38). Several validation studies of other dietary assess-
ment methods observed underreporting of energy intake in
most but not all studies (39–41). However, conclusions about
the validity of the FFQ are difficult to make unless 2 underly-
ing assumptions hold true: that the subjects are in energy bal-
ance and that the short-term (14 d) TEE measurement repre-
sents habitual usual energy expenditure. The DLW method
permits validation of reported energy intake only if subjects are
in energy balance (41). Subjects in this study were instructed
not to change their dietary and lifestyle habits during the 14 d
of sampling. Mean weight changes during the sampling period
were small (< 1 kg); therefore, the subjects likely were in
energy balance. With respect to the second assumption, there
was only one 14-d TEE measurement period; however, the FFQ
was designed to assess usual dietary intake over the preceding
year. The observed discrepancies between TEE and reported

<p>| TABLE 4 |
| Energy and protein intakes estimated from the food-frequency questionnaire (FFQ) and reference methods (urinary nitrogen and doubly labeled water), differences between the methods, the ratio of the methods, and correlation coefficients between the methods: EPIC-Potsdam Study |</p>
<table>
<thead>
<tr>
<th>FFQ</th>
<th>Reference methods</th>
<th>Difference (FFQ – reference)</th>
<th>FFQreference (%)</th>
<th>Variance ratio</th>
<th>Pearson’s r</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total protein (g), n = 134</td>
<td>75.3 ± 19.4</td>
<td>97.4 ± 18.4</td>
<td>–22.1 ± 20.6</td>
<td>0.79</td>
<td>1.076</td>
</tr>
<tr>
<td>Energy (MJ), n = 28</td>
<td>9.05 ± 2.13</td>
<td>11.23 ± 2.38</td>
<td>–2.49 ± 2.27</td>
<td>0.80</td>
<td>—</td>
</tr>
</tbody>
</table>

1EPIC, European Prospective Investigation into Cancer and Nutrition. 2x±SD.
energy intake might therefore partly be explained by the incomplete coverage of the FFQ data by the DLW measurement and the lack of data on within-person variability.

A second biomarker, urinary nitrogen, was used to validate protein intake reported on the FFQ. Previous studies comparing urinary nitrogen with FFQ data observed correlation coefficients between 0.07 and 0.54 (10, 30, 42, 43), indicating that our data were within the range observed previously. Similar to reported energy intakes, protein intakes were underreported when compared with urinary nitrogen. The use of urinary nitrogen as a reference method for dietary intake requires stable nitrogen balance, e.g., no gain or loss of tissue. For >90% of the study population, the estimated weight change during the 1-y study period was ≤2 kg. However, larger short-term changes in weight might have occurred during the study period. Furthermore, it was assumed that extrarenal nitrogen losses are a constant proportion of overall nitrogen excretion (12). This might not be the case for all subjects. These limitations might therefore partly explain the observed discrepancies between estimated protein excretion and intake.

Relative validity of the FFQ was assessed in comparison with twelve 24-h dietary recalls administered over 1 y to obtain information on performance of the FFQ in our East German study population. Procedures and data analysis were adapted to the previous validation study of the FFQ in the other German EPIC study

FIGURE 2. Differences between energy intake estimated from the food-frequency questionnaire and total energy expenditure estimated with the doubly labeled water method plotted against the mean from the 2 methods in the European Prospective Investigation into Cancer and Nutrition (EPIC) Potsdam Study (n = 28).

FIGURE 3. Differences between energy intake estimated from the 24-h dietary recalls and total energy expenditure estimated with the doubly labeled water method plotted against the mean from the 2 methods in the European Prospective Investigation into Cancer and Nutrition (EPIC) Potsdam Study (n = 28).
population in Heidelberg, where twelve 24-h dietary recalls were obtained in face-to-face interviews (15). Limitations of the applied reference instrument and study procedures were described and discussed previously (15, 16); therefore, only aspects specific to the Potsdam validation study are discussed here.

For the Potsdam validation study, a computer-supported, structured, face-to-face 24-h dietary recall interview was administered by one trained interviewer. In contrast with the Heidelberg Study, for which 24-h dietary recall data were not available for Fridays and Saturdays, we obtained data for Saturdays using a 24-h dietary recall applied on Monday. However, no data were obtained for Fridays because of logistic reasons. Because dietary habits on Fridays typically differ from those on other weekdays, our 24-h dietary recall might have underestimated some nutrients as well as energy intake. Crude correlation coefficients were similar in the 2 studies. Weaker correlations were observed in Potsdam than in Heidelberg only for intakes of alcohol and dietary fiber. Comparison of cross-classifications indicated similar results with the FFQ in our study population; there were no significant differences between the results of the Heidelberg and Potsdam studies. Note, however, that methodologic differences in the 2 validation studies did not allow an unrestricted comparison. Our correlation coefficients were in a range comparable with those of other validation studies (10) and thus we concluded that the relative validity of the FFQ was acceptable.

The validation approach in which a second dietary assessment instrument is used as a reference method assumes uncorrelated errors between the test and reference methods. However, when the dietary assessment instruments used in the present validation study were compared with biomarkers, measurement error between the test methods (FFQ, 24-h recalls) and the reference methods (DLW, urinary nitrogen) were correlated. The validation of dietary assessment instruments has been discussed not only in terms of the appropriateness of the reference instrument but also with respect to the statistics used to compare the validity of the methods (44, 45). As pointed out by Bland and Altman (34), it is inappropriate to use correlation coefficients to analyze agreement between methods. Therefore, the mean difference is calculated to obtain information on bias in the group estimate, and limits of agreement (±2 SD of mean difference) indicate the scatter of individual results. Applying this analytic concept to the energy data, we showed a strongly biased mean of differences and a wide scatter of differences between reported energy intakes and total energy expenditure, despite significant correlations between reported energy intake and TEE. The wide scattering of the differences showed clearly that some subjects underreported their energy intakes more so than did others. This would not have been detected if only 24-h dietary recall data were used as a reference, as shown in Figure 4.

A limitation in the interpretation of our results with respect to the validity of the FFQ in the entire study population is the fact that participants in the validation study were a selected, highly motivated sample of the study population. In addition, 16 subjects dropped out of the validation study. For these reasons, there may have been smaller within-person variation and, therefore, stronger correlations in the validation study than in the entire study population.

Even though some uncertainties remain about the absolute amount of underreporting of energy intake on the FFQ, we observed a systematic increase in measurement error with increasing absolute reported energy intakes. This observation could have serious implications on risk estimates of diet-disease relations. Differential underreporting depending on subject characteristics and selective underreporting of certain macronutrients (46) are issues to be considered in this regard. Our data indicated, for example, that underreporting was higher with higher energy intakes. Because subjects with a high BMI have higher energy requirements, underreporting could be related to obesity, as described previously (46–49). These assumptions are supported by the strong correlations observed in our data between measurement error and BMI. However, the sample size was too small to detect significant determinants of underreporting. Several investigators have noted the importance of obtaining data on random and systematic measurement error of the exposure assessment for correcting linear regression (50) or logistic
regression estimates (51). A detailed characterization of measurement error is also considered to be important in the use and interpretation of energy-adjustment models because measurement error may unpredictably distort the estimated effects of energy-adjustment models (52).

In summary, energy intake was underreported on the FFQ when compared with TEE, and protein intakes were underreported on the FFQ when compared with urinary nitrogen. These observations were similar to those described for other dietary assessment instruments. Correlation coefficients between macronutrient intakes estimated with the FFQ and the 24-h dietary recalls were within the range observed in a previous validation study (16), indicating comparable relative validity for the FFQ used in the present study population.

We thank the interviewers of the EPIC-Potsdam Study.

REFERENCES

38. Sawaya AL, Tucker K, Tsay R, et al. Evaluation of four methods for determining energy intake in young and older women: comparison...