Using a Correction Factor to Correct for Overreporting in a Food-Frequency Questionnaire Does Not Improve Biomarker-Assessed Validity of Estimates for Fruit and Vegetable Consumption

Rik P. Bogers, Pieter C. Dagnelie, Klaas R. Westerterp, Arnold D. M. Kester, Jacob D. van Klaveren, Aalt Bast and Piet A. van den Brandt

Department of Epidemiology, Maastricht University, Maastricht, The Netherlands; Department of Human Biology, Maastricht University, Maastricht, The Netherlands; Department of Methodology & Statistics, Maastricht University, Maastricht, The Netherlands; State Institute for Quality Control of Agricultural Products, Wageningen, The Netherlands; and Department of Pharmacology & Toxicology, Maastricht University, Maastricht, The Netherlands

ABSTRACT To correct for overreporting of fruit and vegetable (FV) consumption in a food-frequency questionnaire, summary questions about consumption of main FV groups are often used to calculate correction factors. This study compared the ability to rank people according to their FV intake of those summary questions and the sum of questions on individual FV items within categories, and of corrected or uncorrected estimates of specific sorts of FV. Healthy middle-age women (n = 161) completed a food-frequency questionnaire about FV consumption during the previous month and gave a single fasting blood sample. Correction factors were calculated as the reported frequency on a summary question divided by the summed frequencies of all items in a category. Plasma carotenoids and vitamin C served as biomarkers of FV consumption. Significant correlations between FV consumption and biomarkers were observed (e.g., Spearman’s correlation coefficient \( r \) with total carotenoids/vitamin C: 0.32/0.34 for vegetables, 0.30/0.25 for fruits). Summary estimates of cooked, raw and total vegetable consumption correlated higher with biomarkers than sum estimates. For fruits no differences in correlations between sum and summary estimates were observed. Applying a correction factor on the consumption of carrots and total cabbage resulted in lower correlations with relevant biomarkers. For broccoli/cauliflower, Brussels sprouts and citrus fruits, correlations with biomarkers did not change after correction. We conclude that summary questions may suffice to rank individuals according to their intake of main FV categories, and that correction for overreporting of individual FV items is probably not advisable when ranking individuals according to intake of these items. J. Nutr. 133: 1213–1219, 2003.

KEY WORDS: food-frequency questionnaire • validity • biomarkers • fruits • vegetables

Many human observational studies suggest that consumption of fruit and vegetables (FV) or associated micronutrients is beneficial in the prevention of cancer (1–4) and cardiovascular disease (5–7).

For epidemiological studies aimed at further establishing associations between FV consumption and disease, a food-frequency questionnaire (FFQ) that adequately classifies persons according to their usual FV intake is a useful tool. Such a questionnaire not only must be suitable for estimating total FV consumption, but also must cover consumption of all relevant sorts of FV to calculate intake of specific (groups of) FV and of essential nutrients and other bioactive compounds. However, estimates of consumption derived from a FFQ are never free of errors and result in biased estimates of reported consumption frequency and portion size. Most subjects have difficulties in estimating the frequency of consumption of specific food items (in this case specific fruits or vegetables). The sum of all items within a certain food category usually results in an estimate that is higher than the true overall frequency of consumption. On the other hand, estimates derived from the sum of items might better represent true intake, because in that case a more appropriate weight belonging to each specific item instead of a mean weight for all the items can be used in the calculations (e.g., an apple weighing more than a tangerine).

In studies where the intake of specific items or nutrients is of interest, a common practice to correct for this overreporting is to include a summary question (how often do you eat vegetables?) preceding or following questions on intake of specific items within a food category. In data analysis, frequencies of intake of specific items are then calibrated so that the sum of all items equals the overall frequency (10). However, the question arises whether in the case of vegetables this is a...
valid method, because more than one type of vegetable may be
served at the same meal. Also for fruits, different sorts may be
eaten during a single day. Amanatidis et al. (9) found that
applying a correction factor did not substantially affect the
ranking of subjects according to intake of various micronutri-
ients. However, they did not have an independent reference
method against which validity of corrected and uncorrected
estimates could be determined.

The concept of validity consists of several aspects, each of
which can be addressed depending on the purpose and context
of the method used (11,12). The type of validity we refer to in
the present study concerns the relative ranking of subjects by
two different methods. A correlation coefficient is the typical
measure used to evaluate this aspect of validity (12).

One way of validating FFQ is by using biochemical markers
of dietary intake. Although there are several drawbacks in the
use of biomarkers (13), they provide an objective measure of
intake of which the errors are largely independent of the errors
associated with FFQ (14). Frequently used biomarkers for FFQ
intake are plasma carotenoids and vitamin C, which have
been shown to be responsive to intake of FV (15–18). Signifi-
cant (although moderate) correlations between plasma or
serum concentrations of carotenoids and carotenoid intake
(19–24) or FV consumption (24–28) have been observed, and
plasma/serum vitamin C was found to be significantly corre-
lated with FV intake (26,28) or vitamin C intake (29). This
study focuses on two questions. First, are summary questions on
FV consumption better able to rank subjects according to their
usual FV consumption than the sum of specific items within
each category? This question was studied for the categories
total vegetables (excluding potatoes), cooked vegetables (ex-
cluding potatoes), raw vegetables, fruits and fruit juice. Sec-
ond, do calibrated estimates of three (groups of) FV (cabbage,
carrots and citrus fruits) lead to a better ranking of subjects
than uncalibrated estimates? Plasma carotenoids (α-carotene,
β-carotene, lutein, lycopene, β-cryptoxanthin) and vitamin C
were used to serve as independent estimates of FV consump-
tion as criterion variables against which validity was assessed.
At the same time, this study can be regarded as a validation
study of the FFQ we developed for estimating FV consump-
tion.

SUBJECTS AND METHODS

Study design. The study was part of a larger project aimed at
increasing FV consumption in children aged 7–10 y and their
mothers, for which we received approval by the medical ethical
committee of the University Hospital in Maastricht. Participants completed an
FFQ about FV consumption in the previous month, after which blood
was taken from the mothers for the analysis of biomarkers. FV
consumption was calculated both with summary questions about FV
categories (cooked vegetables, raw vegetables, fruits, fruit juice) and
with extensive lists of single items within each category. Consump-
tion of specific fruits and vegetables was calculated with and without
correction for overreporting. All data presented in this publication
were collected in March 2001.

Participants. From the population registry of the municipality of
Maastricht, a random sample was obtained containing 2000 addresses
of mothers with children aged 7–10 y. A letter was sent inviting the
mother and the selected child in each family to participate. An
additional 1100 letters were distributed in elementary schools in
Maastricht. The women had to be apparently healthy nonsmokers
who agreed not to use vitamin supplements from 1 mo before the first
blood collection to the end of the study period. We explained to the
participants that the purposes of the study were to develop question-
naires to assess both FFV intake and determinants of FFV consumption
(the latter, in another part of the project, are not described here).
A total of 207 volunteers were recruited (6.7% response rate), of whom
163 women (79%) fully completed the study, making an overall
response rate of 5.3%. Two subjects having FFQ with >20% missing
values were excluded from further analyses, leaving a total number of
161 subjects.

Food-frequency questionnaire. The FFQ was designed to mea-
sure FV intake during the past month and was completed in March
2001. We distributed the questionnaires by mail and asked the participants
to fill out the FFQ on the day before blood collection.

The FFQ we used was a modified version of the VEG-FFQ (from
“vegetable”) developed at the Dutch State Institute for Quality
Control of Agricultural Products (30). This VEG-FFQ was a 172-item
semiquantitative food-frequency questionnaire that was developed
and first used as a part of the Third Dutch National Food Consump-
tion Survey in 1998 (31). Questions were asked separately for summer
and winter. In the design of the VEG-FFQ, existing Dutch FFQ
(32–37) served as guidelines. Specific features of the VEG-FFQ were
as follows: the use of household measures and natural units in ques-
tions on amounts eaten; ordering of the questions according to the
Dutch meal pattern; questions at the level of meals; addition of
overall questions on each product category (e.g., fruits) for correction
for overreporting; grouping products within the same food category
and ordering foods within a category according to frequency of consump-
tion of products; taking into account seasonal differences in
FV consumption. The food categories included in the VEG-FFQ were
soup, warm meals, potatoes, mashed vegetables, salads, cooked vege-
tables, fruits (including apple sauce), drinks and snacks. Questions
on foods were noted as consumption frequency and portion size,
except for sweet peppers, mushrooms, tomatoes and onions, which
were estimated by asking the amount eaten per month or week. The
seven frequency categories used were: never or less than 1 d/mo, 1
d/mo, 2–3 d/mo, and 1, 2–3, 4–5 and 6–7 d/wk.

Before the present study the VEG-FFQ had been used in 1592 out
of 5958 participants of the Third Dutch National Food Consumption
Survey in which subjects recorded their 48-h food intake on two
consecutive days. These 1592 participants completed the VEG-FFQ
4–8 wk after the 48-h records. It was found that mean FV intake
calculated using the VEG-FFQ corresponded well with mean intake
calculated from the 48-h records. After correction for overreporting
using summary questions on food categories, mean vegetable intake
was 105 g/d according to the VEG-FFQ and 123 (95) g/d according to
the records; for fruit intake, this was 114 and 105 (114) g/d,
respectively (30). Further validation was not performed.

For the present study, we slightly modified the VEG-FFQ to fit
into the study design. Separate questions on consumption of FV in
summer and winter were combined into questions on FV consump-
tion during the past month. Open-ended questions on average por-
tion size were changed to multiple-choice questions so that the
questionnaires could be read by an optical scanner. The result was a
106-item semiquantitative FFQ. Calculated FV categories were fruits
(including fresh orange and grapefruit juice), cooked vegetables (not
including potatoes), raw vegetables, total vegetables (the sum of
cooked and raw vegetables) and fruit juice (from a carton or bottle).
For each of these categories, a summary question was asked, followed
by a list of items belonging to a category that we used for calculating
the sum estimates. The numbers of items were 17 (fruits), 21 (cooked
vegetables), 14 (raw vegetables) and 5 (fruit juice).

Assessment of biomarkers in blood. Blood was taken shortly
after the participants completed the FFQ. After an overnight fast, 14
mL venous blood was collected between 0700 and 1100 h in two
K3-EDTA-coated plastic tubes (Vacutainer Systems; Becton Dickin-
son, Plymouth, UK). Both tubes were immediately placed on ice in
the dark.

Total ascorbic acid was determined according to the method of
Speek et al. (38). Briefly, blood was collected in a 4-mL tube in which
0.08 mL EGTA-glutathione solution had been injected before blood
sampling. A 250-µL aliquot of blood was stabilized in 1 mL trichlo-
roacetic acid and centrifuged after 20 to 60 min (15,000 × g, 10 min,
4°C). A pilot study showed that ascorbic acid in trichloroacetic acid
plasma is stable for at least 60 min of incubation. Of the
supernatant, 350 µL was transferred in amber Eppendorf vials (Epp-
dendorf-Netherl-Hinz-GmbH, Hamburg, Germany) and stored at
−80°C. The samples were analyzed within 2 wk with HPLC with

BOGERS ET AL.
fluorometric detection. For this purpose, L-ascorbic acid was enzymatically oxidized to dehydro-L-ascorbic acid and the latter was condensed with o-phenylenediamine to its fluorescent quinoxaline derivative. The HPLC mobile phase consisted of 0.08 mol/L potassium dihydrogen phosphate and methanol (41:1 v/v).

Plasma carotenoids were analyzed according to Hess et al. (39), with modifications described by Oostenbrug et al. (40). Briefly, after blood was drawn in a 10-mL tube it was centrifuged (1000 × g, 10 min, 4°C), and the plasma was transferred into amber Eppendorf vials and stored at −80°C. The plasma was analyzed within 5 mo using HPLC with UV detection. First, 500 µL double-distilled water was added to 500 µL plasma and this mixture was denatured with 1 mL of a mixture of absolute ethanol and methanol (1:1 v/v). Carotenoids were twice extracted into 3 mL hexane. The pooled supernatant was evaporated to dryness under nitrogen at 37°C and resuspended in 200 µL ethanol-dioxan (1:1 v/v), after which 300 µL acetonitrile was added. The HPLC mobile phase consisted of acetonitrile, tetrahydrofuran, methanol and double-distilled water (68:22:6:28 v/v/v/v).

Calculations of FV intake and statistical analysis. Statistical analyses were performed by use of the SPSS program (version 10.0; SPSS, Chicago, IL). Standard amounts as indicated in a Dutch table for measures and weights (41,42) were used for converting household measures to portion sizes. Missing values for portion sizes on a particular item were substituted with mean portion sizes for the item concerned. FV intake (in g/d) was calculated as the product of frequency of intake and portion size. Mean values were used to substitute missing values for FV intake. Besides being included under the heading “raw vegetables,” consumption of sweet peppers and tomatoes was recorded as the absolute amount eaten both raw and cooked per week or month. After subtraction of raw sweet peppers and tomatoes we added these amounts to cooked vegetables. We supposed that onions and mushrooms, which were also included as raw and cooked amounts eaten per month, were eaten predominantly cooked, and added these two items to cooked vegetables. The correction factor for calibrating items was calculated as the reported frequency on the summary question divided by the sum of frequencies of all items belonging to a category. Corrected intake of items was calculated by multiplying intake of an item by the correction factor. Corrected intake of a category was the sum of all corrected items in a category. The relative validity of corrected and uncorrected intakes was determined for carrots, citrus fruits and cabbage (total cabbage, Brussels sprouts and broccoli/cauliflower), because of their known high contents of α-carotene, β-cryptoxanthin and lutein, respectively.

Agreement between the summary question and the sum of individual items within a category was determined by cross-classifying quartiles of FV intake according to the sum of items in a category with quartiles according to the summary question. Validity was calculated as correlation coefficients between FV intake and biomarkers. We used Spearman’s correlation coefficients because of the skewed distribution of most food consumption data. To compare the validity of two estimates of FV consumption, equality of correlations was tested using a formula stated by Olkin and Siotani (43,44).

RESULTS

Table 1 provides a summary of some of the participants’ characteristics. Table 2 presents a comparison between mean FV consumption derived from one summary question or from summing all items in a category. The greatest difference in intake between these two methods was for cooked vegetables: 110 g/d using the summary question (summary amount) vs. 204 g/d using the sum of items (sum amount). For cooked vegetables the correlation between frequencies (r = 0.35) and intake (r = 0.49) measured with either method was lowest; however, the mean correction factor of 0.89 was closer to 1.00, and the mean difference between sum and summary frequencies was lower than for raw vegetables and fruit. For fruit juice the sum amount approached the summary amount most closely (67 vs. 79 g/d), and for fruits this was 156 vs. 195 g/d. The mean correction factor for fruit juice was almost 1.00 because most people drank only orange juice, so that the sum frequency consisted of only one item.

Table 3 illustrates the agreement in results when the summary question or the sum of items in a category was used to calculate FV intake. The best agreement was found for fruits and fruit juice, where 95 and 93% of the subjects were classified in the same or adjacent quartile, respectively. Again, the lowest agreement was observed for cooked vegetables (83% in the same or adjacent quartile).

Table 4 details the descriptive statistics for the biomarkers. The carotenoid most abundant in plasma was β-cryptoxanthin, and the least abundant was α-carotene. The median values corresponded well with mean values. Table 5 presents the results of the validation of the two methods for calculating FV consumption against plasma carotenoids and vitamin C. Values of correlations with biomarkers were consistently higher for summary amounts than for sum amounts. Some of these differences were statistically significant, especially for total vegetable consumption. In contrast, for fruit consumption the correlations with biomarkers were comparable for the summary and the sum amounts. For cooked vegetables, the highest correlation was with vitamin C (Spearman’s correlation coefficient r = 0.35). Consumption of raw vegetables correlated highest with α-carotene and total carotenoids (r = 0.31 for both), and fruit consumption with β-cryptoxanthin (r = 0.43). Consumption of fruit juice was not significantly correlated with any of the biomarkers, regardless of whether it was expressed as summary or sum amount.

Table 6 shows the relative validity of the estimated consumption of cabbage, carrots and citrus fruits when a correction factor was used or not. For carrots, correlations with α- and β-carotene were significantly higher when consumption was not corrected for overreporting; this was also the case for α-carotene and consumption of total cabbage. However, there were no differences in correlations with biomarkers between corrected and uncorrected consumption of broccoli/cauliflower and Brussels sprouts, two specific sorts of cabbage. The correlation between consumption of citrus fruits and its most important marker, β-cryptoxanthin, did not change when consumption was corrected for overreporting.

DISCUSSION

The present study indicates that, based on biomarkers as a reference for validation of a FFQ, estimates of consumption of vegetables based on one summary question (“summary estimate”) are better able to rank persons according to their FV intake than estimates obtained by summing all separate items across each category (“sum estimate”). Our data suggest that this applies to both raw and, even more so, cooked vegetables.
For fruits, the correlations with biomarkers for summary and sum estimates are comparable. Furthermore, when intake of individual fruits and vegetables is estimated in a FFQ, correction for overreporting seems unnecessary or may even lower correlations with biomarkers.

Several possible explanations can be given as to why the summary estimate is as valid as the sum estimate for fruits, whereas this is not the case for vegetables. First, for fruits portion size is easier to indicate because, contrary to vegetables, whereas this is not the case for vegetables. First, for fruits individual sorts of fruits are mostly served in natural units. Second, it may be easier for subjects to estimate consumption frequencies of specific fruits because the number of different sorts of fruits eaten is usually smaller than the number of different vegetables (31). However, the mean correction factor (based on frequencies) for fruits was close to 0.5, suggesting a high degree of overestimating the frequency of consumption of individual fruits relative to the summary frequency. Rather than overestimating the frequency of consumption of individual sorts of fruits, this could be attributed to the fact that people eat more than one sort of fruit a day. The finding that the sum amount was only 25% higher than the summary amount supports this idea. In contrast, the mean correction factor was closest to unity for cooked vegetables, although consumption of cooked vegetables (expressed in g/d) showed the largest difference between summary and sum estimates. This was the result of the great between-subject variation in the correction factor for cooked vegetables, where, not contrary to fruits, almost every subject reported a higher sum than summary frequency.

The observation that correction for overreporting may not improve or even lower correlations with biomarkers may be attributable to the fact that the correction factor includes estimate errors of all separate items. The sum of these errors may likely be higher than the estimate error of a single item, which would reduce the validity of an estimate of a specific fruit or vegetable when correction for overreporting is applied. As already mentioned, the consumption frequency of fruits may be easier to estimate, and as a result the error of the correction factor for fruits would probably be less than that of vegetables.

We used biomarkers as a reference method because the errors in biomarkers (e.g., biological variation between persons in uptake or metabolism of vitamins) are to a high degree statistically independent of the errors in intake estimates based on questionnaires. Although the correlations we observed are moderate, the use of biomarkers for our purpose can be justified. Because the variation in our criterion measure (the biomarkers) remains the same when we calculate correlations with either of the two dietary measures, differences between the two correlation coefficients (biomarkers vs. summary/sum

**TABLE 2**

Comparison between mean fruit and vegetable intake derived from one summary question or from summing all items in a category, and Spearman’s correlation coefficients between these two methods1,2

<table>
<thead>
<tr>
<th></th>
<th>Summary3</th>
<th>Sum3</th>
<th>Spearman’s r (P)4</th>
<th>Summary frequency3</th>
<th>Sum frequency3</th>
<th>Spearman’s r (P)4</th>
<th>Correction factor3,5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>g/d</td>
<td>g/d</td>
<td>times/d</td>
<td>times/d</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total vegetables</td>
<td>151 ± 69</td>
<td>264 ± 123</td>
<td>0.59 (0.00)</td>
<td>1.09 ± 0.45</td>
<td>1.83 ± 1.03</td>
<td>0.61 (0.00)</td>
<td>0.69 ± 0.30</td>
</tr>
<tr>
<td>Cooked</td>
<td>110 ± 51</td>
<td>204 ± 94</td>
<td>0.49 (0.00)</td>
<td>0.65 ± 0.22</td>
<td>0.83 ± 0.39</td>
<td>0.35 (0.00)</td>
<td>0.89 ± 0.42</td>
</tr>
<tr>
<td>Raw</td>
<td>41 ± 38</td>
<td>60 ± 69</td>
<td>0.80 (0.00)</td>
<td>0.44 ± 0.35</td>
<td>1.01 ± 0.82</td>
<td>0.80 (0.00)</td>
<td>0.61 ± 0.27</td>
</tr>
<tr>
<td>Fruit</td>
<td>156 ± 116</td>
<td>195 ± 128</td>
<td>0.82 (0.00)</td>
<td>0.69 ± 0.28</td>
<td>1.34 ± 0.78</td>
<td>0.71 (0.00)</td>
<td>0.59 ± 0.27</td>
</tr>
<tr>
<td>Fruit juice</td>
<td>67 ± 84</td>
<td>79 ± 95</td>
<td>0.81 (0.00)</td>
<td>0.32 ± 0.33</td>
<td>0.37 ± 0.35</td>
<td>0.81 (0.00)</td>
<td>0.97 ± 0.60</td>
</tr>
</tbody>
</table>

1 Assessed with a food-frequency questionnaire in 161 women.
2 Values are means ± sd.
3 All differences (paired t test and Wilcoxon signed rank test) between summary and sum amounts and frequencies are significant at P < 0.001.
4 P-values are in parentheses.
5 Calculated as (frequency of consumption indicated in summary question)/(sum of frequencies of consumption of all items in a category).
6 Sum and summary significantly different at P = 0.002.

**TABLE 3**

Cross-classification of quartiles derived from calculating fruit and vegetable intake as the sum of all items in a category by quartiles of intake using a summary question1

<table>
<thead>
<tr>
<th></th>
<th>Same quartile using summary</th>
<th>Adjacent quartile using summary</th>
<th>Two quartiles’ difference using summary</th>
<th>Extreme opposites using summary</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total vegetables</td>
<td>47.2</td>
<td>39.2</td>
<td>10.6</td>
<td>3.1</td>
</tr>
<tr>
<td>Cooked</td>
<td>42.2</td>
<td>41.0</td>
<td>17.9</td>
<td>3.1</td>
</tr>
<tr>
<td>Raw</td>
<td>56.5</td>
<td>36.6</td>
<td>6.7</td>
<td>0.0</td>
</tr>
<tr>
<td>Fruits</td>
<td>59.7</td>
<td>35.5</td>
<td>4.9</td>
<td>0.0</td>
</tr>
<tr>
<td>Fruit juice</td>
<td>67.1</td>
<td>25.9</td>
<td>4.9</td>
<td>1.8</td>
</tr>
</tbody>
</table>

1 Assessed with a food-frequency questionnaire in 161 women.

**TABLE 4**

Percentiles of plasma concentrations of vitamin C and carotenoids1

<table>
<thead>
<tr>
<th></th>
<th>P10</th>
<th>P25</th>
<th>P50</th>
<th>P75</th>
<th>P90</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>μmol/L</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vitamin C</td>
<td>39.93</td>
<td>47.89</td>
<td>56.24</td>
<td>65.64</td>
<td>72.65</td>
</tr>
<tr>
<td>Total carotenoids</td>
<td>1.24</td>
<td>1.49</td>
<td>1.90</td>
<td>2.20</td>
<td>2.94</td>
</tr>
<tr>
<td>Lutein</td>
<td>0.24</td>
<td>0.29</td>
<td>0.38</td>
<td>0.48</td>
<td>0.61</td>
</tr>
<tr>
<td>β-Cryptoxanthin</td>
<td>0.30</td>
<td>0.39</td>
<td>0.51</td>
<td>0.71</td>
<td>1.04</td>
</tr>
<tr>
<td>Lycopene</td>
<td>0.14</td>
<td>0.22</td>
<td>0.26</td>
<td>0.36</td>
<td>0.43</td>
</tr>
<tr>
<td>α-Carotene</td>
<td>0.06</td>
<td>0.09</td>
<td>0.13</td>
<td>0.19</td>
<td>0.24</td>
</tr>
<tr>
<td>β-Carotene</td>
<td>0.21</td>
<td>0.29</td>
<td>0.45</td>
<td>0.65</td>
<td>1.02</td>
</tr>
</tbody>
</table>

1 P10 to P90 are the 10th to the 90th percentiles, n = 161 women.
Our findings of lower summary than sum amounts is in line with food consumption studies in which FV intake measured by short frequency-based screening modules was compared with intake as determined by 3-d diet records (45,46), repeated 24-h recalls (46,47) or a more extensive FFQ (48). All these studies showed that the short instruments underestimated FV consumption relative to the reference methods used.

In contrast, screening modules gave slightly better (45) or similar (47) correlations with dietary records or recalls in comparison with an extensive FFQ, indicating that they were able to rank people according to their FV intake. However, because of correlated errors in the screener and reference method these correlations could be overestimated (49). We are aware of only one study in which summary and sum measures of FV consumption were compared with plasma carotenoids. Contrary to our findings, no substantial differences in correlations between the two measures of consumption were found (46). The different results may be attributed

### TABLE 5

**Spearman’s rank correlations of plasma carotenoid and vitamin C concentrations with fruit and vegetable intake calculated either with a summary question or as the sum of the items in a category**

<table>
<thead>
<tr>
<th></th>
<th>Vitamin C</th>
<th>Total carotenoids</th>
<th>Lutein</th>
<th>β-Cryptoxanthin</th>
<th>Lycopene</th>
<th>α-Carotene</th>
<th>β-Carotene</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total vegetables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Summary^3</td>
<td>0.34 (0.00)</td>
<td>0.32 (0.00)*</td>
<td>0.30 (0.00)</td>
<td>0.19 (0.02)*</td>
<td>0.00 (0.96)</td>
<td>0.37 (0.00)*</td>
<td>0.26 (0.00)*</td>
</tr>
<tr>
<td>Sum^3</td>
<td>0.23 (0.00)</td>
<td>0.22 (0.01)*</td>
<td>0.25 (0.00)</td>
<td>0.09 (0.25)*</td>
<td>0.03 (0.67)</td>
<td>0.20 (0.01)*</td>
<td>0.14 (0.08)*</td>
</tr>
<tr>
<td><strong>Cooked vegetables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Summary</td>
<td>0.35 (0.00)*</td>
<td>0.23 (0.00)</td>
<td>0.24 (0.00)</td>
<td>0.14 (0.07)</td>
<td>0.04 (0.62)</td>
<td>0.24 (0.00)</td>
<td>0.15 (0.05)</td>
</tr>
<tr>
<td>Sum</td>
<td>0.16 (0.04)*</td>
<td>0.19 (0.02)</td>
<td>0.21 (0.01)</td>
<td>0.08 (0.34)</td>
<td>0.06 (0.48)</td>
<td>0.14 (0.08)</td>
<td>0.11 (0.15)</td>
</tr>
<tr>
<td><strong>Raw vegetables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Summary</td>
<td>0.19 (0.02)</td>
<td>0.31 (0.00)</td>
<td>0.22 (0.01)</td>
<td>0.23 (0.00)</td>
<td>0.02 (0.84)</td>
<td>0.31 (0.00)</td>
<td>0.25 (0.00)</td>
</tr>
<tr>
<td>Sum</td>
<td>0.14 (0.08)</td>
<td>0.21 (0.01)</td>
<td>0.20 (0.01)</td>
<td>0.11 (0.18)</td>
<td>0.01 (0.88)</td>
<td>0.20 (0.01)</td>
<td>0.17 (0.03)</td>
</tr>
<tr>
<td><strong>Fruit</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Summary</td>
<td>0.25 (0.00)</td>
<td>0.30 (0.00)</td>
<td>0.10 (0.21)</td>
<td>0.43 (0.00)</td>
<td>0.03 (0.69)</td>
<td>0.15 (0.07)</td>
<td>0.18 (0.02)</td>
</tr>
<tr>
<td>Sum</td>
<td>0.24 (0.00)</td>
<td>0.34 (0.00)</td>
<td>0.15 (0.06)</td>
<td>0.42 (0.00)</td>
<td>0.01 (0.89)</td>
<td>0.17 (0.04)</td>
<td>0.22 (0.01)</td>
</tr>
<tr>
<td><strong>Fruit juice</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Summary</td>
<td>0.15 (0.06)</td>
<td>0.09 (0.28)</td>
<td>−0.02 (0.77)</td>
<td>0.08 (0.34)</td>
<td>0.02 (0.79)</td>
<td>0.02 (0.78)*</td>
<td>0.05 (0.52)</td>
</tr>
<tr>
<td>Sum</td>
<td>0.14 (0.08)</td>
<td>0.05 (0.56)</td>
<td>0.01 (0.93)</td>
<td>0.08 (0.31)</td>
<td>0.05 (0.50)</td>
<td>−0.10 (0.19)*</td>
<td>−0.02 (0.78)</td>
</tr>
</tbody>
</table>

1 Assessed with a food-frequency questionnaire in 161 women.
2 P-values are in parentheses.
3 Uncorrected and corrected refer to uncorrected consumption and consumption after multiplication with a correction factor based on frequencies of consumption. * Uncorrected and corrected correlation coefficients significantly different from each other at P < 0.05.

### TABLE 6

**Spearman’s rank correlations of plasma carotenoid and vitamin C concentrations with consumption of cabbage, carrots and citrus fruits with and without correction for under- or overreporting**

<table>
<thead>
<tr>
<th></th>
<th>Vitamin C</th>
<th>Total carotenoids</th>
<th>Lutein</th>
<th>β-Cryptoxanthin</th>
<th>Lycopene</th>
<th>α-Carotene</th>
<th>β-Carotene</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Carrots</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uncorrected^3</td>
<td>0.19 (0.02)</td>
<td>0.29 (0.00)*</td>
<td>0.17 (0.03)</td>
<td>0.12 (0.12)</td>
<td>0.04 (0.58)</td>
<td>0.39 (0.00)*</td>
<td>0.23 (0.00)*</td>
</tr>
<tr>
<td>Corrected^3</td>
<td>0.24 (0.00)</td>
<td>0.21 (0.01)*</td>
<td>0.13 (0.11)</td>
<td>0.09 (0.26)</td>
<td>0.05 (0.56)</td>
<td>0.28 (0.00)*</td>
<td>0.15 (0.06)*</td>
</tr>
<tr>
<td><strong>Cabbage</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uncorrected</td>
<td>0.16 (0.04)</td>
<td>0.20 (0.01)</td>
<td>0.19 (0.01)</td>
<td>0.11 (0.17)</td>
<td>−0.10 (0.20)</td>
<td>0.21 (0.01)*</td>
<td>0.16 (0.04)</td>
</tr>
<tr>
<td>Corrected</td>
<td>0.13 (0.10)</td>
<td>0.11 (0.15)</td>
<td>0.15 (0.06)</td>
<td>0.08 (0.31)</td>
<td>−0.07 (0.40)</td>
<td>0.10 (0.23)*</td>
<td>0.10 (0.21)</td>
</tr>
<tr>
<td><strong>Brussels sprouts</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uncorrected</td>
<td>0.05 (0.52)</td>
<td>0.09 (0.24)</td>
<td>0.22 (0.01)</td>
<td>0.04 (0.61)</td>
<td>−0.08 (0.35)</td>
<td>0.10 (0.20)</td>
<td>0.05 (0.50)</td>
</tr>
<tr>
<td>Corrected</td>
<td>0.06 (0.43)</td>
<td>0.07 (0.40)</td>
<td>0.21 (0.01)</td>
<td>0.05 (0.55)</td>
<td>−0.08 (0.30)</td>
<td>0.08 (0.31)</td>
<td>0.04 (0.58)</td>
</tr>
<tr>
<td><strong>Broccoli/cauliflower</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uncorrected</td>
<td>0.18 (0.02)</td>
<td>0.14 (0.07)</td>
<td>0.20 (0.01)</td>
<td>0.04 (0.63)</td>
<td>−0.14 (0.09)</td>
<td>0.12 (0.13)</td>
<td>0.11 (0.19)</td>
</tr>
<tr>
<td>Corrected</td>
<td>0.18 (0.02)</td>
<td>0.10 (0.22)</td>
<td>0.19 (0.02)</td>
<td>0.02 (0.85)</td>
<td>−0.10 (0.21)</td>
<td>0.05 (0.56)</td>
<td>0.05 (0.51)</td>
</tr>
<tr>
<td><strong>Citrus fruits</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uncorrected</td>
<td>0.21 (0.01)</td>
<td>0.24 (0.00)</td>
<td>0.08 (0.35)</td>
<td>0.54 (0.00)</td>
<td>−0.02 (0.85)</td>
<td>−0.02 (0.82)</td>
<td>0.02 (0.76)</td>
</tr>
<tr>
<td>Corrected</td>
<td>0.24 (0.00)</td>
<td>0.21 (0.01)</td>
<td>0.06 (0.46)</td>
<td>0.54 (0.00)</td>
<td>−0.06 (0.48)</td>
<td>−0.03 (0.70)</td>
<td>−0.00 (0.96)</td>
</tr>
</tbody>
</table>

1 Assessed with a food-frequency questionnaire in 161 women.
2 P-values are in parentheses.
3 Uncorrected and corrected refer to uncorrected consumption and consumption after multiplication with a correction factor based on frequencies of consumption. * Uncorrected and corrected correlation coefficients significantly different from each other at P < 0.05.
either to the lower mean FV consumption in their population (50) or to the smaller number of items these authors used for calculating sum amounts, which could lead to a sum estimate closer to the summary estimate (51).

The correlations we found between FV consumption and biomarkers were moderate, but comparable to those from several other studies in which FV intake was related to plasma carotenoids and/or vitamin C. Correlations observed in these studies were approximately between 0.2 and 0.5, with the exception of lycopene, for which much lower correlations were found (24–28), as we did. A strong point of our study is that the results are not biased by the effect of smoking, nor by the use of vitamin supplements on concentrations of carotenoids and vitamin C in blood, given that all participants were nonsmokers who agreed not to use vitamin supplements. We are aware of the fact that plasma concentrations of carotenoids and vitamin C vary over time (20,27), and that a single measurement may not represent the 1-mo average. Repeated blood measurements would have improved the correlations we found (24), as we did. A strong point of our study is that different correlation coefficients are reported with (25,26,52) or without (52–54) adjustment for blood cholesterol levels, we decided not to include this adjustment because these studies showed that correlations between plasma or serum carotenoids and FV or carotenoid intake changed only marginally after adjusting for cholesterol (25,26,52–54). It is also possible that adjustment for energy intake would yield different correlation coefficients. Because we wanted to assess FV intake elaborately, we made the concession not to determine energy intake because this would have resulted in a far more extended questionnaire, increasing the risk that participants would complete it less accurately. Before generalizing the results of the present study, attention should be paid to the high degree of selective participation by subjects who were probably interested in healthy eating. How exactly this has affected our findings is difficult to say. On the one hand, our study population could have completed the questionnaire more carefully than would have been the case for the general population. On the other hand, it is also possible that our study population, being more aware of the healthy properties of FV than the average Dutch population, would be biased toward overreporting their FV intake. In addition, studying a more representative population could also have increased the observed correlation coefficients as a result of the greater range in FV intake we would have measured.

In conclusion, for the purpose of ranking people according to their intake of main groups of FV, it would appear that a small number of brief, general questions on FV intake suffice. This has implications in the design of FFQ for use found between FV intake and biomarkers, although we expected this increase would be minor because of the short study period of 1 mo.

Although in literature correlation coefficients are reported with (25,26,52) or without (52–54) adjustment for blood cholesterol levels, we decided not to include this adjustment because these studies showed that correlations between plasma or serum carotenoids and FV or carotenoid intake changed only marginally after adjusting for cholesterol (25,26,52–54). It is also possible that adjustment for energy intake would yield different correlation coefficients. Because we wanted to assess FV intake elaborately, we made the concession not to determine energy intake because this would have resulted in a far more extended questionnaire, increasing the risk that participants would complete it less accurately.

Before generalizing the results of the present study, attention should be paid to the high degree of selective participation by subjects who were probably interested in healthy eating. How exactly this has affected our findings is difficult to say. On the one hand, our study population could have completed the questionnaire more carefully than would have been the case for the general population. On the other hand, it is also possible that our study population, being more aware of the healthy properties of FV than the average Dutch population, would be biased toward overreporting their FV intake. In addition, studying a more representative population could also have increased the observed correlation coefficients as a result of the greater range in FV intake we would have measured.

In conclusion, for the purpose of ranking people according to their intake of main groups of FV, it would appear that a small number of brief, general questions on FV intake suffice. This has implications in the design of FFQ for use found between FV intake and biomarkers, although we expected this increase would be minor because of the short study period of 1 mo.

Although in literature correlation coefficients are reported with (25,26,52) or without (52–54) adjustment for blood cholesterol levels, we decided not to include this adjustment because these studies showed that correlations between plasma or serum carotenoids and FV or carotenoid intake changed only marginally after adjusting for cholesterol (25,26,52–54). It is also possible that adjustment for energy intake would yield different correlation coefficients. Because we wanted to assess FV intake elaborately, we made the concession not to determine energy intake because this would have resulted in a far more extended questionnaire, increasing the risk that participants would complete it less accurately.

Before generalizing the results of the present study, attention should be paid to the high degree of selective participation by subjects who were probably interested in healthy eating. How exactly this has affected our findings is difficult to say. On the one hand, our study population could have completed the questionnaire more carefully than would have been the case for the general population. On the other hand, it is also possible that our study population, being more aware of the healthy properties of FV than the average Dutch population, would be biased toward overreporting their FV intake. In addition, studying a more representative population could also have increased the observed correlation coefficients as a result of the greater range in FV intake we would have measured.

In conclusion, for the purpose of ranking people according to their intake of main groups of FV, it would appear that a small number of brief, general questions on FV intake suffice. This has implications in the design of FFQ for use found between FV intake and biomarkers, although we expected this increase would be minor because of the short study period of 1 mo.

Although in literature correlation coefficients are reported with (25,26,52) or without (52–54) adjustment for blood cholesterol levels, we decided not to include this adjustment because these studies showed that correlations between plasma or serum carotenoids and FV or carotenoid intake changed only marginally after adjusting for cholesterol (25,26,52–54). It is also possible that adjustment for energy intake would yield different correlation coefficients. Because we wanted to assess FV intake elaborately, we made the concession not to determine energy intake because this would have resulted in a far more extended questionnaire, increasing the risk that participants would complete it less accurately.

Before generalizing the results of the present study, attention should be paid to the high degree of selective participation by subjects who were probably interested in healthy eating. How exactly this has affected our findings is difficult to say. On the one hand, our study population could have completed the questionnaire more carefully than would have been the case for the general population. On the other hand, it is also possible that our study population, being more aware of the healthy properties of FV than the average Dutch population, would be biased toward overreporting their FV intake. In addition, studying a more representative population could also have increased the observed correlation coefficients as a result of the greater range in FV intake we would have measured.

In conclusion, for the purpose of ranking people according to their intake of main groups of FV, it would appear that a small number of brief, general questions on FV intake suffice. This has implications in the design of FFQ for use found between FV intake and biomarkers, although we expected this increase would be minor because of the short study period of 1 mo.

Although in literature correlation coefficients are reported with (25,26,52) or without (52–54) adjustment for blood cholesterol levels, we decided not to include this adjustment because these studies showed that correlations between plasma or serum carotenoids and FV or carotenoid intake changed only marginally after adjusting for cholesterol (25,26,52–54). It is also possible that adjustment for energy intake would yield different correlation coefficients. Because we wanted to assess FV intake elaborately, we made the concession not to determine energy intake because this would have resulted in a far more extended questionnaire, increasing the risk that participants would complete it less accurately.

Before generalizing the results of the present study, attention should be paid to the high degree of selective participation by subjects who were probably interested in healthy eating. How exactly this has affected our findings is difficult to say. On the one hand, our study population could have completed the questionnaire more carefully than would have been the case for the general population. On the other hand, it is also possible that our study population, being more aware of the healthy properties of FV than the average Dutch population, would be biased toward overreporting their FV intake. In addition, studying a more representative population could also have increased the observed correlation coefficients as a result of the greater range in FV intake we would have measured.

In conclusion, for the purpose of ranking people according to their intake of main groups of FV, it would appear that a small number of brief, general questions on FV intake suffice. This has implications in the design of FFQ for use found between FV intake and biomarkers, although we expected this increase would be minor because of the short study period of 1 mo.

Although in literature correlation coefficients are reported with (25,26,52) or without (52–54) adjustment for blood cholesterol levels, we decided not to include this adjustment because these studies showed that correlations between plasma or serum carotenoids and FV or carotenoid intake changed only marginally after adjusting for cholesterol (25,26,52–54). It is also possible that adjustment for energy intake would yield different correlation coefficients. Because we wanted to assess FV intake elaborately, we made the concession not to determine energy intake because this would have resulted in a far more extended questionnaire, increasing the risk that participants would complete it less accurately.

Before generalizing the results of the present study, attention should be paid to the high degree of selective participation by subjects who were probably interested in healthy eating. How exactly this has affected our findings is difficult to say. On the one hand, our study population could have completed the questionnaire more carefully than would have been the case for the general population. On the other hand, it is also possible that our study population, being more aware of the healthy properties of FV than the average Dutch population, would be biased toward overreporting their FV intake. In addition, studying a more representative population could also have increased the observed correlation coefficients as a result of the greater range in FV intake we would have measured.

In conclusion, for the purpose of ranking people according to their intake of main groups of FV, it would appear that a small number of brief, general questions on FV intake suffice. This has implications in the design of FFQ for use found between FV intake and biomarkers, although we expected this increase would be minor because of the short study period of 1 mo.
Validation of Fruit and Vegetable Questionnaire


