

## Prediction of fruit and vegetable intake from biomarkers using individual participant data of diet-controlled intervention studies

Olga W. Souverein<sup>1\*</sup>, Jeanne H. M. de Vries<sup>1</sup>, Riitta Freese<sup>2</sup>, Bernhard Watzl<sup>3</sup>, Achim Bub<sup>3</sup>, Edgar R. Miller III<sup>4</sup>, Jacqueline J. M. Castenmiller<sup>5</sup>, Wilrike J. Pasman<sup>6</sup>, Karin van het Hof<sup>7</sup>, Mridula Chopra<sup>8</sup>, Anette Karlsen<sup>9</sup>, Lars O. Dragsted<sup>10</sup>, Renate Winkels<sup>1</sup>, Catherine Itsiopoulos<sup>11</sup>, Laima Brazionis<sup>12</sup>, Kerin O'Dea<sup>13</sup>, Carolien A. van Loo-Bouwman<sup>14</sup>, Ton H. J. Naber<sup>15</sup>, Hilko van der Voet<sup>16</sup> and Hendriek C. Boshuizen<sup>1</sup>

<sup>1</sup>Division of Human Nutrition, Wageningen University, PO Box 8129, 6700 EV Wageningen, The Netherlands

<sup>2</sup>Division of Nutrition, Department of Food and Environmental Sciences, University of Helsinki, Helsinki, Finland

<sup>3</sup>Department of Physiology and Biochemistry of Nutrition, Max Rubner-Institut, Federal Research Institute of Nutrition and Food, Karlsruhe, Germany

<sup>4</sup>Johns Hopkins University, Baltimore, MD, USA

<sup>5</sup>Netherlands Food and Consumer Product Safety Authority, Utrecht, The Netherlands

<sup>6</sup>TNO, Zeist, The Netherlands

<sup>7</sup>Unilever, Vlaardingen, The Netherlands

<sup>8</sup>School of Pharmacy and Biomedical Science, University of Portsmouth, Portsmouth, UK

<sup>9</sup>Department of Nutrition, Faculty of Medicine, Institute of Basic Medical Sciences, University of Oslo, Blindern, Oslo, Norway

<sup>10</sup>Department of Nutrition, Exercise and Sports, Faculty of Science, University of Copenhagen, Frederiksberg C, Denmark

<sup>11</sup>Faculty of Health Sciences, Latrobe University, Bundoora, VIC 3086, Australia

<sup>12</sup>Department of Medicine, University of Melbourne, Saint Vincent's Hospital, VIC 3065, Australia

<sup>13</sup>Sansom Institute of Health Research, University of South Australia, Adelaide, SA 5001, Australia

<sup>14</sup>Department of Gastroenterology and Hepatology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands

<sup>15</sup>Department of Internal Medicine and Gastroenterology, Tergooi, Hilversum, The Netherlands

<sup>16</sup>Biometris, Wageningen University and Research Centre, Wageningen, The Netherlands

(Submitted 14 March 2014 – Final revision received 2 December 2014 – Accepted 19 January 2015 – First published online 8 April 2015)

### Abstract

Fruit and vegetable consumption produces changes in several biomarkers in blood. The present study aimed to examine the dose–response curve between fruit and vegetable consumption and carotenoid ( $\alpha$ -carotene,  $\beta$ -carotene,  $\beta$ -cryptoxanthin, lycopene, lutein and zeaxanthin), folate and vitamin C concentrations. Furthermore, a prediction model of fruit and vegetable intake based on these biomarkers and subject characteristics (i.e. age, sex, BMI and smoking status) was established. Data from twelve diet-controlled intervention studies were obtained to develop a prediction model for fruit and vegetable intake (including and excluding fruit and vegetable juices). The study population in the present individual participant data meta-analysis consisted of 526 men and women. Carotenoid, folate and vitamin C concentrations showed a positive relationship with fruit and vegetable intake. Measures of performance for the prediction model were calculated using cross-validation. For the prediction model of fruit, vegetable and juice intake, the root mean squared error (RMSE) was 258.0 g, the correlation between observed and predicted intake was 0.78 and the mean difference between observed and predicted intake was  $-1.7$  g (limits of agreement:  $-466.3$ ,  $462.8$  g). For the prediction of fruit and vegetable intake (excluding juices), the RMSE was 201.1 g, the correlation was 0.65 and the mean bias was  $2.4$  g (limits of agreement:  $-368.2$ ,  $373.0$  g). The prediction models which include the biomarkers and subject characteristics may be used to estimate average intake at the group level and to investigate the ranking of individuals with regard to their intake of fruit and vegetables when validating questionnaires that measure intake.

**Key words:** Fruits and vegetables: Prediction models: Vitamin C: Folate: Carotenoids

**Abbreviations:** FV, fruit and vegetable intake (excluding juices); FVJ, fruit, vegetable and juice intake; MFP, multivariable fractional polynomials; RMSE, root mean squared error.

\* **Corresponding author:** Dr O. W. Souverein, fax +31 317 482782, email olga.souverein@gmail.com

A high consumption of fruit and vegetables has been associated with a reduced risk of several chronic diseases, including cancer and CVD<sup>(1–3)</sup>. Therefore, intervention studies that aim to increase the consumption of fruit and vegetables using advice or counselling are often conducted. To investigate the success of an intervention, the subjects are asked to report or recall their consumption of fruits and vegetables. However, because it is highly likely that the subject is aware of the intervention (i.e. the advice or counselling), the report or recall is likely to be biased. Objective measures, such as measuring subjects' serum/plasma concentrations of carotenoids, have been used to investigate whether an intervention led to an increase in fruit and vegetable consumption as compared to the control group<sup>(4–6)</sup>, but these biomarkers do not quantify the increase in fruit and vegetable intake caused by the intervention.

The validation of fruit and vegetable intake currently relies on self-reporting instruments. However, self-reported dietary intake instruments are found to be biased and to have correlated errors in comparison to recovery biomarkers, such as doubly labelled water and urinary N excretion<sup>(7–10)</sup>. Therefore, if we were able to quantify fruit and vegetable intake based on biomarkers rather than on self-reporting, the comparison of self-reported intake with this biomarker-based intake estimate would provide us with a better idea of true validity. No recovery biomarker is available for fruit and vegetable intake. Therefore, it would be useful to find a predictive biomarker that can be related to the true intake of fruits and vegetables<sup>(11,12)</sup>.

It is not accurate to relate, for instance, an increase in  $\beta$ -carotene concentration with an exact increase in fruit and vegetable consumption. Single biomarkers and the sum of carotenoids have previously been shown to have low correlations with self-reported intakes of fruits and vegetables<sup>(13–21)</sup>. Therefore, in order to ascertain the full range of fruit and vegetable intake, it is worthwhile to investigate whether a combination of biomarkers, possibly in combination with other factors, provides more reliable results. Baldrick *et al.*<sup>(22)</sup> found that the carotenoids and vitamin C are the most consistently responsive biomarkers for fruit and vegetable intake. In addition, serum/plasma folate may be used as a biomarker of fruit and vegetable intake, even though it is a less sensitive marker, especially in countries where fortification with folate is mandatory<sup>(23,24)</sup>. In order to be able to use biomarkers to quantify the consumption of fruits and vegetables, the dose–response relationship between fruit and vegetable intake and the respective biomarkers must be present. Because dietary intake recorded by subjects is often biased, a cross-sectional study with such data will not provide us with an unbiased estimate of the dose–response curve. In contrast, for diet-controlled intervention studies in which fruits and vegetables are provided to the participants, the intake data does not rely solely on self-reporting. In these studies, the combination of information about the amounts provided, information from supervised consumption and self-reported information on compliance may lead to a less biased estimate of fruit and vegetable intake. We therefore conducted an individual participant data meta-analysis of such studies, covering a wide range of fruit and vegetable intakes. The first aim of the present

study was to investigate the dose–response curve between fruit and vegetable consumption and multiple biomarkers, namely, serum carotenoids ( $\alpha$ -carotene,  $\beta$ -carotene,  $\beta$ -cryptoxanthin, lycopene, lutein and zeaxanthin), serum/plasma folate and serum/plasma vitamin C. The second aim was to establish a prediction model of fruit and vegetable intake based on these biomarkers which may be used as a predictive biomarker or to estimate group-level intake.

## Methods

### Search strategy

The aim of the literature search was to find diet-controlled intervention studies (i.e. food provision studies or partly supervised feeding studies) conducted with adult subjects in which reports on the amount of consumed fruits and vegetables were supported by information on the amounts provided and in which significant efforts were made to maximise compliance. The following diet-controlled intervention studies were included: (1) studies in which all foods and drinks were provided to the subjects during the intervention, and (2) studies in which all fruits and vegetables consumed were provided to the subjects. In addition, carotenoids or folate concentrations in the blood after intervention were measured, and papers were published in the English language. The search was conducted in Scopus, in Pubmed and by a manual search of reference lists. Search terms in the titles and abstracts included 'fruit' and 'vegetables' combined with 'intervention', 'trial' and 'feeding study'. These terms were then combined with 'biomarkers', 'biological markers', 'carotenoids', ' $\alpha$ -carotene', ' $\beta$ -carotene', ' $\beta$ -cryptoxanthin', 'zeaxanthin', 'lycopene', 'lutein', 'folate' and 'bioavailability'. The search included studies published before October 2012.

Papers were first screened based on their titles and abstracts. Then, the full text of the papers that were considered potentially relevant were read and judged for relevancy. Next, the full text of the papers was retrieved and judged using inclusion and exclusion criteria. The exclusion criteria were: (1) intervention studies in which the intervention consisted of dietary advice or counselling (and therefore foods were not provided to the subjects by the investigators); (2) intervention studies in which not all fruits and vegetables were provided (i.e. the provision consisted of additional fruits and vegetables on top of normal fruit and vegetable consumption) or in which fruits and vegetables were provided as supplements (e.g. capsules), juices or extracts; (3) intervention studies in which the intervention involved a single ingestion of the intervention food(s) or an intervention period of 6 d or fewer; and (4) studies that were conducted in children, adolescents, institutionalised elderly or pregnant or lactating women.

### Data

The current contact details of each study's corresponding author, first author or other authors were searched on the Internet.

Authors were contacted by email and asked whether they were willing to send the original data of the study. These authors were offered a co-authorship on the present paper. We requested individual participant data (where available) of subject characteristics (sex, age, height, weight (or BMI) and smoking status), serum/plasma values of biomarkers and intake of fruits and vegetables (or intervention group coding).

In addition, we collected information on: (1) the study design (whether it was a parallel or crossover study, whether a run-in period was included and, where applicable, whether a wash-out period was included); (2) the dietary intervention (the duration of the dietary intervention and the daily intake of fruits and vegetables, carotenoids or folate); and (3) the serum/plasma measurements (whether blood was drawn after a fasting period and which methods were used for sample analysis).

### Statistical analysis

Outliers, which were defined as all observations above  $(Q3 + 4 \times IQR)$  (where  $Q3$  refers to the third quartile and  $IQR$  refers to the interquartile range), were removed from the dataset. The median number of outliers per biomarker was 1 (range: 0–7).

**Dose–response curves.** The dose–response curve between log-transformed biomarker concentrations (dependent variable) and fruit and vegetable intake (independent variable) and between biomarker concentrations and the corresponding micronutrient was estimated using fractional polynomials<sup>(25,26)</sup>. To account for the one crossover study and for between-study heterogeneity, the final parameter estimates were calculated using mixed models with study and subjects as random effects. Therefore, the estimated variance components refer to differences between studies, differences between individuals (to account for the crossover study) and residual variance.

To obtain predictions on the original scale rather than on the logarithmic scale, we applied the following back-transformation:

$$E(Y) = \exp\left(\beta_0 + \sum_{k=1}^p \beta_k X_k + \frac{1}{2} \sigma^2\right),$$

where  $Y$  is the biomarker concentration on the original scale,  $E(Y)$  is the expectation of  $Y$ ,  $X$  is the fruit and vegetable intake,  $\beta$  refers to the regression coefficients of the dose–response model and  $\sigma^2$  is the sum of the variance components estimated in the mixed model.

Several covariates were tested to see whether they statistically significantly predicted the biomarker concentrations. Covariates that were tested included age, BMI, sex and smoking. In addition, the interaction between fruit and vegetable intake and these covariates was tested. The covariates and interactions were tested by including them one at a time in separate fractional polynomial regression models.

**Prediction models of fruit and vegetable intake.** We developed three different prediction models based on what we learned from the dose–response curves. The models were estimated using linear regression: (1) a pre-specified model in which all continuous variables were added as linear terms,

(2) a pre-specified model in which the shape of all continuous variables was established using multivariable fractional polynomials (MFP; referred to as the MFP model), and (3) a reduced model that included only the statistically significant predictors which were selected using MFP (referred to as the reduced MFP model). The MFP models were analysed using STATA/SE version 11.0 for Windows. Interactions between the subject characteristics (age, BMI, sex and smoking status) and the biomarkers ( $\alpha$ -carotene,  $\beta$ -carotene, lutein + zeaxanthin, lycopene and  $\beta$ -cryptoxanthin) were tested for inclusion in the model in four separate models (including (1) main effects + age  $\times$  biomarkers; (2) main effects + BMI  $\times$  biomarkers; (3) main effects + sex  $\times$  biomarkers; and (4) main effects + smoking status  $\times$  biomarkers). All interactions were included as linear terms. Interactions with  $P < 0.05$  were considered relevant for inclusion in the prediction model. These interactions were then tested together in the model, and a backward selection was applied until all interactions included in the model had a  $P$  value of  $< 0.05$ .

Because data on predictors and outcomes were not complete, we used a multiple imputation approach in which ten multiple imputed datasets were created. The power and selection of the predictors was established in all ten imputed datasets separately, and the final model was established by majority voting<sup>(27)</sup>.

The validation of the fruit, vegetable and juice intake (FVJ) and fruit and vegetable intake (excluding juices; FV) prediction models was assessed using tenfold cross-validation. First, the data was imputed as it was earlier, and then the data was randomly separated into ten parts. One part was left out to construct the training set (i.e. the remaining nine parts), and the prediction models were fitted to each of the imputed datasets using linear regression models. The regression coefficients were combined using normal procedures to obtain the regression coefficients for the test data. The out-of-sample data (the test set) was used to calculate the predicted values for each individual by multiplying the regression coefficients and the observed values of the predictors in each of the imputed test sets. The final predicted values were calculated by averaging the predicted values over the ten imputed test sets. Each of the parts was left out once, so the procedure was repeated ten times. These predicted values were compared to the observed values as an estimate of the model performance using three different measures: (1) the root mean squared error (RMSE) =  $\sqrt{1/n \sum (Y - \hat{Y})^2}$ , (2) the correlation between observed intake and predicted intake, and (3) the mean difference (observed intake minus predicted intake) with the corresponding limits of agreement at the individual level (i.e. mean difference  $\pm 1.96 \times SD_{\text{difference}}$ ). Unless otherwise indicated, all analyses were performed using SAS version 9.2 (SAS Institute, Inc.).

## Results

### Search and data retrieval

A total of 1002 studies were found of which twenty-seven qualified for inclusion in the present meta-analysis<sup>(28–54)</sup>. Of these twenty-seven papers, eight publications described a study

population that was also involved in another publication. Therefore, the authors of a total of nineteen unique diet-controlled intervention studies were contacted for cooperation in retrieving individual data. The flowchart of the selection of studies is shown in Fig. 1. A total of twelve authors responded positively to the request and made their data available for the present analysis. A summary of study characteristics of these studies is given in Table 1, and an overview of the data of these studies is presented in Tables 2 and 3. The data of four studies were unfortunately unavailable, and three authors did not respond to our request. Information from these studies is available in online supplementary Table SA.

For six studies, specific groups were not useful in the present analysis<sup>(36,38,41,49,50,52)</sup>, and for one study<sup>(44)</sup>, data of a subset of participants was received. For the study by Miller *et al.*<sup>(44)</sup>, intake of fruits and vegetables in serves was converted to g/d by multiplying the number of serves by 80g. For the study by Itsiopoulou *et al.*<sup>(40)</sup>, intake of fruits and

vegetables was known for fifteen subjects. For the remaining twelve subjects, vegetable intake was imputed as the mean of the intake reported in the paper (i.e. 466 g/d vegetables and 162 g/d fruits). Where necessary,  $\alpha$ -carotene,  $\beta$ -carotene and lycopene were converted from  $\mu\text{g/ml}$  to  $\mu\text{mol/l}$ .

### Dose-response analysis

The estimated dose-response curves between the different biomarkers and FVJ are shown in Fig. 2, and the dose-response curves between the biomarkers and FV are shown in Fig. 3. All biomarkers show a positive dose-response relationship with fruit and vegetable intake. The regression equations that were obtained are shown in online supplementary Table SB.

The *P* values of the covariate and interaction analyses are shown in online supplementary Table SC. Age and smoking were significant predictors for all carotenoids but not for plasma folate. BMI was a significant predictor for  $\alpha$ -carotene,

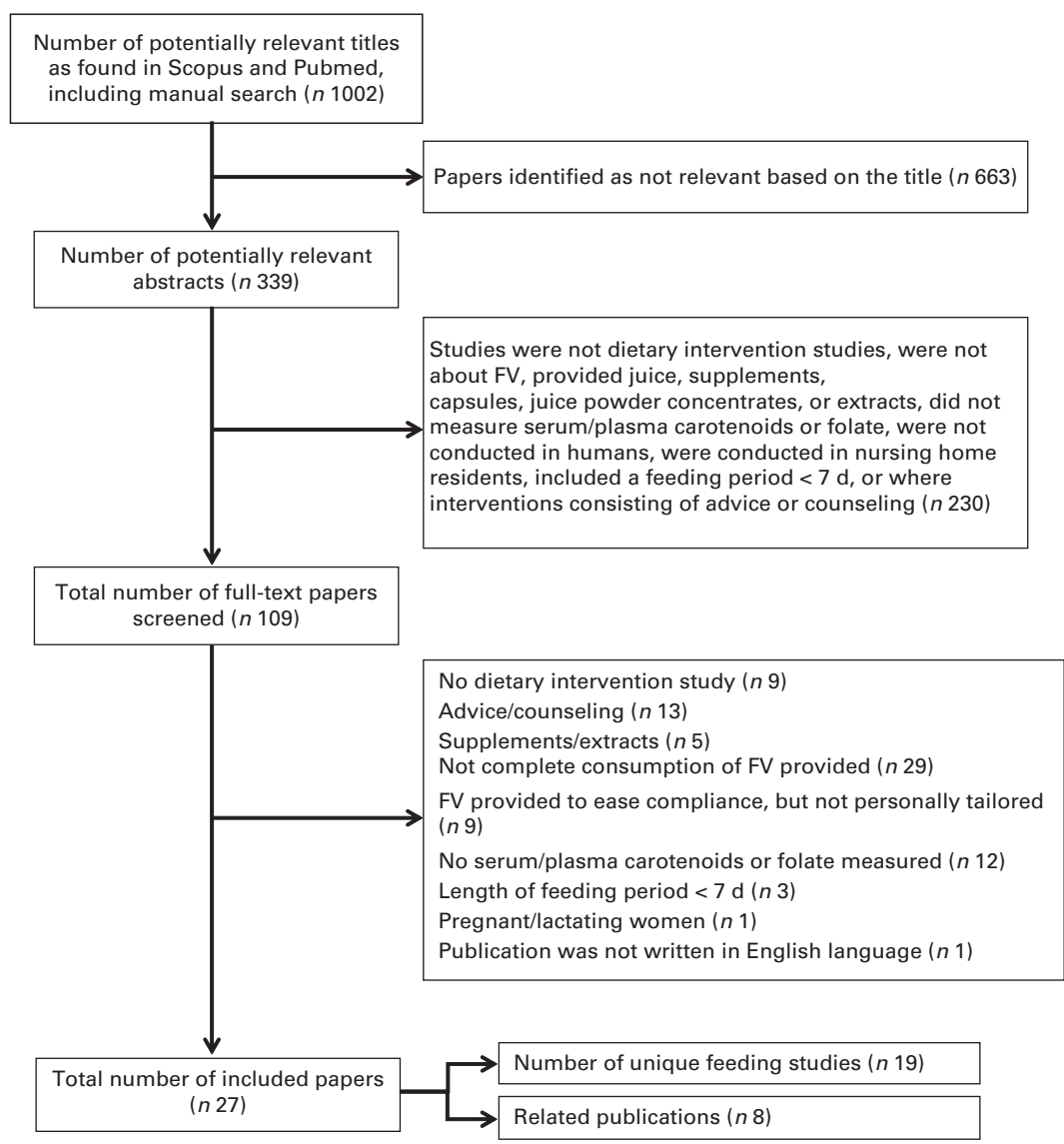


Fig. 1. Flow diagram of study selection process.

**Table 1.** Overview of study characteristics of included studies

| Author  | Year | n*      | Study design and dietary intervention                                 | Checks on compliance/intake   | Duration (d) | F&V intake of included groups (g/d)   |                                 |                                |                                 |                                 |
|---|------|---------|---|---|--------------|---|---------------------------------|--------------------------------|---------------------------------|---------------------------------|
|   |      |         |   |   |              | Group†  | FV                              |                                | FVJ                             |                                 |
|   |      |         |   |   |              |   | Mean                            | SD                             | Mean                            | SD                              |
| Broekmans <i>et al.</i> <sup>(33)</sup>   | 2000 | 47 (47) | Complete diet; parallel intervention                                  | Evening meal under supervision at the institute, remaining parts were weighed and recorded. The remainder of the daily diet was handed out to the volunteers. Consumption was checked by a questionnaire.   | 28           | A: Low (P)<br>B: High (P)   | 100<br>565                      |                                | 100<br>765                      |                                 |
| Castenmiller <i>et al.</i> <sup>(35,36)</sup>                                   | 1999 | 58 (72) | Complete diet with list of free choice; parallel intervention         | Subjects received a hot meal at the university, and foods for their other meals and snacks were packed to be taken home. The daily selection of free choice foods was recorded in diaries.  | 21           | A: Control (P)<br>B: Whole leaf spinach (P)<br>C: Minced spinach (P)<br>D: Liquefied spinach (P)<br>E: Liquefied spinach plus dietary fibre (P) | 491<br>484<br>471<br>473<br>468 | 137<br>117<br>108<br>100<br>90 | 728<br>722<br>712<br>711<br>711 | 172<br>146<br>135<br>129<br>122 |
| Chopra <i>et al.</i> <sup>(37)</sup>  | 2000 | 34 (32) | F&V provided; crossover intervention                                  | Participants were provided with food items. Most of these were consumed during lunch at the university during the weekdays. Researchers relied on participants for extra consumption during the rest of the day and on weekends.  | 7            | A: Red week (P)<br>B: Green week (P)  | 350<br>350                      |                                | 350<br>350                      |                                 |
| Dragsted <i>et al.</i> <sup>(38)</sup> ; Moller <i>et al.</i> <sup>(46)</sup>   | 2003 | 31 (43) | Complete diet; parallel intervention                                  | All of the food was provided free of charge throughout the intervention. In addition, plasma α- and β-carotene and ascorbate were used as markers to assure that the groups differed.   | 24           | A: Fruveg (P)<br>B: Placebo (P)   | 480<br>0                        |                                | 600<br>0                        |                                 |
| Freese <i>et al.</i> <sup>(39)</sup> ; Misikangas <i>et al.</i> <sup>(45)</sup> | 2001 | 77 (77) | Complete diet with list of free choice; parallel intervention         | During the intervention, food consumption was controlled by serving lunch at the department on weekdays and by asking the volunteers to mark in their study diaries if any study foods were not eaten. Also, biomarkers were used to check compliance.  | 42           | A: PUFA – low FBV (P)<br>B: PUFA – high FBV (P)<br>C: MUFA – low FBV (P)<br>D: MUFA – high FBV (P)  | 217<br>807<br>235<br>809        | 32<br>166<br>51<br>138         | 505<br>1057<br>549<br>1059      | 73<br>217<br>119<br>181         |
| Itsiopoulos <i>et al.</i> <sup>(40)</sup>                                       | 2011 | 27 (27) | Diet provided in excess of intake; crossover intervention             | Compliance was checked with 7 d diet diaries, and participants were interviewed every 2 weeks when they returned to pick up supplies of food. Participants were asked to select the foods they ate during the previous 2 weeks from a booklet. Plasma fatty acids, carotenoids and body weight were measured as markers of compliance.  | 84           | Mediterranean diet (R)  | 768                             | 216                            | 768                             | 216                             |
| Bøhn <i>et al.</i> <sup>(29)</sup> ; Karlsen <i>et al.</i> <sup>(41)</sup>      | 2010 | 33 (33) | Diet provided in excess of energy requirements; parallel intervention | A detailed questionnaire was completed at each weekly follow-up to record compliance. All participants were instructed to bring the remaining food items to the weekly follow-up. Individual counselling was given to the participants to help them consume the provided food items. Dietary intake during the intervention period was recorded using a 7 d food record with a picture book, and it was completed during the final week of the intervention period. | 56           | Antioxidant-rich diet (R)   | 525                             | 242                            | 1491                            | 509                             |

Table 1. Continued

| Author   | Year | n*       | Study design and dietary intervention                          | Checks on compliance/intake  | Duration (d) | F&V intake of included groups (g/d)                         |      |     |      |     |     |
|--|------|----------|--|--|--------------|---|------|-----|------|-----|-----|
|  |      |          |  |  |              | Group†  | FV   |     | FVJ  |     |     |
|  |      |          |  |  |              |   | Mean | SD  | Mean | SD  |     |
| Miller III <i>et al.</i> <sup>(44)</sup>                                       | 2005 | 60 (103) | Complete diet; parallel intervention                           | Meals were prepared in a metabolic kitchen and served in an outpatient dining facility. Throughout the 3 months of feeding, participants agreed to eat only the food provided to them and nothing else.  | 90           | A: DASH diet (P)<br>B: Control diet (P)                     | –    | –   | 768  | 288 |     |
| Van het Hof <i>et al.</i> <sup>(49)‡</sup>                                     | 1999 | 43 (54)  | Complete diet with list of free choice; parallel intervention  | Hot meals were provided at lunchtime under supervision from Monday to Friday. Other foods consumed during these days and during the weekends were eaten at home, and compliance was checked via diaries. Volunteers were carefully instructed on how to prepare the foods.                       | 28           | A: Low-vegetable diet (P)<br>B: High-vegetable diet (P)     | 255  | 605 | 455  | 805 |     |
| Van Loo-Bouwman <i>et al.</i> <sup>(50)</sup>                                  | 2009 | 24 (24)  | Complete diet with list of free choice; crossover intervention | Hot meals were provided at lunchtime under supervision from Monday to Friday. Other foods consumed during these days and during the weekends were eaten at home, and compliance was checked via diaries.   | 21           | Mixed diet (vegetables and fruits high in β-carotene) (P)   | 329  | 100 | 654  | 182 |     |
| Briviba <i>et al.</i> <sup>(32)</sup> ,<br>Watzl <i>et al.</i> <sup>(51)</sup> | 2005 | 63 (63)  | F&V provided; parallel intervention                            | Each study participant was provided with a box of F&V. F&V which were not consumed during the study period had to be returned. Daily intake of F&V was assessed via a specific F&V protocol throughout the study period. During two 4 d periods, the whole food intake was assessed via diaries. | 28           | A: 2 serves/d (P)<br>B: 5 serves/d (P)<br>C: 8 serves/d (P) | –    | –   | 250  | 565 | 955 |
| Winkels <i>et al.</i> <sup>(52)</sup>  | 2007 | 29 (72)  | Complete diet with list of free choice; parallel intervention  | All foods were provided. Participants were asked to report all free-choice items and any deviations in diaries.  | 28           | Food folate group (P)                                       | 476  | –   | 876  | –   |     |

Predicting fruit and vegetable intake

F&V, fruit and vegetables; FV, fruit and vegetable intake, excluding juices; FVJ, fruit, vegetable and juice intake; FBV, fruit, berries and vegetables.

\* The number of individuals used in the present analysis. In brackets, the number of individuals reported in the original publication. For several studies, specific intervention groups were not useful in the present analysis<sup>(36,38,41,49,50,52)</sup>, and for one study<sup>(44)</sup>, data of a subset of participants was received.

† In brackets, indication of whether the amount of fruits and vegetables reported in the table and used in the analysis was the amount provided to the subjects (indicated by 'P') or whether the amount relied partly on self-reporting (indicated by 'R').

‡ The folate data of that study were no longer available<sup>(34)</sup>.

**Table 2.** Baseline characteristics of the included studies (Mean values and standard deviations)

| Study   | n   | Age (years) |      | BMI (kg/m <sup>2</sup> ) |     | Sex (% male) | Smoking (%) | Plasma folate (nmol/l) |     | Vitamin C (μmol/l) |      |
|---|-----|-------------|------|--------------------------|-----|--------------|-------------|------------------------|-----|--------------------|------|
|   |     | Mean        | SD   | Mean                     | SD  |              |             | Mean                   | SD  | Mean               | SD   |
| Broekmans <i>et al.</i> <sup>(33)</sup>   | 47  | 49.3*       | 5.1  | 25.7                     | 3.1 | 51.1         | 25.5        | 13.7                   | 7.1 | 49.4               | 18.6 |
| Castenmiller <i>et al.</i> <sup>(35,36)</sup>                                   | 58  | 22.8        | 7.7  | 22.1                     | 2.1 | 39.7         | 0           | 15.3                   | 4.2 | –                  | –    |
| Chopra <i>et al.</i> <sup>(37)</sup>  | 34  | 37.2        | 8.7  | –                        | –   | 0            | –           | –                      | –   | –                  | –    |
| Dragsted <i>et al.</i> <sup>(38)</sup> ; Moller <i>et al.</i> <sup>(46)</sup>   | 31  | 27.3        | 7.3  | 23.1                     | 2.3 | 48.4         | 0           | 10.8                   | 4.0 | –                  | –    |
| Freese <i>et al.</i> <sup>(39)</sup> ; Misikangas <i>et al.</i> <sup>(45)</sup> | 77  | 25.1        | 6.6  | 22.6                     | 3.2 | 26.0         | 5.2         | 10.0                   | 4.1 | 51.9               | 16.5 |
| Itsiopoulos <i>et al.</i> <sup>(40)</sup>                                       | 27  | 59.1        | 7.1  | 30.2                     | 3.7 | 59.3         | –           | –                      | –   | –                  | –    |
| Bøhn <i>et al.</i> <sup>(29)</sup> ; Karlsen <i>et al.</i> <sup>(41)</sup>      | 33  | 56.7        | 6.4  | 24.8                     | 2.7 | 100          | 100         | –                      | –   | 46.7               | 17.0 |
| Miller III <i>et al.</i> <sup>(44)</sup>  | 60  | 52.0*       | 10.0 | 29.6*                    | 4.4 | 44*          | 14*         | –                      | –   | –                  | –    |
| Van het Hof <i>et al.</i> <sup>(49)</sup>                                       | 43  | 22.4        | 6.4  | 22.4                     | 2.1 | 27.9         | 0           | –                      | –   | 66.6               | 17.4 |
| Van Loo-Bouwman <i>et al.</i> <sup>(50)</sup>                                   | 24  | 22.0        | 4.0  | 21.8                     | 2.2 | 41.7         | 0           | –                      | –   | –                  | –    |
| Briviba <i>et al.</i> <sup>(32)</sup> ; Watzl <i>et al.</i> <sup>(51)</sup>     | 63  | 31.2        | 9.0  | 23.7                     | 2.7 | 100          | 0           | –                      | –   | 83.7               | 16.6 |
| Winkels <i>et al.</i> <sup>(52)</sup>   | 29  | 23.3        | 4.8  | 22.6                     | 2.8 | 24.1         | 13.8        | 12.1                   | –   | –                  | –    |
| Total population  | 526 | 30.9        | 13.8 | 23.6                     | 3.4 | 47.9         | 13.1        | 12.1                   | 5.2 | 60.8               | 22.2 |

\* These data are taken from the original publication, but they were not available for the present analysis.

β-carotene, lutein, β-cryptoxanthin and lycopene. Sex was only a significant predictor for lutein, zeaxanthin and lutein+zeaxanthin. The interactions between these covariates and the intake of fruits and vegetables were relevant ( $P < 0.1$ ) in most instances. The smoking × fruit and vegetable interaction was only a significant predictor for about half of the biomarkers, but this may be a result of the relatively low number of smokers included in the present sample.

Where possible, the dose–response relationship between the biomarkers and the intake of the micronutrient was also investigated (online supplementary Fig. SA). The available sample size was largest for β-carotene ( $n$  316) and smallest for lutein + zeaxanthin ( $n$  35). The sample size of zeaxanthin was too low to warrant analysis. All curves showed a positive relationship between intake and serum or plasma concentrations except

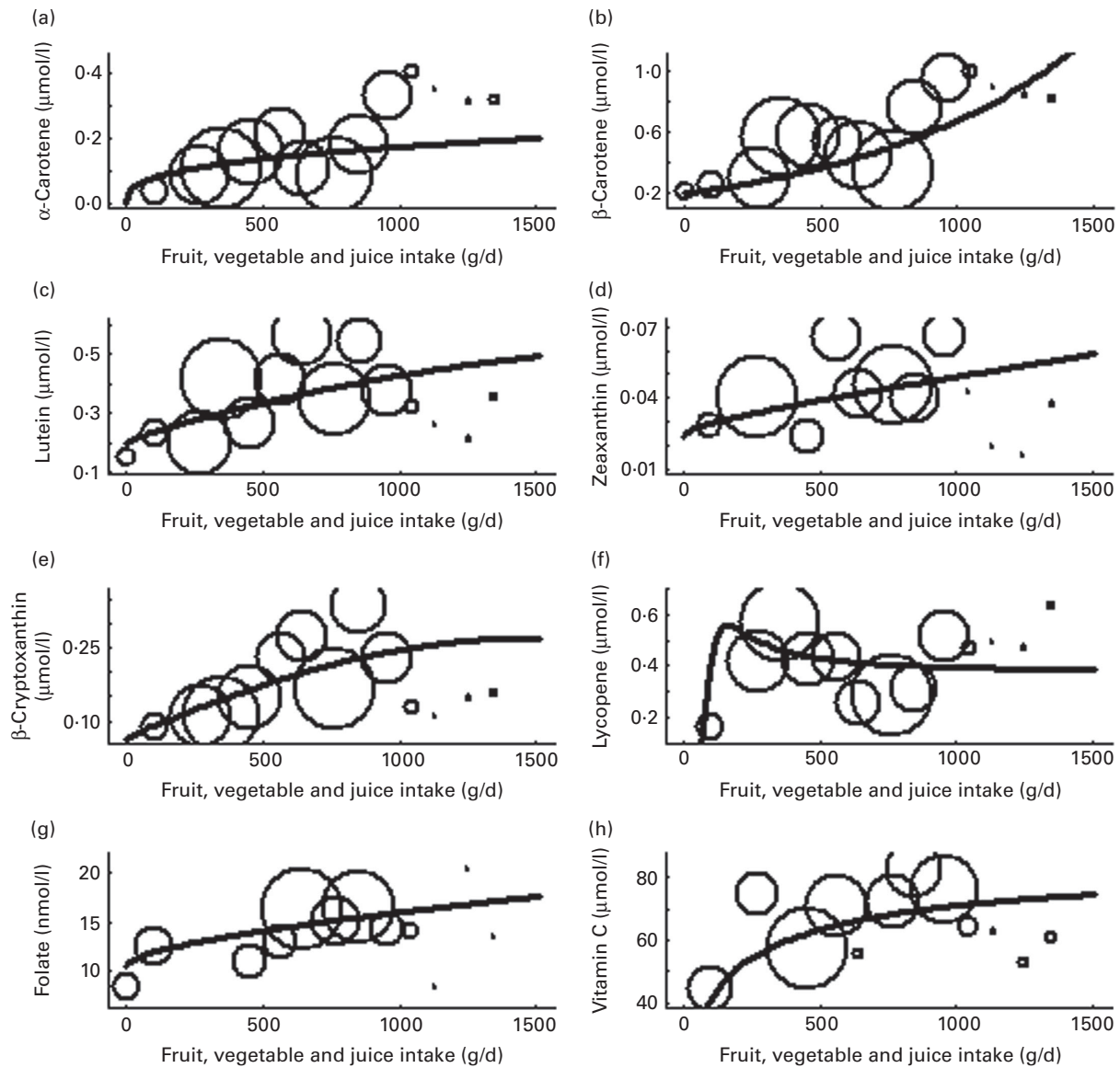
lutein at high intakes. There is no biological evidence for the drop that is visible in the lutein curve. Because there were very few data available for lutein intake of more than 15 mg/d, this part of the curve is not considered reliable.

#### Prediction model

The regression coefficients of the final prediction model are presented in Table 4, and the performance measures are presented in Table 5. The power and variable selection process of the MFP and the reduced MFP model is shown in online supplementary Tables SD and SE. For FVJ, the reduced MFP model showed the lowest RMSE (i.e. 258.0 g) and the highest correlation between observed and predicted (i.e. 0.78) as compared to the linear model and the full pre-specified MFP

**Table 3.** Baseline characteristics of the included studies (Mean values and standard deviations)

| Study   | α-Carotene (μmol/l) |      | β-Carotene (μmol/l) |      | β-Cryptoxanthin (μmol/l) |      | Lycopene (μmol/l) |      | Lutein (μmol/l) |      | Zeaxanthin (μmol/l) |      | Lutein + zeaxanthin (μmol/l) |      |
|---|---------------------|------|---------------------|------|--------------------------|------|-------------------|------|-----------------|------|---------------------|------|------------------------------|------|
|   | Mean                | SD   | Mean                | SD   | Mean                     | SD   | Mean              | SD   | Mean            | SD   | Mean                | SD   | Mean                         | SD   |
| Broekmans <i>et al.</i> <sup>(33)</sup>   | 0.06                | 0.04 | 0.40                | 0.21 | 0.14                     | 0.10 | 0.45              | 0.27 | 0.34            | 0.15 | 0.04                | 0.03 | 0.37                         | 0.16 |
| Castenmiller <i>et al.</i> <sup>(35,36)</sup>                                   | 0.07                | 0.03 | 0.27                | 0.14 | 0.28                     | 0.12 | 0.19              | 0.10 | 0.22            | 0.07 | 0.03                | 0.01 | 0.25                         | 0.08 |
| Chopra <i>et al.</i> <sup>(37)</sup>  | 0.10                | 0.07 | 0.38                | 0.29 | 0.13                     | 0.09 | 0.34              | 0.16 | 0.23            | 0.10 | –                   | –    | –                            | –    |
| Dragsted <i>et al.</i> <sup>(38)</sup> ; Moller <i>et al.</i> <sup>(46)</sup>   | –                   | –    | 0.36                | 0.23 | –                        | –    | –                 | –    | 0.26            | 0.12 | –                   | –    | –                            | –    |
| Freese <i>et al.</i> <sup>(39)</sup> ; Misikangas <i>et al.</i> <sup>(45)</sup> | 0.20                | 0.10 | 0.60                | 0.30 | 0.10                     | 0.05 | 0.62              | 0.19 | 0.26            | 0.10 | –                   | –    | –                            | –    |
| Itsiopoulos <i>et al.</i> <sup>(40)</sup>                                       | 0.08                | 0.05 | 0.31                | 0.20 | 0.16                     | 0.14 | 0.43              | 0.20 | –               | –    | –                   | –    | 0.35                         | 0.13 |
| Bøhn <i>et al.</i> <sup>(29)</sup> ; Karlsen <i>et al.</i> <sup>(41)</sup>      | 0.07                | 0.06 | 0.35                | 0.29 | 0.15                     | 0.10 | 0.56              | 0.26 | 0.16            | 0.07 | 0.04                | 0.02 | 0.20                         | 0.08 |
| Miller III <i>et al.</i> <sup>(44)</sup>  | 0.05                | 0.05 | 0.23                | 0.13 | 0.07                     | 0.04 | 0.28              | 0.15 | 0.16            | 0.06 | 0.04                | 0.02 | 0.19                         | 0.07 |
| Van het Hof <i>et al.</i> <sup>(49)</sup>                                       | 0.08                | 0.06 | 0.40                | 0.19 | 0.34                     | 0.21 | 0.27              | 0.12 | 0.17            | 0.07 | 0.04                | 0.02 | 0.20                         | 0.09 |
| Van Loo-Bouwman <i>et al.</i> <sup>(50)</sup>                                   | 0.10                | 0.06 | 0.75                | 0.36 | 0.34                     | 0.14 | –                 | –    | –               | –    | –                   | –    | –                            | –    |
| Briviba <i>et al.</i> <sup>(32)</sup> ; Watzl <i>et al.</i> <sup>(51)</sup>     | 0.13                | 0.08 | 0.55                | 0.31 | 0.23                     | 0.12 | 0.55              | 0.25 | 0.26            | 0.10 | 0.06                | 0.02 | 0.33                         | 0.14 |
| Winkels <i>et al.</i> <sup>(52)</sup>   | –                   | –    | –                   | –    | –                        | –    | –                 | –    | –               | –    | –                   | –    | –                            | –    |
| Total population  | 0.10                | 0.08 | 0.42                | 0.28 | 0.18                     | 0.15 | 0.42              | 0.24 | 0.23            | 0.11 | 0.04                | 0.02 | 0.27                         | 0.13 |



**Fig. 2.** Dose–response curves between serum carotenoids ((a)  $\alpha$ -carotene, (b)  $\beta$ -carotene, (c) lutein, (d) zeaxanthin, (e)  $\beta$ -cryptoxanthin, (f) lycopene), (g) plasma/serum folate and (h) vitamin C and fruit, vegetable and juice intake. The  $\circ$  indicate the individual data points, and their sizes are proportional to the number of individuals for each specific intake (i.e. the larger the circle, the more individuals were available for analysis).

model. The mean difference of the reduced MFP model ( $-1.7$  g) was slightly higher than those of the other two models (linear model:  $-1.6$  g; MFP model:  $-1.5$  g), but the limits of agreement were markedly smaller than those of the other two models. Bland–Altman plots are presented in online supplementary Fig. SB.

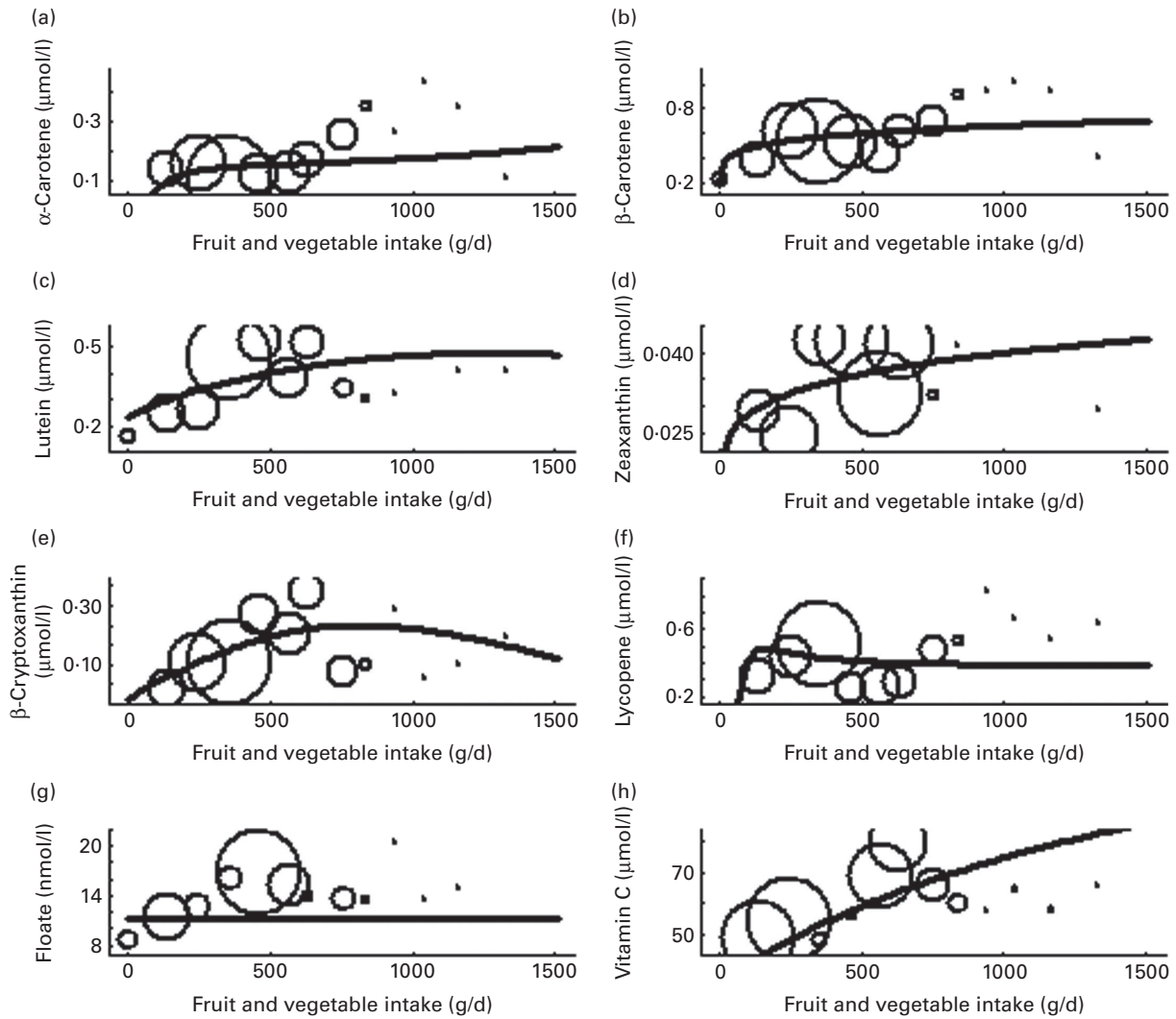
For FV, the MFP model was the best model. It showed the lowest RMSE (201.1 g), the highest correlation (0.65) and the lowest mean bias (2.4 g) with the smallest limits of agreement ( $-368.2$ , 373.0 g).

The prediction model for FV showed a somewhat lower correlation and a higher absolute mean difference than the model for fruit and vegetable intake including juices. Therefore, we investigated whether a model including a predictor variable that represented juice intake (in g/d) would improve the prediction for fruit and vegetable intake when juices were

excluded. However, this did not markedly change the results. The MFP model including juice as a predictor variable had an RMSE of 202.8 g, a correlation of 0.64, a mean bias of 0.2 g (limits of agreement:  $-374.1$ , 374.6 g). Therefore, the simpler model without juice as a predictor variable is preferred as a prediction model for FV.

In order to compare the performance of the prediction model with the current practice of using the sum of carotenoids or any single biomarker, we calculated the correlation coefficients between the observed intakes and the sum of carotenoids and those between observed intakes and single biomarkers (Table 6). For FVJ, the correlations ranged between 0.04 and 0.32, which was much lower than the 0.65 in the prediction model. Also for FV, the correlations (between 0.15 and 0.38) were lower than that in the prediction model (0.64).





**Fig. 3.** Dose–response curves between serum carotenoids ((a)  $\alpha$ -carotene, (b)  $\beta$ -carotene, (c) lutein, (d) zeaxanthin, (e)  $\beta$ -cryptoxanthin, (f) lycopene), (g) plasma/serum folate and (h) vitamin C and fruit and vegetable intake (excluding juices). The  $\circ$  indicate the individual data points, and their sizes are proportional to the number of individuals for each specific intake (i.e. the larger the circle, the more individuals were available for analysis).

To indicate the value of the prediction model for individual studies, an additional cross-validation was performed by leaving one entire study out of the training set. The study that was left out comprised the test set. Table 7 shows the RMSE and mean difference with the limits of agreement for the reduced MFP model for FVJ and the MFP model for FV. These show that there is a difference between how well the prediction models perform in each study. The study by Karlsen *et al.*<sup>(41)</sup> shows a worse performance for FVJ but not for FV. This is most likely caused by the relatively high intake of fruits, vegetables and juices in that study (see Table 1).

### Discussion

The first part of the present research showed that all investigated biomarkers (carotenoids and folate) had a positive relationship with fruit and vegetable intake, and they are therefore useful for predicting fruit and vegetable intake. Several covariates were significantly associated with the

biomarkers. The next aim was to develop a prediction model for fruit and vegetable intake based on objective variables, such as biomarkers and subject characteristics. Among the three models for predicting FVJ that were investigated, the reduced MFP model showed the best performance in cross-validation, and the MFP model showed the best performance for FV.

The sum of carotenoids has been used in an attempt to combine biomarkers into a single estimate for fruit and vegetable intake in various studies. The sum of carotenoids was positively correlated with self-reported fruit and vegetable intake<sup>(14–21,55,56)</sup>. In the present study, the correlations between the predicted values, which can easily be calculated in future research by multiplying observed values from biomarkers and subject characteristics with the corresponding  $\beta$  coefficients from Table 4 and then adding these together, and the observed fruit and vegetable intake (both including and excluding juices) was markedly higher than the correlations between the observed intakes and the sum of

**Table 4.** The predictors on the multiple completed\* datasets (*n* 492† in each completed dataset) from a linear regression analysis

(Regression coefficients, standard errors, and powers)

| Predictors                                   | Linear model |       | MFP model |       |       | Reduced MFP model |       |       |
|--|--------------|-------|-----------|-------|-------|-------------------|-------|-------|
|  | $\beta$      | SE    | $\beta$   | SE    | Power | $\beta$           | SE    | Power |
| <b>FVJ</b>                                   |              |       |           |       |       |                   |       |       |
| Constant                                     | -172.8       | 158.9 | -1691.4   | 526.9 | -     | 1043.2            | 180.0 | -     |
| $\alpha$ -Carotene ( $\mu\text{mol/l}$ )     | 479.8        | 142.2 | 607.8     | 133.4 | 0.5   | 674.1             | 90.1  | 0.5   |
| $\beta$ -Carotene ( $\mu\text{mol/l}$ )      | 123.1        | 53.1  | 101.5     | 50.9  | 1     | -                 | -     | -     |
| Lutein + zeaxanthin ( $\mu\text{mol/l}$ )    | 193.2        | 68.8  | 154.6     | 70.6  | 1     | -153.7            | 36.8  | -0.5  |
| $\beta$ -Cryptoxanthin ( $\mu\text{mol/l}$ ) | 162.1        | 138.5 | 141.2     | 138.3 | 1     | -                 | -     | -     |
| Lycopene ( $\mu\text{mol/l}$ )               | -13.8        | 87.4  | -78.0     | 82.2  | 1     | -                 | -     | -     |
| Folate (nmol/l)‡                             | 158.9        | 38.9  | 49.9      | 11.1  | 2     | 48.9              | 10.9  | 2     |
| Vitamin C ( $\mu\text{mol/l}$ )              | 0.91         | 0.93  | 0.78      | 0.96  | 1     | -                 | -     | -     |
| BMI ( $\text{kg/m}^2$ )                      | 7.6          | 7.9   | 10.2      | 7.2   | 1     | -                 | -     | -     |
| Female sex                                   | -40.2        | 27.3  | -55.3     | 28.1  | -     | -63.5             | 29.2  | -     |
| Age (years)§                                 | 39.4         | 24.4  | -1711.6   | 596.0 | 0     | -992.9            | 341.0 | 0     |
|  |              |       | 1982.9    | 676.6 | 0.5   | 470.2             | 149.4 | 0     |
| Smoking                                      | -367.4       | 248.6 | -278.6    | 195.3 | -     | -232.2            | 187.4 | -     |
| Smoking $\times$ folate                      | 38.1         | 13.7  | 31.3      | 10.5  | 1     | 28.4              | 10.3  | 1     |
| <b>FV</b>                                    |              |       |           |       |       |                   |       |       |
| Constant                                     | -274.2       | 166.5 | -304.9    | 164.2 | -     | -85.5             | 141.5 | -     |
| $\alpha$ -Carotene ( $\mu\text{mol/l}$ )     | 939.2        | 205.0 | 830.9     | 219.9 | 1     | -                 | -     | -     |
| $\beta$ -Carotene ( $\mu\text{mol/l}$ )      | 104.1        | 45.9  | 95.4      | 45.1  | 1     | 300.2             | 65.2  | 1     |
| Lutein + zeaxanthin ( $\mu\text{mol/l}$ )    | 276.8        | 69.5  | 414.4     | 90.2  | 1     | -158.3            | 29.3  | -0.5  |
|  |              |       | -562.4    | 140.2 | 1     | -                 | -     | -     |
| $\beta$ -Cryptoxanthin ( $\mu\text{mol/l}$ ) | 146.1        | 105.7 | 74.4      | 100.7 | 1     | -                 | -     | -     |
| Lycopene ( $\mu\text{mol/l}$ )               | -764.1       | 306.0 | -782.8    | 295.8 | 1     | -                 | -     | -     |
| Folate (nmol/l)‡                             | 74.0         | 34.7  | 59.6      | 33.0  | 1     | 62.5              | 33.4  | 1     |
| Vitamin C ( $\mu\text{mol/l}$ )              | 1.7          | 0.7   | 1.4       | 0.6   | 1     | 1.6               | 0.6   | 1     |
| BMI ( $\text{kg/m}^2$ )                      | 4.9          | 6.6   | 5.6       | 6.2   | 1     | 16.4              | 3.8   | 1     |
| Female sex                                   | 42.0         | 41.5  | -57.3     | 21.8  | -     | -42.8             | 22.4  | -     |
| Age (years)§                                 | 63.6         | 12.4  | 1.1       | 0.2   | 3     | 53.1              | 14.4  | 1     |
| Smoking                                      | 8.5          | 45.5  | 19.8      | 43.8  | -     | -                 | -     | -     |
| Age $\times$ $\alpha$ -carotene              | -22.0        | 5.3   | -19.1     | 5.4   | 1     | -                 | -     | -     |
| BMI $\times$ lycopene                        | 29.0         | 11.9  | 28.6      | 11.6  | 1     | -                 | -     | -     |
| Sex $\times$ lutein + zeaxanthin             | -215.0       | 82.2  | -         | -     | -     | -                 | -     | -     |
| Age $\times$ $\beta$ -carotene               | -            | -     | -         | -     | -     | -5.0              | 2.1   | 1     |

FVJ, fruit, vegetable and juice intake; FV, fruit and vegetable intake, excluding juices.

\* Completed datasets refers to the data after multiple imputation.

† The study of Chopra *et al.*<sup>(37)</sup> could not be used in the present analysis because of an estimation problem.

‡ Folate is scaled as folate/10.

§ Age is scaled as age/10.

carotenoids or any of the single biomarkers. Despite the model's good performance on average, there was some residual variation as well as an overestimation of low fruit and vegetable intake and an underestimation of high fruit and vegetable intake. Not all fruits and vegetables contain the same concentration of carotenoids and folate, and other foods in the diet also contain these nutrients. Therefore, the type of fruits and vegetables eaten and the diet as a whole

influence the final biomarker concentrations in the blood. The present study tried to capture 'normal' diet effects as much as possible by excluding those studies that provided only a single type of fruit or vegetable and by including intervention arms that focused on carotenoid-rich or folate-rich and carotenoid-poor or folate-poor fruits and vegetables. In order to obtain the large-sample benefits of a meta-analysis, these different study types were grouped together. This was

**Table 5.** Performance measures of the different prediction models as calculated by cross-validation

|                   | FVJ   |             |  |                     | FV    |             |  |                     |
|-------------------|-------|-------------|--|---------------------|-------|-------------|--|---------------------|
|                   | RMSE  | Correlation | Mean difference between observed and predicted | Limits of agreement | RMSE  | Correlation | Mean difference between observed and predicted | Limits of agreement |
| Linear model      | 265.7 | 0.77        | -1.6   | -478.4, 475.2       | 205.6 | 0.64        | 4.4  | -372.3, 381.1       |
| MFP model         | 260.0 | 0.78        | -1.5   | -467.6, 464.7       | 201.1 | 0.65        | 2.4  | -368.2, 373.0       |
| Reduced MFP model | 258.0 | 0.78        | -1.7   | -466.3, 462.8       | 205.2 | 0.61        | 6.8  | -382.3, 396.0       |

FVJ, fruit, vegetable and juice intake; FV, fruit and vegetable intake, excluding juices; RMSE, root mean squared error; MFP, multivariable fractional polynomials.

**Table 6.** Pearson correlations between fruit and vegetable intake and biomarkers

| Biomarker   | FVJ  | FV   |
|---|------|------|
| $\alpha$ -Carotene at follow-up ( $\mu\text{mol/l}$ )             | 0.29 | 0.26 |
| $\beta$ -Carotene at follow-up ( $\mu\text{mol/l}$ )              | 0.27 | 0.24 |
| Cryptoxanthin at follow-up ( $\mu\text{mol/l}$ )                  | 0.08 | 0.16 |
| Lycopene at follow-up ( $\mu\text{mol/l}$ )                       | 0.19 | 0.24 |
| Combined lutein and zeaxanthin at follow-up ( $\mu\text{mol/l}$ ) | 0.08 | 0.15 |
| Sum of carotenoids ( $\mu\text{mol/l}$ )                          | 0.23 | 0.33 |
| Serum/plasma folate at follow-up (nmol/l)                         | 0.32 | 0.26 |
| Serum/plasma vitamin C at follow-up ( $\mu\text{mol/l}$ )         | 0.04 | 0.38 |

FVJ, fruit, vegetable and juice intake; FV, fruit and vegetable intake, excluding juices.

done because a number of studies were included, so we assumed that the applied regression analysis would average out the effects of individual studies and that at least the first approximation would not depend on the types of fruits and vegetables included. Obviously, the assumption is not true in an absolute sense, seeing as carrots, for example, contain more carotenoids than some other vegetables, and this will thus require further investigation.

Another source of variability may come from the different intervention durations. We excluded studies with a duration of less than 7 d because we assumed that it would take approximately 1 week to obtain a new steady state for the carotenoids after the change in diet was induced by the intervention<sup>(57)</sup>. The actual duration of the studies included in the prediction models was much longer (Table 1).

Differences in the analytical methods used in the different studies may be another source of residual variation. In particular, folate levels were analysed using different assays, e.g. immunoassay and radioassay. Also, laboratory variability may be caused by different specimen collection and storage<sup>(58)</sup> techniques, among many other possible sources.

Sex, age, BMI and smoking impact on serum carotenoids, serum vitamin C and plasma folate levels as well as several other covariates, such as serum cholesterol, serum TAG and the consumption of alcohol, fat and energy, may also be related to the biomarkers<sup>(59–63)</sup>. It may be of interest to investigate whether these covariates could significantly improve the

prediction model. However, the present data did not allow us to investigate this thoroughly.

Although significant efforts were made in all individual studies to encourage compliance to the study protocol (e.g. the supervised consumption of meals; see Table 1), the true intake of fruits and vegetables could not always be determined with absolute certainty because they relied on self-reports of compliance. In quite a number of the individual studies, compliance was investigated with, e.g., questionnaires or diaries, and most often this self-reported compliance was high.

Unfortunately, no external validation data was available for the prediction model. We chose to use all of the data from the diet-controlled intervention studies that were available to us to develop the models. To perform an external validation, data from other or new diet-controlled intervention studies would have to be obtained. Because this would be very complicated and because the data from such studies would preferably be used to develop or improve the present model rather than to just validate it, we mimicked independent data by using cross-validation to calculate the measures of performance<sup>(64)</sup>.

The use of individual participant data from diet-controlled intervention studies made it possible to model the dose–response curves and the prediction models for a large range of fruit and vegetable intake with a relatively large number of subjects using a more objective assessment of intake. However, between-study differences may have influenced the study results. In the dose–response analysis, we took clustering into account by using mixed-effects models<sup>(65)</sup>. For the prediction model, the marginal predictions (i.e. using only the fixed effects because the (unknown) random effects cannot be used in predictions for new subjects) from the random intercept linear regression model performed somewhat worse in cross-validation than the predictions from the standard regression model (data not shown), and we therefore chose to present the standard regression model. Bouwmeester *et al.*<sup>(66)</sup> found similar performance measures for a standard logistic regression model and a random intercept logistic regression model in a study on surgical patients that were clustered by anaesthesiologist. Recently, Debray *et al.*<sup>(67)</sup> developed an approach to deal with risk prediction in new

**Table 7.** Performance measures of the best-performing prediction models per study as calculated by cross-validation

|   | FVJ (reduced MFP model) |  |                     | FV (MFP model) |  |                     |
|---|-------------------------|--|---------------------|----------------|--|---------------------|
|   | RMSE                    | Mean difference between observed and predicted | Limits of agreement | RMSE           | Mean difference between observed and predicted | Limits of agreement |
| Broekmans <i>et al.</i> <sup>(33)</sup>   | 340.9                   | –127.9   | –743.2, 487.5       | 209.8          | –88.3  | –457.4, 280.8       |
| Castenmiller <i>et al.</i> <sup>(35,36)</sup>                                   | 188.2                   | 10.1   | –358.4, 378.6       | 126.8          | 17.0   | –224.7, 258.8       |
| Dragsted <i>et al.</i> <sup>(38)</sup> ; Moller <i>et al.</i> <sup>(46)</sup>   | 303.4                   | –198.9   | –631.7, 233.9       | 191.9          | –80.1  | –407.9, 247.6       |
| Freese <i>et al.</i> <sup>(39)</sup> ; Misikangas <i>et al.</i> <sup>(45)</sup> | 274.7                   | 94.7   | –410.3, 599.7       | 304.0          | 150.2  | –368.1, 668.5       |
| Itsiopoulos <i>et al.</i> <sup>(40)</sup>                                       | 271.0                   | 4.8  | –492.4, 502.0       | 253.6          | 129.6  | –289.5, 548.8       |
| Böhn <i>et al.</i> <sup>(29)</sup> ; Karlsen <i>et al.</i> <sup>(41)</sup>      | 673.8                   | 555.8  | –159.4, 1271.0      | 228.7          | 33.0   | –408.2, 474.2       |
| Miller III <i>et al.</i> <sup>(44)</sup>  | 242.0                   | 46.7   | –326.1, 419.6       | 236.4          | 50.4   | –370.8, 471.7       |
| Van het Hof <i>et al.</i> <sup>(49)</sup>                                       | 125.5                   | 27.0   | –170.3, 224.2       | 88.9           | 16.0   | –146.1, 178.0       |
| Van Loo-Bouwman <i>et al.</i> <sup>(50)</sup>                                   | 181.4                   | 0.48   | –305.9, 306.9       | 195.1          | –156.1   | –331.4, 19.2        |
| Briviba <i>et al.</i> <sup>(32)</sup> ; Watzl <i>et al.</i> <sup>(51)</sup>     | 270.1                   | –141.1   | –576.3, 294.1       | 210.6          | –64.8  | –441.2, 311.7       |
| Winkels <i>et al.</i> <sup>(52)</sup>   | 241.1                   | 145.9  | –121.3, 413.0       | 133.5          | 7.5  | –101.4, 116.5       |

FVJ, fruit, vegetable and juice intake; MFP, multivariable fractional polynomials; FV, fruit and vegetable intake, excluding juices; RMSE, root mean squared error.

patients that takes into account the random intercept after the model has been developed using individual participant data meta-analysis with mixed-effects modelling. In the present study, the performance of the conditional predictions was not considerably better than the performance of the standard predictions in an apparent validation (i.e. an internal validation based on the entire data, not using cross-validation) (data not shown).

In conclusion, the relatively strong correlations between predictions and actual intake indicate that the present prediction models may be used to investigate the ranking of individuals with regard to their intake of fruits and vegetables when validating questionnaires that measure intake (e.g. FFQ or 24 h recall). Furthermore, the low mean bias show that the models have good potential to be used to estimate average fruit and vegetable intake on a group level. The large limits of agreement indicate that the prediction models should not be used to estimate individual fruit and vegetable intake.

### Supplementary material

To view supplementary material for the present article, please visit <http://dx.doi.org/10.1017/S0007114515000355>

### Acknowledgements

The present research was financially supported by ZonMW (project number 200400014). ZonMW had no role in the design, analysis or writing of the present article.

The authors declare that there is no conflict of interest.

The authors' responsibilities were as follows: H. C. B. designed the research; R. F., B. W., A. B., E. R. M., J. J. M. C., W. J. P., K. v. d. H., M. C., A. K., L. O. D., R. W., C. I., L. B., K. O., C. A. v. L.-B. and T. H. J. N. provided essential data that was used for the present study; J. H. M. d. V. and H. v. d. V. provided essential advice; O. W. S. performed the statistical analysis; O. W. S. and H. C. B. wrote the paper; O. W. S. and H. C. B. had primary responsibility for final content. All authors read and approved the final manuscript.

### References

- Boeing H, Bechthold A, Bub A, *et al.* (2012) Critical review: vegetables and fruit in the prevention of chronic diseases. *Eur J Nutr* **51**, 637–663.
- Hung HC, Joshupura KJ, Jiang R, *et al.* (2004) Fruit and vegetable intake and risk of major chronic disease. *J Natl Cancer Inst* **96**, 1577–1584.
- Riboli E & Norat T (2003) Epidemiologic evidence of the protective effect of fruit and vegetables on cancer risk. *Am J Clin Nutr* **78**, Suppl. 3, 559S–569S.
- Macdonald HM, Hardcastle AC, Duthie GG, *et al.* (2009) Changes in vitamin biomarkers during a 2-year intervention trial involving increased fruit and vegetable consumption by free-living volunteers. *Br J Nutr* **102**, 1477–1486.
- Newman VA, Flatt SW & Pierce JP (2008) Telephone counseling promotes dietary change in healthy adults: results of a pilot trial. *J Am Diet Assoc* **108**, 1350–1354.
- Rock CL, Moskowitz A, Huizar B, *et al.* (2001) High vegetable and fruit diet intervention in premenopausal women with cervical intraepithelial neoplasia. *J Am Diet Assoc* **101**, 1167–1174.
- Day N, McKeown N, Wong M, *et al.* (2001) Epidemiological assessment of diet: a comparison of a 7-day diary with a food frequency questionnaire using urinary markers of nitrogen, potassium and sodium. *Int J Epidemiol* **30**, 309–317.
- Kipnis V, Midthune D, Freedman L, *et al.* (2002) Bias in dietary-report instruments and its implications for nutritional epidemiology. *Public Health Nutr* **5**, 915–923.
- Kipnis V, Midthune D, Freedman LS, *et al.* (2001) Empirical evidence of correlated biases in dietary assessment instruments and its implications. *Am J Epidemiol* **153**, 394–403.
- Kipnis V, Subar AF, Midthune D, *et al.* (2003) Structure of dietary measurement error: results of the OPEN biomarker study. *Am J Epidemiol* **158**, 14–21, discussion 22–16.
- Tasevska N, Midthune D, Potischman N, *et al.* (2011) Use of the predictive sugars biomarker to evaluate self-reported total sugars intake in the Observing Protein and Energy Nutrition (OPEN) study. *Cancer Epidemiol Biomarkers Prev* **20**, 490–500.
- Tasevska N, Runswick SA, McTaggart A, *et al.* (2005) Urinary sucrose and fructose as biomarkers for sugar consumption. *Cancer Epidemiol Biomarkers Prev* **14**, 1287–1294.
- Andersen LF, Veierod MB, Johansson L, *et al.* (2005) Evaluation of three dietary assessment methods and serum biomarkers as measures of fruit and vegetable intake, using the method of triads. *Br J Nutr* **93**, 519–527.
- Bogers RP, Dagnelie PC, Westertep KR, *et al.* (2003) 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. *J Nutr* **133**, 1213–1219.
- Bogers RP, Van Assema P, Kester AD, *et al.* (2004) Reproducibility, validity, and responsiveness to change of a short questionnaire for measuring fruit and vegetable intake. *Am J Epidemiol* **159**, 900–909.
- Brantsaeter AL, Haugen M, Rasmussen SE, *et al.* (2007) Urine flavonoids and plasma carotenoids in the validation of fruit, vegetable and tea intake during pregnancy in the Norwegian Mother and Child Cohort Study (MoBa). *Public Health Nutr* **10**, 838–847.
- Carlsen MH, Karlsen A, Lillegaard IT, *et al.* (2011) Relative validity of fruit and vegetable intake estimated from an FFQ, using carotenoid and flavonoid biomarkers and the method of triads. *Br J Nutr* **105**, 1530–1538.
- Jansen MC, Van Kappel AL, Ocke MC, *et al.* (2004) Plasma carotenoid levels in Dutch men and women, and the relation with vegetable and fruit consumption. *Eur J Clin Nutr* **58**, 1386–1395.
- Jilcott SB, Keyserling TC, Samuel-Hodge CD, *et al.* (2007) Validation of a brief dietary assessment to guide counseling for cardiovascular disease risk reduction in an underserved population. *J Am Diet Assoc* **107**, 246–255.
- Resnicow K, Odom E, Wang T, *et al.* (2000) Validation of three food frequency questionnaires and 24-hour recalls with serum carotenoid levels in a sample of African-American adults. *Am J Epidemiol* **152**, 1072–1080.
- Toft U, Kristoffersen L, Ladelund S, *et al.* (2008) Relative validity of a food frequency questionnaire used in the Inter99 study. *Eur J Clin Nutr* **62**, 1038–1046.
- Baldrick FR, Woodside JV, Ellborn JS, *et al.* (2011) Biomarkers of fruit and vegetable intake in human intervention studies: a systematic review. *Crit Rev Food Sci Nutr* **51**, 795–815.
- Brevik A, Vollset SE, Tell GS, *et al.* (2005) Plasma concentration of folate as a biomarker for the intake of fruit and vegetables: the Hordaland Homocysteine Study. *Am J Clin Nutr* **81**, 434–439.

24. Willett WC (2013) *Nutritional Epidemiology*, 3rd ed. Oxford: Oxford University Press.
25. Royston P & Altman DG (1994) Regression using fractional polynomials of continuous covariates – parsimonious parametric modeling. *Appl Stat* **43**, 429–467.
26. Sauerbrei W & Royston P (1999) Building multivariable prognostic and diagnostic models: transformation of the predictors by using fractional polynomials. *J R Stat Soc Ser A Stat Soc* **162**, 71–94.
27. Vergouwe Y, Royston P, Moons KGM, *et al.* (2010) Development and validation of a prediction model with missing predictor data: a practical approach. *J Clin Epidemiol* **63**, 205–214.
28. Appel LJ, Miller ER III, Jee SH, *et al.* (2000) Effect of dietary patterns on serum homocysteine: results of a randomized, controlled feeding study. *Circulation* **102**, 852–857.
29. Bøhn SK, Myhrstad MC, Thoresen M, *et al.* (2010) Blood cell gene expression associated with cellular stress defense is modulated by antioxidant-rich food in a randomised controlled clinical trial of male smokers. *BMC Med* **8**, 54.
30. Bowen PE, Garg V, Stacewicz-Sapuntzakis M, *et al.* (1993) Variability of serum carotenoids in response to controlled diets containing six servings of fruits and vegetables per day. *Ann N Y Acad Sci* **691**, 241–243.
31. Brevik A, Andersen LF, Karlsen A, *et al.* (2004) Six carotenoids in plasma used to assess recommended intake of fruits and vegetables in a controlled feeding study. *Eur J Clin Nutr* **58**, 1166–1173.
32. Briviba K, Bub A, Möseneder J, *et al.* (2008) No differences in DNA damage and antioxidant capacity between intervention groups of healthy, nonsmoking men receiving 2, 5, or 8 servings/d of vegetables and fruit. *Nutr Cancer* **60**, 164–170.
33. Broekmans WMR, Klöpping-Ketelaars IAA, Schuurman CRWC, *et al.* (2000) Fruits and vegetables increase plasma carotenoids and vitamins and decrease homocysteine in humans. *J Nutr* **130**, 1578–1583.
34. Brouwer IA, Van Dusseldorp M, West CE, *et al.* (1999) Dietary folate from vegetables and citrus fruit decreases plasma homocysteine concentrations in humans in a dietary controlled trial. *J Nutr* **129**, 1135–1139.
35. Castenmiller JJ, van de Poll CJ, West CE, *et al.* (2000) Bioavailability of folate from processed spinach in humans. Effect of food matrix and interaction with carotenoids. *Ann Nutr Metab* **44**, 163–169.
36. Castenmiller JJ, West CE, Linssen JP, *et al.* (1999) The food matrix of spinach is a limiting factor in determining the bioavailability of  $\beta$ -carotene and to a lesser extent of lutein in humans. *J Nutr* **129**, 349–355.
37. Chopra M, O'Neill ME, Keogh N, *et al.* (2000) Influence of increased fruit and vegetable intake on plasma and lipoprotein carotenoids and LDL oxidation in smokers and nonsmokers. *Clin Chem* **46**, 1818–1829.
38. Dragsted LO, Pedersen A, Hermetter A, *et al.* (2004) The 6-a-day study: effects of fruit and vegetables on markers of oxidative stress and antioxidative defense in healthy nonsmokers. *Am J Clin Nutr* **79**, 1060–1072.
39. Freese R, Alftan G, Jauhiainen M, *et al.* (2002) High intakes of vegetables, berries, and apples combined with a high intake of linoleic or oleic acid only slightly affect markers of lipid peroxidation and lipoprotein metabolism in healthy subjects. *Am J Clin Nutr* **76**, 950–960.
40. Itsiopoulos C, Brazionis L, Kaimakamis M, *et al.* (2011) Can the Mediterranean diet lower HbA1c in type 2 diabetes? Results from a randomized cross-over study. *Nutr Metab Cardiovasc Dis* **21**, 740–747.
41. Karlsen A, Svendsen M, Seljeflot I, *et al.* (2011) Compliance, tolerability and safety of two antioxidant-rich diets: a randomised controlled trial in male smokers. *Br J Nutr* **106**, 557–571.
42. Martini MC, Campbell DR, Gross MD, *et al.* (1995) Plasma carotenoids as biomarkers of vegetable intake: The University of Minnesota cancer prevention research unit feeding studies. *Cancer Epidemiol Biomarkers Prev* **4**, 491–496.
43. Miller ER III, Appel LJ & Risby TH (1998) Effect of dietary patterns on measures of lipid peroxidation: results from a randomized clinical trial. *Circulation* **98**, 2390–2395.
44. Miller ER III, Erlinger TP, Sacks FM, *et al.* (2005) A dietary pattern that lowers oxidative stress increases antibodies to oxidized LDL: results from a randomized controlled feeding study. *Atherosclerosis* **183**, 175–182.
45. Misikangas M, Freese R, Turpeinen AM, *et al.* (2001) High linoleic acid, low vegetable, and high oleic acid, high vegetable diets affect platelet activation similarly in healthy women and men. *J Nutr* **131**, 1700–1705.
46. Moller P, Vogel U, Pedersen A, *et al.* (2003) No effect of 600 grams fruit and vegetables per day on oxidative DNA damage and repair in healthy nonsmokers. *Cancer Epidemiol Biomarkers Prev* **12**, 1016–1022.
47. Silaste ML, Rantala M, Alftan G, *et al.* (2003) Plasma homocysteine concentration is decreased by dietary intervention. *Br J Nutr* **89**, 295–301.
48. Silaste ML, Rantala M, Alftan G, *et al.* (2004) Changes in dietary fat intake alter plasma levels of oxidized, low-density lipoprotein and lipoprotein(a). *Arterioscler Thromb Vasc Biol* **24**, 498–503.
49. van het Hof KH, Brouwer IA, West CE, *et al.* (1999) Bioavailability of lutein from vegetables is 5 times higher than that of  $\beta$ -carotene. *Am J Clin Nutr* **70**, 261–268.
50. Van Loo-Bouwman CA, West CE, Van Breemen RB, *et al.* (2009) Vitamin A equivalency of  $\beta$ -carotene in healthy adults: limitation of the extrinsic dual-isotope dilution technique to measure matrix effect. *Br J Nutr* **101**, 1837–1845.
51. Watzl B, Kulling SE, Möseneder J, *et al.* (2005) A 4-wk intervention with high intake of carotenoid-rich vegetables and fruit reduces plasma C-reactive protein in healthy, nonsmoking men. *Am J Clin Nutr* **82**, 1052–1058.
52. Winkels RM, Brouwer IA, Siebelink E, *et al.* (2007) Bioavailability of food folates is 80% of that of folic acid. *Am J Clin Nutr* **85**, 465–473.
53. Yeon JY, Kim HS & Sung MK (2012) Diets rich in fruits and vegetables suppress blood biomarkers of metabolic stress in overweight women. *Prev Med* **54**, S109–S115.
54. Yeum KJ, Booth SL, Sadowski JA, *et al.* (1996) Human plasma carotenoid response to the ingestion of controlled diets high in fruits and vegetables. *Am J Clin Nutr* **64**, 594–602.
55. Crispim SP, Geelen A, Souverein OW, *et al.* (2011) Biomarker-based evaluation of two 24-h recalls for comparing usual fish, fruit and vegetable intakes across European centers in the EFCOVAL Study. *Eur J Clin Nutr* **65**, Suppl. 1, S38–S47.
56. Kristal AR, Vizenor NC, Patterson RE, *et al.* (2000) Precision and bias of food frequency-based measures of fruit and vegetable intakes. *Cancer Epidemiol Biomarkers Prev* **9**, 939–944.
57. Chopra M, McLoone U, O'Neill M, *et al.* (1996) Fruit and vegetable supplementation – effect on *ex vivo* LDL oxidation in humans. In *Natural Antioxidants and Food Quality in Atherosclerosis and Cancer Prevention*, pp. 150–155 [JT Kumpulainen and JT Salonen, editors]. London: The Royal Society of Chemistry.



58. Blanck HM, Bowman BA, Cooper GR, *et al.* (2003) Laboratory issues: use of nutritional biomarkers. *J Nutr* **133**, Suppl. 3, 888S–894S.
59. Brady WE, Mares-Perlman JA, Bowen P, *et al.* (1996) Human serum carotenoid concentrations are related to physiologic and lifestyle factors. *J Nutr* **126**, 129–137.
60. Drewnowski A, Rock CL, Henderson SA, *et al.* (1997) Serum  $\beta$ -carotene and vitamin C as biomarkers of vegetable and fruit intakes in a community-based sample of French adults. *Am J Clin Nutr* **65**, 1796–1802.
61. Maiani G, Caston MJ, Catasta G, *et al.* (2009) Carotenoids: actual knowledge on food sources, intakes, stability and bioavailability and their protective role in humans. *Mol Nutr Food Res* **53**, Suppl. 2, S194–S218.
62. Tucker KL, Selhub J, Wilson PW, *et al.* (1996) Dietary intake pattern relates to plasma folate and homocysteine concentrations in the Framingham Heart Study. *J Nutr* **126**, 3025–3031.
63. van Kappel AL, Steghens JP, Zeleniuch-Jacquotte A, *et al.* (2001) Serum carotenoids as biomarkers of fruit and vegetable consumption in the New York Women's Health Study. *Public Health Nutr* **4**, 829–835.
64. Efron B (1983) Estimating the error rate of a prediction rule: improvement on cross-validation. *J Am Stat Assoc* **78**, 316–331.
65. Abo-Zaid G, Guo B, Deeks JJ, *et al.* (2013) Individual participant data meta-analyses should not ignore clustering. *J Clin Epidemiol* **66**, 865–873, e864.
66. Bouwmeester W, Twisk JW, Kappen TH, *et al.* (2013) Prediction models for clustered data: comparison of a random intercept and standard regression model. *BMC Med Res Methodol* **13**, 19.
67. Debray TP, Moons KG, Ahmed I, *et al.* (2013) A framework for developing, implementing, and evaluating clinical prediction models in an individual participant data meta-analysis. *Stat Med* **32**, 3158–3180.