

Consistency of food intake factors by different dietary assessment methods and population groups

Per Togo^{1,3*}, Berit L. Heitmann^{1,2}, Thorkild I. A. Sørensen² and Merete Osler⁴

¹Research Unit for Dietary Studies and

²Danish Epidemiology Science Centre at the Institute of Preventive Medicine, Copenhagen University Hospital, Copenhagen, Denmark

³Copenhagen County Research Centre for Prevention and Health, Glostrup University Hospital, Glostrup, Denmark

⁴Institute of Public Health, University of Copenhagen, Copenhagen, Denmark

(Received 25 May 2002 – Revised 27 April 2003 – Accepted 23 May 2003)

Several studies have used factor analysis to identify food intake patterns in epidemiological studies as an alternative to nutrient-based analyses, but few have validated the factors in a larger population. Our present objectives were: to compare the factor scores based on a food-frequency questionnaire (FFQ) with scores based on a 7 d diet record; to examine the consistency of the factor score correlations across strata of age, BMI, energy intake, education, physical activity and smoking and to compare factors identified in two sub-populations. In 879 men and 927 women, of the total population sample of 3785, scores on food intake factors, three for men ('green', 'sweet' and 'traditional') and two for women ('green' and 'sweet-traditional'), identified in data from the FFQ and the diet record, were compared. The loadings of foods on the factors were very similar and the correlations between the corresponding factor scores, based on the two dietary assessment methods, were: for men 'green' 0.61, 'sweet' 0.55, 'traditional' 0.34; for women, 'green' 0.61, 'sweet-traditional' 0.57. Stratification did not significantly modify the correlations, with a few inconsistent exceptions. Factors obtained in a different subsample of the population, for which there was only data from the FFQ, were almost identical to the factors found in the subsample, who provided both FFQ and diet record information with regard to food loadings and model fit. In conclusion, the food intake factors identified were reproducible using two different dietary assessment methods and, furthermore, independent of stratification.

Diet: Factor analysis: Observational studies: Validation

In nutritional epidemiology, the identification of food intake patterns has become an important alternative to the nutrient-oriented approach (Jacobson & Stanton, 1986; Williams *et al.* 1999; Hu, 2002; Quatromoni *et al.* 2002). One strong argument for the use of food intake patterns rather than nutrient intakes is that nutrient interaction and action of known or unknown substances in the daily diet are probably better accounted for by the use of food intake patterns. Food intake patterns identify the consistent components in the actual intake of a combination of foods and the most frequent 'exposures' from the diet. One commonly used method to identify food intake patterns is exploratory factor analysis, which is based on analyses of the correlation-covariance matrix of a number of food variables. The factors are weighted combinations of foods, which best explain the variance in the food intake (the correlation matrix). Factor loadings are the correlations between the foods and the factor, and the individual factor scores are estimates of factor relationship to the individual's food intake, and hence, the factor scores

reflect the values of each of the foods that identifies the factors. Two variants of factor analyses were used for the present study: exploratory factor analysis, in which all foods load on all factors (the common *a posteriori* approach); confirmatory factor analysis, in which only the foods decided on *a priori*, e.g. by the magnitude of their loading in an exploratory factor analysis, are included (for example, see Loehlin, 1998).

Comparisons of scores with macronutrient and energy (E) intake, as well as a small number of background variables have typically been applied to validate the resulting factors and individual factor scores. Dietary assessment methods, such as food-frequency questionnaires (FFQ) or diet records (DR), have been used interchangeably in nutritional epidemiology (Kant, 1996; Togo *et al.* 2001) and as a basis for factor analysis. One study investigated the reproducibility of two factors using different dietary assessment methods on a small sample (*n* 127) from the Health Professionals Follow-up study (Hu *et al.* 1999). Exploratory factor analyses were applied to data from

Abbreviations: DR, diet record; E, energy; FFQ, food-frequency questionnaire.

* **Corresponding author:** Dr Per Togo, present address, Kong Oscars Gade 9, 2.th., DK-2100 Copenhagen Ø, Denmark, fax +45 3332 4240, email per.togo@dadlnet.dk

two 126-item FFQ (condensed into forty items for the analyses) and two 7 d DR: correlations between scores based on the two methods ranged between 0.45 and 0.74. Another recent paper of the European Prospective Investigation into Cancer and Nutrition (Potsdam cohort) described tests of the usefulness of simplified measures of dietary patterns derived from an exploratory factor analysis in the form of a sum of the standardised value of the six most loading foods on each factor (Schulze *et al.* 2003). High correlation coefficients were found between the original factor scores and the simplified sum-scores (>0.95) and the simplified measures were proposed as a way to test factors in different studies and populations.

The present study repeats, in a different sample, modifications of some of the analyses described in the two studies by Hu *et al.* (1999) and Schulze *et al.* (2003). The present study adds the use of confirmatory factor analysis as a method to test and reproduce factors in a different population, in order to compute factor scores based on *a priori*-defined food intake patterns. Thus, the purpose of the present study was to test two hypotheses: (1) that similar food intake patterns will be identified using data from a twenty-six-item FFQ and data from a 7 d DR; (2) that a model of food intake pattern, identified by exploratory factor analysis in one population group, can be readily identified, by exploratory as well as confirmatory factor analysis, in another population group.

Material and methods

Subjects

The data used for the present analysis originates from the Danish contribution to the WHO-initiated survey Monitoring of Trends and Determinants in Cardiovascular Diseases (MONICA; MONICA Data Centre, 2000). The population invited for the baseline examination (MONICA 1, 1982–4) at the Copenhagen County Research Centre for Prevention and Health (Glostrup University Hospital, Glostrup, Denmark) included 4807 Danish citizens, aged 30, 40, 50 or 60 years. The study population was a random sample of the Danish population, selected from the Central Person Register among all citizens who lived in the western part of the Copenhagen County. Of those invited, 79% (1845 women and 1940 men) attended MONICA 1. In total, 38% of the invited or 48% of the attending participants (879 men and 927 women) were eligible for the dietary assessment comparison part of the present study, as they had completed both types of dietary assessments, the FFQ and the DR (referred to as group A). Most of the remaining subjects (833 women and 971 men, 48% of the subjects attending in MONICA 1) filled in the FFQ but not the DR. Data from those subjects were used to test the reproducibility of the FFQ-based factors (referred to as group B).

Dietary information

The dietary information was obtained using a twenty-six-item FFQ and a 7 d weighed DR. In the FFQ, the subjects were asked to state how often, on average,

they had consumed each food within the last year. Answers could be given as: never, once per month or less, twice per month, once per week, 2–3 times per week, once per d, 2–3 times per d, 4 times per d or more. The foods and food groups included in the FFQ are listed in Table 1. Five foods were omitted because of a large proportion ($>50\%$) of 'never' or 'once per month or less' answers (Table 1), which made the foods less suitable for the comparison with the 7 d of recording in the DR, and gave them a skewed distribution that could cause difficulty in factor analyses.

All participants were asked to complete a DR. Subjects who accepted (903 men and 949 women) were asked to complete the DR in a normal week within the 3 weeks following the baseline investigation. Differences between subjects who did or did not fill in the DR have been described earlier (Jørgensen, 1992). In short, subjects who filled in the DR were more likely to be women, older and non-smokers. There were only small differences in diet, measured by the FFQ, between the subjects who did and those who did not fill in the DR (results not shown). We have no information on the reasons why approximately half of the population chose not to record their diet. The DR were completed using a pre-coded form containing 111 foods and food groups (with space for additional individual foods), together with a guide on the size of standard measures if weighing the food was not possible. The food intake reported using the DR was converted into nutrients and E using the program Micro-Camp® (Mørup, 1986), which is based on the Danish nutrient composition tables (Levnedsmiddelstyrelsen, 1985). Protein, carbohydrate and fat (% E) and total E intake (kJ/d) were calculated as described earlier and E intake divided in quartiles for stratification. E density was calculated as a mean value (total E intake (kJ)/total weight (g) of all foods and beverages recorded in the DR).

To enable an equivalent factor analysis on the FFQ and DR data, the output from the DR (g food per week) was merged by summation into twenty-one food groups corresponding to the twenty-one (of the twenty-six) groups used in the analysis based on the FFQ. Those twenty-one summary variables were categorised into eight intake levels (octiles). The number of categories was chosen to match the number of intake levels on the FFQ. Table 1 shows the foods that were summed into each food group and correlations between the FFQ- and DR-derived category variables. Table 1 includes a list of the five food groups and twenty-two individual foods that were less frequently eaten or difficult to group and therefore omitted.

Other covariates

All subjects went through a general health examination in addition to filling in an extensive questionnaire on lifestyle and health. BMI was calculated from weight and height (kg/m^2) measured during the health examination using standardised methods. The WHO cut-off points (18.5, 25.0 and $30.0 \text{ kg}/\text{m}^2$; World Health Organization, 2000) were used to define the BMI groups underweight, normal weight and overweight and obesity; due to few subjects in the underweight group, the lower limit for normal

Table 1. Food groups on the food-frequency questionnaire (FFQ) and foods from 7 d diet records merged into summary variables*†

| Food group variables (FFQ) | Foods from 7 d diet record‡§ | Correlations | |
|---|--|--------------|-------|
| | | Men | Women |
| Milk or yoghurt¶ | Skimmed milk, semi-skimmed milk, butter milk, butter milk dish, quark 5% (cheese) full-fat milk, cream (13%), double cream, drinking chocolate, fruit yoghurt, junket (of full-fat milk); sour cream (18%) | 0.68 | 0.66 |
| Rye bread with whole grains and/or bran | Rye bread with whole grains | 0.68 | 0.63 |
| Jam, marmalade or honey | Honey, jam or marmalade | 0.59 | 0.60 |
| Fruit | Apples pears, avocado, banana, fruit (boiled), fruit (preserved), fruit (raw), fruit pudding, orange, raisins, strawberry | 0.57 | 0.51 |
| Cheese¶ | 30% cheese, 40% cheese, 45% cheese | 0.56 | 0.50 |
| Vegetable fat margarine or spread | Maize oil, soyabean oil, sunflower seed oil, 'plant' margarine | 0.53 | 0.58 |
| Candy or chocolate | Chocolate, sweets, confectionery, candy or liquorice | 0.47 | 0.47 |
| Cake, biscuits and baked goods | Biscuits, cookies, 'Danish pastry', pancake | 0.46 | 0.48 |
| White (wheat) bread | Bread rolls, rusk (toasted roll), white bread | 0.45 | 0.47 |
| Wheat bread with whole grains and/or bran | Crisp bread, wheat bread with whole grain | 0.42 | 0.48 |
| Soft drink or ice cream | Ice cream, soft drink (with sugar) | 0.41 | 0.39 |
| Eggs | Egg white, egg whole and scrambled, egg yolk | 0.43 | 0.43 |
| Butter, lard and hard margarine | Butter, olive oil, hard margarine, lard from pork | 0.42 | 0.44 |
| Pâté and meat for bread | Brisket of beef, pâté, salami-type sausage, salted meat for bread | 0.39 | 0.41 |
| Raw vegetables | Carrots, olives (green), preserved vegetables (assorted), raw vegetables (assorted), tomatoes | 0.38 | 0.39 |
| Potatoes | New potatoes, old potatoes, mashed potatoes | 0.38 | 0.36 |
| Fish | Cod or trout, cod roe, caviar, eel, fillet of fish, garfish, halibut, herring, mackerel, mince from fish, mussel, plaice, salmon, shrimps, squid, tuna | 0.28 | 0.34 |
| Boiled vegetables | Boiled vegetables (assorted), sweetcorn, green peas, soyabean, green beans | 0.27 | 0.27 |
| Rice | Rice | 0.27 | 0.35 |
| Sausages | Sausage (boiled) | 0.19 | 0.25 |
| Meat** | Beef (fatty), beef (lean), lamb, pork (breaded fried), pork (fatty), pork (lean), pork (medium fat), bacon, rissole, veal, liver, kidney, heart, chicken, duck or hare, goose, hen, turkey | 0.10 | 0.14 |

* For details of subjects and procedures, see p. 668.

† Food groups and foods were sorted by Spearman rank correlation coefficients (for men) between FFQ variables and categorised summary variables (based on diet records, categorised in octiles).

‡ Omitted foods and drinks on the diet record: roasted onions, added sugar, mustard, gravy, sauce thickened with roux, sauce Béarnaise, mayonnaise, dressing, nuts, sunflower seeds, potato crisps, French fries, pizza, stew, consommé, tomato soup, mayonnaise salads (two different), vegetables in white sauce, sponge pudding, lemon mousse, rice porridge, coffee, tea, water, sparkling water, beer (ordinary, light or strong), wine, strong liquor, dessert wine.

§ Omitted food groups due to low-frequent intake (> 50% of subjects stated an intake less than once per month): oatmeal porridge or hulled grain (barley oats, muesli, oatmeal), low-fat spread or margarine (diet margarine, low-fat margarine), juice, rye bread, ordinary (rye bread, ordinary dark, light rye bread, rye bread porridge 'oellebroed'), spaghetti and pasta.

|| Spearman rank correlations, all had a *P* value < 0.01.

¶ '%' Refers to fat (g) per 100 g milk product (not % energy) if given in the 7 d diet record.

** Meat was not subdivided in the FFQ except for 'pâté and meat for bread', 'sausages' and 'fish'. Poultry and offal were added to the new 'meat' variable for the diet record data as listed.

weight was increased to 20. Educational level in three categories (of originally eight) in addition to the obligatory 7–10 years of school was used: (1) no further education; (2) semi-skilled worker, acquired skills or 1 year (theoretical) education or basic vocational courses, apprentice or short (2–3 years) theoretical education; (3) medium-length theoretical education (3–4 years) or longer education (5+ years) or student and/or in-training. Physical activity level during leisure time was classified as: (1) mostly sitting; (2) walking, gardening etc.; (3) low-level sports; (4) competitive sports. (Due to too few subjects in the latter category, the categories (3) and (4) were merged.)

Smoking habits were determined by the question: Do you smoke? (1) yes, daily; (2) occasionally; (3) no; and, if not smoking: Did you smoke previously? (1) yes, daily; (2) occasionally; (3) never. The questions were merged into one variable with the categories: (1) never; (2) ex-smoker (including ex- and occasionally); (3) smoker (including occasionally).

Factor analyses

The number of factors was determined by scree-plots based on principal component analyses (Fig. 1) and the

combination of foods on the factors. A factor solution with three factors for men and two for women was found to be the most reasonable and, consequently, all subsequent analyses were done separately for men and women.

Two exploratory analyses with non-orthogonal rotation (see Appendix) were conducted, using FFQ and DR data in the participants who completed both dietary assessments (group A). The loadings of food items on each factor are given in Table 2. Factor-factor correlations within each factor model were allowed in order to accommodate the likely possibility that, for example, a prudent dietary factor was negatively associated with an imprudent factor; correlations are given in the text.

Two confirmatory factor models (see Appendix) were tested in the same participant (group A), based on the foods loadings >0.3 on the factors in exploratory factor analyses on FFQ and DR data. The included foods and loadings based on the two dietary assessment methods are shown in Table 3; model fits and factor correlations are given in the text.

For comparison, individual scores on the factors based on the two kinds of data were computed in both the exploratory analyses and by using the confirmatory factor models. In Table 4, the Pearson correlations between scores based on different dietary assessment methods, as well as different factor analysis methods, are shown (score computation is described briefly in the Appendix).

The factors identified in group A were compared with the factors identified in the data for group B. This was done by conducting three analyses: (1) An exploratory factor analysis on group B using all twenty-one food groups; (2) A confirmatory factor analysis on group B, using the model developed in group A. In this analysis,

the number of factors and the specific foods loading on them were the same for group B as for group A. The loading values were allowed to be different as illustrated in the two scatter-plots in Figs 2 and 3; (3) A confirmatory analysis on both groups at once, in which all loadings were fixed to the same value for group B as found in the analysis on group A. This allowed the determination of a difference in mean score between the two groups (mean score is standardised to 0, by default, in other analyses).

Additional statistical methods

Spearman rank correlation was used for the comparison of the FFQ food groups and the DR food groups, whereas Pearson correlations were computed between the factor scores and the total E intake and overall E density. Partial correlations were calculated between factor scores and the relative E contribution by macronutrients, adjusted for age and total E intake as assessed by the DR. Pearson correlations between the factor scores based on the two dietary assessment methods were computed overall, and for each category of the covariates: age; BMI group; total E intake in quartiles; physical activity; education; smoking. The difference of correlations across strata was tested between all strata of each variable using Fisher transformation of correlations and the method described in Kleinbaum *et al.* (1998) and in the Appendix.

Data were analysed using SPSS statistical software package (version 10.0; SPSS Inc., Chicago, IL, USA), Mplus statistical software package (version 2.01; Mplus, Los Angeles, CA, USA; Muthen & Muthen, 2001) and Microsoft® Excel 2000 (Microsoft, Redmond, WA, USA).

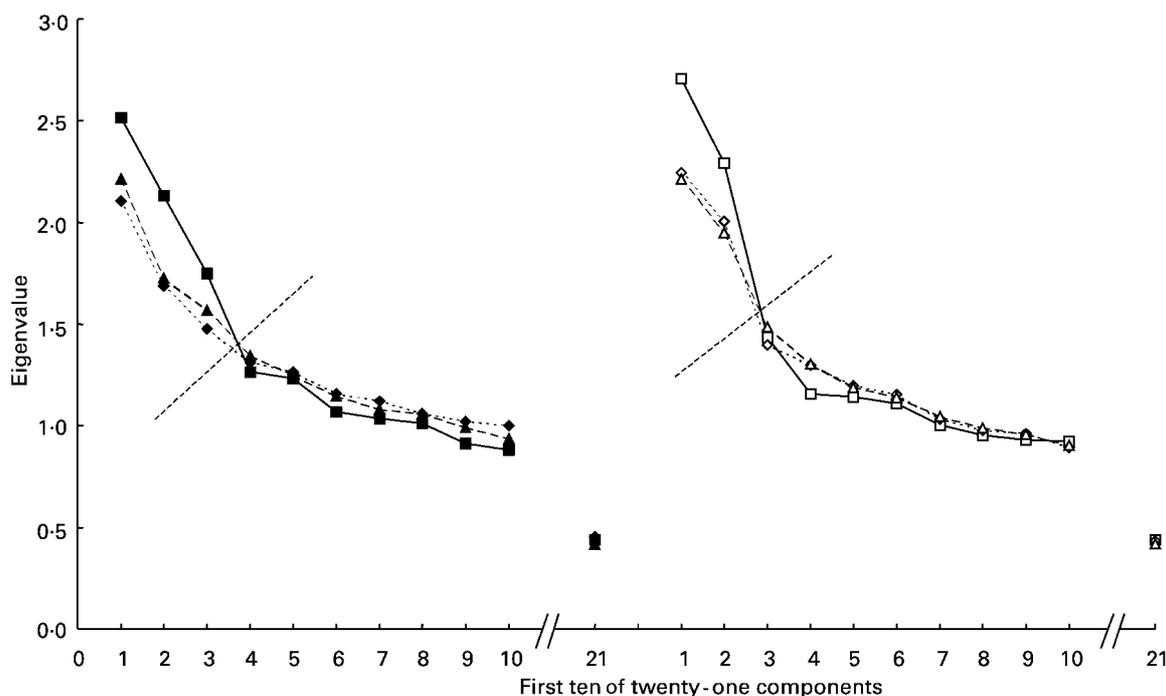


Fig. 1. Scree-plots of Eigenvalues by number of components for exploratory factor analysis on the three kinds of dietary input food-frequency questionnaire (FFQ), 7 d diet records (DR; g per week) and 7 d DR (g per week, categorised into octiles). Men: ■, FFQ; ◆, 7 d DR; ▲, 7 d DR categorised into octiles. Women: □, FFQ; ◇, 7 d DR; △, 7 d DR categorised into octiles. For details of subjects and procedures, see pp. 668–670.

Table 2. Factor loadings estimated by two exploratory factor analyses using food-frequency questionnaire (FFQ) data and diet record (DR) data in octiles*†

| Factor labels... | Men (n 879) | | | | | | Women (n 927) | | | | | |
|---|-------------|---------|---------|--------|---------------|---------|---------------|--------|---------------------|--------|--------|------|
| | 'Green' | | 'Sweet' | | 'Traditional' | | 'Green' | | 'Sweet-Traditional' | | DR-2 | |
| | FFQ-1§ | DR-1 | FFQ-2§ | DR-2 | FFQ-3§ | DR-3 | FFQ-1§ | DR-1 | FFQ-2§ | DR-2 | FFQ-2§ | DR-2 |
| Wheat bread with whole grains and/or bran | 0.60** | 0.45** | 0.07 | 0.20 | -0.22 | -0.38** | 0.62** | 0.47** | -0.07 | -0.29 | | |
| Raw vegetables | 0.60** | 0.65** | -0.09 | 0.05 | 0.01 | 0.13 | 0.58** | 0.56** | -0.13 | -0.07 | | |
| Boiled vegetables | 0.57** | 0.52** | -0.10 | 0.03 | 0.20 | 0.23 | 0.41** | 0.37** | -0.10 | -0.14 | | |
| Rye bread with whole grains and/or bran | 0.56** | 0.54** | 0.05 | 0.05 | -0.17 | 0.06 | 0.57** | 0.50** | -0.12 | -0.03 | | |
| Fruit | 0.50** | 0.54** | 0.20 | 0.17 | 0.02 | 0.03 | 0.56** | 0.56** | 0.10 | 0.09 | | |
| Fish | 0.41** | 0.29 | -0.06 | -0.10 | 0.24 | 0.21 | 0.40** | 0.33** | -0.06 | -0.03 | | |
| Rice | 0.36** | -0.01 | 0.29 | 0.25 | 0.00 | -0.28 | 0.38** | 0.19 | 0.25 | -0.05 | | |
| Cheese | 0.31** | 0.27 | 0.01 | 0.17 | 0.12 | 0.05 | 0.31** | 0.49** | 0.03 | 0.17 | | |
| Milk or yoghurt | 0.30** | 0.25 | 0.13 | 0.25 | 0.17 | 0.06 | 0.34** | 0.34** | 0.11 | -0.01 | | |
| Cake, biscuits and baked goods | 0.07 | 0.05 | 0.75** | 0.67** | 0.00 | 0.20 | 0.23 | 0.25 | 0.61** | 0.56** | | |
| Candy or chocolate | -0.05 | 0.04 | 0.73** | 0.61** | 0.01 | -0.03 | 0.17 | 0.11 | 0.60** | 0.51** | | |
| Soft drink or ice cream | -0.12 | -0.03 | 0.56** | 0.38** | 0.04 | 0.13 | -0.07 | -0.05 | 0.42** | 0.45** | | |
| Jam, marmalade or honey | 0.18 | 0.21 | 0.55** | 0.49** | 0.01 | -0.07 | 0.31** | 0.31** | 0.42** | 0.38** | | |
| Meat¶ | 0.07 | 0.29 | -0.08 | -0.12 | 0.63** | 0.58** | -0.10 | 0.18 | 0.33** | 0.08 | | |
| Pâté and meat for bread | 0.13 | -0.06 | 0.01 | 0.18 | 0.61** | 0.43** | -0.09 | -0.04 | 0.55** | 0.33** | | |
| Potatoes | -0.02 | 0.17 | -0.15 | -0.14 | 0.54** | 0.63** | -0.07 | 0.03 | 0.38** | 0.19 | | |
| Sausages | 0.07 | -0.29 | 0.13 | 0.09 | 0.49** | 0.00 | -0.03 | -0.22 | 0.46** | 0.18 | | |
| White (wheat) bread | -0.38** | -0.38** | 0.24 | 0.37** | 0.41** | 0.27 | -0.42** | -0.21 | 0.47** | 0.67** | | |
| Butter, lard and hard margarine | -0.02 | -0.09 | 0.16 | 0.37** | 0.37** | 0.47** | -0.05 | -0.10 | 0.45** | 0.49** | | |
| Eggs | 0.29 | 0.18 | -0.16 | -0.03 | 0.35** | 0.27 | 0.30** | 0.26 | 0.12 | 0.11 | | |
| Vegetable-fat margarine or spread | 0.10 | 0.14 | 0.14 | -0.04 | 0.20 | -0.16 | 0.12 | 0.28 | 0.27 | 0.19 | | |
| Explained variance (%) | 12.0 | 10.5 | 10.2 | 8.2 | 8.3 | 7.5 | 12.9 | 10.5 | 10.9 | 9.3 | | |

* For details of subjects and procedures, see Table 1 and p. 668.

† Loadings estimated by SPSS on correlation matrix with Kaiser normalisation, using principal components and Promax rotation. (For details of statistical procedures, see pp. 668-670 and Appendix.)

‡ Sorted by loadings on FFQ factors; for men; only foods eaten more than once per month by >50% of subjects according to the FFQ were included in the analyses (oatmeal, juice, light rye-bread, diet margarine and pasta were omitted).

§ FFQ-1, -2 and -3 refer to the first, second, and for men, third, factors based on FFQ data.

|| DR-1, -2, -3 refer to the first, second and third factors based on the DR data.

¶ Except for 'pâté and meat for bread' and sausages.

** Loadings ≥ 0.3.

Table 3. Factor loadings estimated by two confirmatory factor analyses using food-frequency questionnaire (FFQ) data and diet record (DR) data in octiles*†

| | Factor loading | |
|---|----------------|-------|
| | FFQ | DR |
| Men (n 879) | | |
| ‘Green’ (FFQ-1 and DR-1)§ | | |
| Wheat bread with whole grains and/or bran | 0.74 | 0.69 |
| Rye bread with whole grains and/or bran | 0.56 | 0.39 |
| Raw vegetables | 0.47 | 0.39 |
| Fruit | 0.43 | 0.39 |
| Boiled vegetables | 0.40 | 0.27 |
| Rice | 0.31 | |
| Fish | 0.25 | |
| Cheese | 0.24 | |
| Milk or yoghurt | 0.20 | |
| White (wheat) bread | -0.36 | -0.78 |
| ‘Sweet’ (FFQ-2 and DR-2)§ | | |
| Cake, biscuits and baked goods | 0.71 | 0.57 |
| Candy or chocolate | 0.63 | 0.45 |
| Jam, marmalade or honey | 0.46 | 0.51 |
| Soft drink or ice cream | 0.40 | 0.16 |
| White (wheat) bread | | 0.61 |
| Butter, lard and hard margarine | | 0.21 |
| ‘Traditional’ (FFQ-3 and DR-3)§ | | |
| Meat | 0.62 | 0.26 |
| Pâté and meat for bread | 0.57 | 0.44 |
| Potatoes | 0.47 | 0.38 |
| White (wheat) bread | 0.38 | |
| Butter, lard and hard margarine | 0.32 | 0.42 |
| Sausages | 0.31 | |
| Eggs | 0.12 | |
| Wheat bread with whole grains and/or bran | | -0.20 |
| Women (n 927) | | |
| ‘Green’ (FFQ-1 and DR-1)§ | | |
| Wheat bread with whole grains and/or bran | 0.60 | 0.51 |
| Raw vegetables | 0.55 | 0.49 |
| Rye bread with whole grains and/or bran | 0.52 | 0.42 |
| Fruit | 0.44 | 0.44 |
| Boiled vegetables | 0.42 | 0.32 |
| Fish | 0.37 | 0.26 |
| Cheese | 0.28 | 0.33 |
| Rice | 0.21 | |
| Milk or yoghurt | 0.21 | 0.25 |
| Jam, marmalade or honey | 0.21 | 0.35 |
| White (wheat) bread | -0.41 | |
| ‘Sweet–Traditional’ (FFQ-2 and DR-3)§ | | |
| Candy or chocolate | 0.58 | 0.37 |
| Cake, biscuits and baked goods | 0.57 | 0.37 |
| Pâté and meat for bread | 0.47 | 0.20 |
| Butter, lard and hard margarine | 0.43 | 0.43 |
| White (wheat) bread | 0.40 | 0.65 |
| Soft drink or ice cream | 0.39 | 0.37 |
| Potatoes | 0.37 | |
| Sausages | 0.37 | |
| Meat | 0.34 | |
| Jam, marmalade or honey | 0.33 | 0.42 |

* For details of subjects and procedures, see Table 1 and p. 668.

† Confirmatory factor analysis based on food groups loading ≥ 0.3 in exploratory factor analysis on the same data material (FFQ and DR); loadings estimated by Mplus by confirmatory factor analysis on polychoric correlation matrix, weighted least square variable estimates with robust standard errors and mean- and variance-adjusted χ^2 test statistic were used (for details of statistical analysis, see pp. 668–670).

§ FFQ-1, -2 and -3 refer to the first, second, and for men, third, factors based on the FFQ data; DR-1, -2 and 3 refer to the first, second and third factors based on the DR data.

Results

Correlations at the food-group level

Correlations between the intake of foods assessed by the FFQ and the intake of the corresponding groups of foods, assessed by the 7 d DR (Table 1), were similar for men and women, had a median coefficient of 0.43 and were within the range 0.10 (meat)–0.68 (milk products and rye bread with whole grains). We were puzzled by the very low correlation between meat measured by FFQ and the summed meat variable measured by the DR, as seen in Table 1, which, for women, was reflected in the difference of meat-loadings between the two methods (Table 2). The only significant Spearman rank correlations between the FFQ ‘meat’ variable and specific meats on the DR were: for men, higher fat pork (0.10) and rissole (0.09); for women, higher fat pork (0.11), higher fat beef (0.09), rissole (0.08), goose (0.08), lean pork (0.07) (results not shown). However, the low overall correlation may also result from the high clustering of answers in two of the eight food-frequency categories (‘once per d’ or ‘2–3 times per week’) in the FFQ data, in contrast to the broad near-normally distributed summed weight of the meat (‘per week’) in the DR data (results not shown). In recognition of the limited ability of the DR to measure a low frequency of intake, an alternative categorisation of both the FFQ and the DR was also carried out, in which the lower three categories of the FFQ were merged. Correlations were very similar to those found in Table 1 (results not shown).

Exploratory and confirmatory factor analyses using different dietary data

From the literature, we expected to find at least two factors by exploratory factor analysis: a ‘western’ or ‘traditional’ factor covering foods such as meats, potatoes and foods rich in fat; a ‘prudent’ or ‘green’ factor mainly identified by fruits, vegetables and wholegrain products (Togo *et al.* 2001). We identified a factor solution with three factors for men and two for women (Table 2). For men, the factors were labelled ‘green’ (high loadings on wheat bread and rye bread with whole grains and/or bran; raw and boiled vegetables and fruits), ‘sweet’ (cake, biscuits and baked goods; candy or chocolate; soft drink or ice cream and jam, marmalade or honey) and ‘traditional’ (meat; pâté and meat for bread; potatoes; butter; lard and hard margarine). For women, the factors were labelled ‘green’ (high loadings on the same foods as for men) and ‘sweet–traditional’ (cake, biscuits and baked goods; candy or chocolate; pâté and meat for bread white (wheat) bread and butter, lard and hard margarine).

Using the exploratory factor analysis on FFQ data, the total explained variances attributable to each factor were (%): for men ‘green’ 12.0, ‘sweet’ 10.2, ‘traditional’ 8.3; for women, ‘green’ 12.9, ‘sweet–traditional’ 10.9 (and slightly lower for the same factors based on the DR data, Table 2). For men, the correlation between the ‘green’ and ‘traditional’ factor scores, based on the exploratory factor analysis, was -0.2 ($P < 0.01$) regardless of diet assessment method, while correlations between the ‘green’

Table 4. Pearson correlation coefficients between factor scores on similar factors based on food-frequency questionnaire (FFQ) and diet record (DR) data in exploratory and confirmatory factor analyses (FA) respectively*†

| Correlations between factors based on | Men | | | Women | |
|---|---------|---------|---------------|---------|---------------------|
| | 'Green' | 'Sweet' | 'Traditional' | 'Green' | 'Sweet-Traditional' |
| Different dietary assessment methods | | | | | |
| Exploratory FA on FFQ data v. exploratory FA on DR data | 0.61 | 0.55 | 0.34 | 0.61 | 0.57 |
| Confirmatory FA on FFQ data v. confirmatory FA on DR data | 0.57 | 0.53 | 0.37 | 0.64 | 0.56 |
| Different factor analysis methods | | | | | |
| Exploratory FA on FFQ data v. confirmatory FA on FFQ data | 0.95 | 0.94 | 0.91 | 0.95 | 0.96 |
| Exploratory FA on DR data v. confirmatory FA on DR data | 0.82 | 0.90 | 0.90 | 0.94 | 0.94 |

* For details of subjects and procedures, see Table 1 and pp. 668–672.

† All correlations, $P < 0.01$.

and 'sweet' and between the 'sweet' and the 'traditional' factor scores were insignificant. For women, the correlation between the 'green' and 'sweet-traditional' factors was small (-0.1) and only significant using the FFQ data (results not shown).

The major differences (relative to the cut point of ± 0.3) between the loadings on the three factors based on the FFQ and the factors based on the DR data were as follows (Table 2). For men, fish, rice, cheese, milk and yoghurt loaded less on the 'green' factor; white bread, butter, lard and margarine loaded relatively more on the 'sweet' factor; sausages, white bread and eggs loaded less positively and wheat bread with whole grains and/or bran loaded more negatively on the 'traditional' factor, when based on the DR data. For women, rice, white bread and eggs loaded less on the DR-based 'green' factor; meat, potatoes and sausages loaded less on the DR-based 'sweet-traditional' factor.

Factor loadings in the confirmatory factor analyses based on the foods loading > 0.3 in the exploratory factor analyses of the same population are given in Table 3. For men, the model fits estimated as root mean square error of approximation were 0.079 in the FFQ-based data and 0.074 in the DR-based data; for women the model fits had a root mean square error of approximation of 0.085 in the FFQ-based data and 0.074 in the DR-based data (preferably < 0.05 – 0.10 , Loehlin, 1998; results not shown). As expected, the correlations between the exploratory factor analysis and confirmatory factor analysis-based equivalent scores were high when using the same dietary data (Table 4; FFQ, all correlations > 0.9 ; DR, all correlations > 0.8). The correlations between the factor scores based on the confirmatory factor analysis were, in general, higher than the same correlations based on the exploratory factor analysis (using the FFQ and DR data respectively): for men 'green' and 'sweet' 0.2 and 0.5, 'green' and

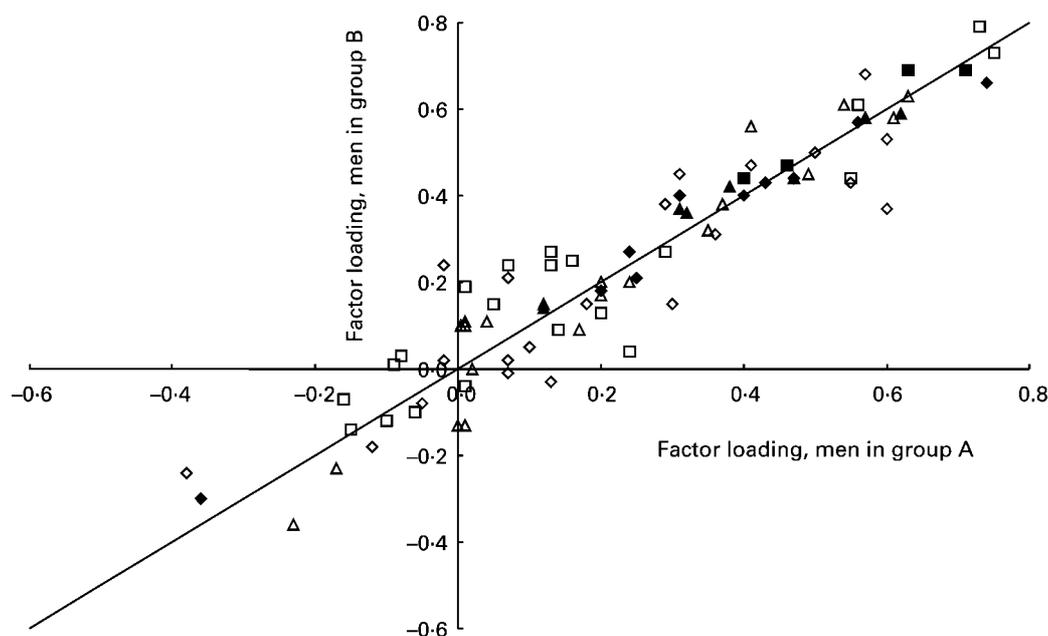


Fig. 2. Factor loadings (each data point represents a food group) on three factors found in exploratory factor analysis (EFA) of food-frequency questionnaire data and confirmatory analysis (CFA), based on male subjects who filled in a diet record as well as a food frequency questionnaire (group A), plotted against factor loadings based on the male subjects who only filled in the food-frequency questionnaire (group B). \diamond , 'Green' by EFA; \blacklozenge , 'green' by CFA; \square , 'sweet' by EFA; \blacksquare , 'sweet' by CFA; \triangle , 'traditional' by EFA; \blacktriangle , 'traditional' by CFA. For details of subjects and procedures, see pp. 668–670.

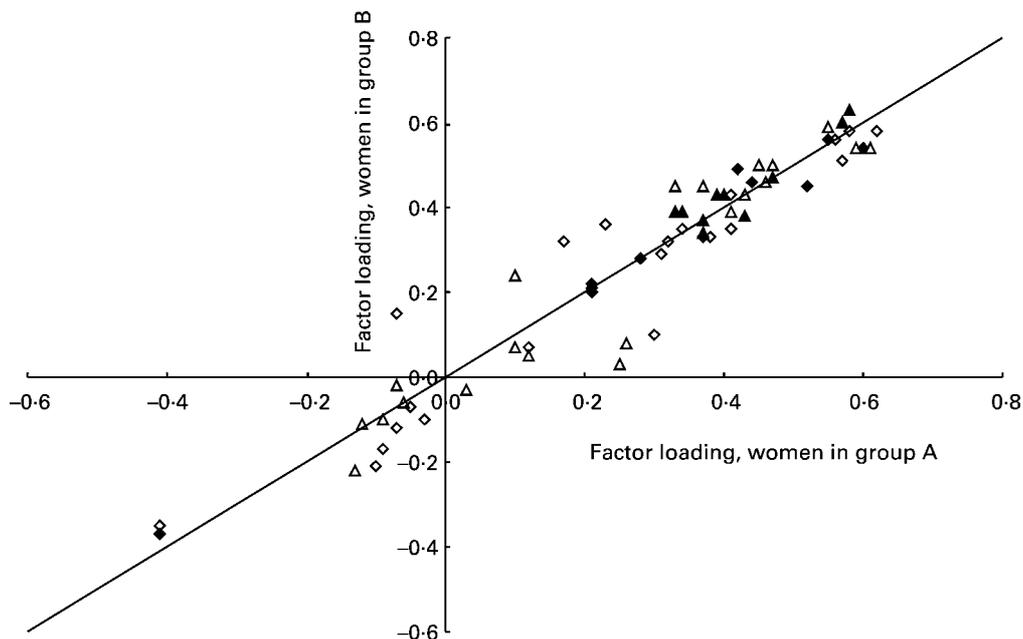


Fig. 3. Factor loadings (each data point represents a food group) on three factors found in exploratory factor analysis (EFA) of food-frequency questionnaire data and confirmatory analysis (CFA), based on female subjects who filled in a diet record as well as a food frequency questionnaire (group A), plotted against factor loadings based on the female subjects who only filled in the food-frequency questionnaire (group B). \diamond , 'Green' by EFA; \blacklozenge , 'green' by CFA; \square , 'sweet' by EFA; \blacksquare , 'sweet' by CFA; \triangle , 'traditional' by EFA; \blacktriangle , 'traditional' by CFA. For details of subjects and procedures, see pp. 668–670.

'traditional' -0.2 and 0.3 , 'sweet' and 'traditional' 0.2 and 0.01 ; for women 'green' and 'sweet-traditional' -0.2 and -0.4 .

Nutritional correlates of the factor scores

Due to the high correlations between the exploratory factor analysis- and confirmatory factor analysis-based factor scores (Table 4), only nutritional correlates of the factor scores based on confirmatory factor analysis are reported in Table 5 (all correlations were practically unchanged by factor analysis method; in Table 5 only confirmatory factor analyses-based factor scores were shown).

For men, the FFQ-based 'sweet' and the 'traditional' scores were weakly, positively correlated with total E based on the DR (0.15 and 0.17). The 'green' score was not significantly correlated with total E intake. The correlations were greater for the factor scores based on the DR data. E density, based on the DR, was not correlated with the 'green' factor score but positively with the 'sweet' and 'traditional' scores when scores were based on the DR. Only the positive correlation between E density and the FFQ-based 'sweet' score was significant. For women, only the 'sweet-traditional' factor was significantly positively correlated with E in both the FFQ and the DR analysis while in the DR the 'green' score was also positively correlated with E intake. Scores on the 'green' factor were inversely correlated and scores on the 'sweet-traditional' factor were positively correlated with E density.

Moderate correlation coefficients were found between the % E from protein, carbohydrate and fat and the factor scores (only adjusted correlations above 0.2 or

below -0.2 are described here; see Table 5 for details). Adjusting for age and total E intake did not considerably alter the coefficients. For men, relatively stronger correlations were seen between the 'sweet' factor score and % E from carbohydrate (positive) as well as between the 'traditional' factor score and % E from fat (positive) and carbohydrate (inverse). For women, the 'green' score was positively correlated with % E from protein and carbohydrate and negatively with % E from fat, while the 'sweet-traditional' score was negatively correlated with % E from protein, positively with % E from fat and non-correlated with % E from carbohydrate.

Correlations between factor scores stratified by covariates

The correlations between the factor scores from the FFQ data and the corresponding factor scores from the categorical DR data were between 0.34 and 0.61 in men and between 0.56 and 0.64 in women (Table 4). Only minor differences were seen in the correlations across subgroups of age, BMI, E intake, education, leisure-time physical activity and smoking. In men, direct comparison of correlations in different strata showed that the 'green' and the 'sweet' score correlations were higher in men with higher education than in the men with basic education. Likewise the 'green' score correlation was higher in the moderately or very active men compared with the sedentary men. For women, the 'green' score correlation was lower in the highly physically active women than in the moderately active or sedentary women. The 'traditional' and 'sweet-traditional' factor score correlations were not changed significantly by stratification (results not shown).

Table 5. Pearson correlation coefficients between factor scores (based on confirmatory factor models) and total energy intake, energy density and partial correlation coefficients between factor scores and protein, carbohydrate and fat intake as % energy (E) (from diet records, DR) controlled for age and total energy†

| | 'Green' | | 'Sweet' | | 'Traditional' | |
|--------------------------|---------|--------|---------|---------|---------------|--------|
| | FFQ-1§ | DR-1 | FFQ-2§ | DR-2 | FFQ-3§ | DR-3 |
| Men | | | | | | |
| Correlation with | | | | | | |
| Energy (kJ) | 0.03† | 0.11* | 0.15* | 0.42* | 0.17* | 0.50* |
| Energy density (kJ/g) | -0.03† | 0.02† | 0.18 | 0.24* | 0.00† | 0.18* |
| Partial correlation with | | | | | | |
| Protein (% E) | 0.19* | 0.10* | -0.16* | -0.18* | -0.13* | -0.05† |
| Carbohydrate (% E) | 0.20* | 0.23* | 0.27* | 0.37* | -0.06* | -0.29* |
| Fat (% E) | -0.13* | -0.15* | 0.07* | -0.00† | 0.14* | 0.37* |
| Women | | | | | | |
| Correlation with | | | | | | |
| Energy (kJ) | -0.02† | 0.25* | 0.32*¶ | 0.47*¶ | | |
| Energy density (kJ/g) | -0.16* | -0.12* | 0.37*¶ | 0.47*¶ | | |
| Partial correlation with | | | | | | |
| Protein (% E) | 0.28* | 0.45* | -0.32*¶ | -0.49*¶ | | |
| Carbohydrate (% E) | 0.15* | 0.22* | 0.02†¶ | -0.00†¶ | | |
| Fat (% E) | -0.29* | -0.36* | 0.21*¶ | 0.28*¶ | | |

FFQ, food-frequency questionnaire.

* $P < 0.05$.

† NS.

‡ For details of subjects and procedures, see Table 1 and pp. 668–672.

§ FFQ-1, -2 and -3 refer to the first, second, and, for men, third, factors based on FFQ data.

|| DR-1, -2 and -3 refer to the first, second and third factors based on the DR data.

¶ Sweet-traditional.

Exploratory and confirmatory factor analyses using different sub-populations

Exploratory factor analyses were also conducted on data from the subjects in group B, and the factors were compared with those identified (as described earlier) in the subjects in group A. There was a good agreement between the loadings, as illustrated in Figs 2 and 3 (R^2 0.80, 0.87 and 0.92 for 'green', 'sweet' and 'traditional' factors in men respectively, and 0.90 for both the 'green' and 'sweet-traditional' factors in women). Two confirmatory factor models were analysed in data from groups A and B; the models had the same number of factors, identified by the same foods (based on group A as described earlier), but allowing different loadings on the factors and correlations between factors). Relatively small differences in model fit were observed. The root mean square errors of approximation were 0.079 and 0.082 for men and 0.085 and 0.105 for women in groups A and B respectively; all loadings were significant in both groups (results not shown). The values of R^2 between the loadings in the two analyses conducted as confirmatory factor analysis were 0.98, 0.94 and 0.97 for the 'green', 'sweet' and 'traditional' factors in the men and 0.98 and 0.86 for the 'green' and 'sweet-traditional' in the women respectively (results not shown). Analyses in which the loadings were fixed to the same value to estimate the mean structure of the two groups showed only minor differences in the mean value of the factor score between the two groups. As expected, the participants in group A, on which the model was based (and who also completed the DR), had slightly higher scores than those in group B (results not shown).

Discussion

The present study had four main results: (1) Using the same subjects, three food-intake factors, defined using FFQ data, were readily identified in data from 7 d DR; (2) Correlations between factor scores based on FFQ and diet-record data varied, but were, in general, fairly high and did not vary much across subgroups defined by age, BMI, E intake, education, physical activity or smoking; (3) The same factors emerged when exploratory factor analysis was conducted in two equally sized subsamples of the population, defined by participation or non-participation in the DR part of the study; (4) The use of confirmatory analysis was illustrated and showed a reasonable model fit that did not vary considerably with use of dietary assessment method or different population groups. In addition, factor scores were more strongly (but not highly) correlated with total E when based on the DR than on the FFQ, and only modest correlations were seen between factor scores and % E from macronutrients.

Agreement with other studies

The three factors 'green', 'sweet' and 'traditional' (although with some variations in foods and labels) were similar to those found in studies by Gex-Fabry *et al.* (1988), Whicelow & Prevost, (1996) and Mullen *et al.* (2000). 'Green' and 'traditional' factors were also identified in five other studies (Barker *et al.* 1990; Slattery *et al.* 1998; Hu *et al.* 2000; Maskarinec *et al.* 2000; Tseng & DeVillis, 2000), while factors resembling the 'green' and 'sweet' factors were found in the studies of

Randall *et al.* (1991), Beaudry *et al.* (1998) and Schulze *et al.* (2001). This suggests that the factors we identified represent common components of the diet that can be identified in different populations. From a statistical point of view, the components or factors represent convenient mathematical constructs of the correlated foods included. However, the factors could also be interpreted as indicators of different traits controlling the individual's food intake. In that context, the factors identified in the present study could represent a higher compliance with dietary recommendations to adapt a Mediterranean diet (the 'green' factor), a craving for sweet and tasty foods (the 'sweet' factor) or a continuation of established dietary habits (the 'traditional' factor).

Variation with the dietary assessment method

In the present study, the factors were identified by exploratory factor analysis on data from a short FFQ and the factors replicated in a structured confirmatory factor analysis. Applying the same factor analysis structure on dietary input assessed by a 7 d DR gave quite similar loadings of the foods on each factor. It may be argued that a great similarity was to be expected, as the analysis was done on similar food groups and in the same individuals. However, the similarity was not given in advance, since measuring the habitual frequency of food intake 'over the previous year' and the actual weight of intake during a limited period, such as 1 week, may not result in similar food intake patterns or factors or similar loadings if the factor structure is fixed. In addition, the known problem of spurious correlations due to correlated errors occurring when comparing two observations on the same individuals (Willett, 1998) is probably not critical in the present study, given the very different principles of the diet assessment methods being compared.

The difference in loadings between the two methods that was seen for fish, rice and sausages may be explained by the relatively low intake of those foods that also loaded relatively low on the 'original' FFQ factors. The negative correlation between white bread (the only significant negatively loading food) and wheat bread with whole grains and/or bran played an important role in men, and to some extent in women, as markers within the 'green' factor. However, for men, the relationship between white bread and the other factors was less clear and seemed to change from a high loading on the 'traditional' factor in the FFQ data, to a high loading on the 'sweet' factor, followed by butter, lard and margarine. This could be explained by a tendency to eat, in particular, white bread (or bread rolls, which were included in the DR summary variable) with butter, and none of the other types of spreads that were used with other types of bread. In general, it would be expected that such 'traditions' in the coupling of foods were reflected more strongly in the factors based on the DR or other direct measurement of the food intake than in a FFQ without portion size (and therefore not quantifiable into mass or E intake) as we used.

Distributional differences, such as for meat, may also explain, in part, other discrepancies between the loading patterns of the FFQ and the DR data in the exploratory

factor analysis in Table 2. The low loadings in the exploratory factor analysis of sausages and, for women, meat also led to the exclusion of those foods in the 'traditional' and 'sweet-traditional' factors in the DR-based confirmatory factor analysis while playing a major role in the 'similar' factor for men in the FFQ data. This could lead to some misclassification depending on which dietary assessment method was used.

In order to exclude the possibility that the categorisation of the food intake based on the DR into octiles had an influence on the number or characteristics of the factors, we also conducted most of the exploratory analyses using DR data expressed as g per week. The results from these analyses gave factors that were very similar to the factors found using the octile data (results not shown).

Associations with macronutrient and energy intakes

The stronger correlations between DR-derived and the FFQ-derived factor scores and total E were expected due to the more direct link between the scores and the mass and E of the foods within the factors. The correlation with E density, however, was less affected by the dietary assessment method behind the scores. This similarity in correlations (in terms of either inverse or positive correlation with the factor scores) also applied for the macronutrient E contribution, while the magnitude of the correlations was only modest. It was not expected that factor scores would entirely predict individual nutrient intake. Hu *et al.* (1999) also pointed out this limitation, and argued against the use of dietary patterns for analyses of disease associations where a single substance (e.g. folic acid) is regarded as having a particular importance. However, even when using the DR as a reference method, the assessment of nutrients may be biased in a number of ways, related to age, obesity, education etc., due to its dependence on self-reporting (Fricker *et al.* 1992; Heitmann, 1993; Schoeller, 1995).

Consistency of factor scores

Despite different dietary assessment methodology, the correlations in the present study between the corresponding factor scores from the two different dietary assessment methods were surprisingly close to the correlations found in a similar study (Hu *et al.* 1999).

Stratification on the selected covariates in the present study produced a consistent general pattern and, due to multiple testing, the few significant differences across strata should be interpreted with caution. The relative robustness of the three-factor scores makes them a useful tool to characterise subjects by the importance of the three factors of the habitual diet ('green', 'sweet' and 'traditional' or a combination of the latter two for women).

In summary, the three food-intake factors appear to have a sound basis in terms of a recognisable content and nutrient profile, and they resemble factors found in other studies in different populations. Hence, we suggest that the food factors found in the present study may be used in analysis of different diet-disease relationships; however, a number of limitations should be considered when applying the factor scores to diet-disease models.

Limitations

One important limitation is the difficulty of investigating a hypothesis that food intake pattern is related to a disease or condition, especially if the association between food, described by other variables, e.g. nutrient E contribution, and the disease is weak or uncertain. The general limitations and usefulness of factor analysis and other methods of defining food intake patterns have been discussed in previous papers (Martinez *et al.* 1998; Slattery & Boucher, 1998; McCann *et al.* 2001; Togo *et al.* 2001; Hu, 2002; Schulze *et al.* 2003). With the use of confirmatory factor analysis, however, the significance of the loading foods was tested, and the same factor structure applied on results from the same population, assessed by a different method (i.e. the DR). Thereby, it was possible to reduce some of the limitations, i.e. the subjectivity involved in identifying the factors and the explorative nature of the method. This would particularly be the case if the factors were tested in their present form (given in the confirmatory factor analysis of Table 3) in a different study as true *a priori* factors.

In conclusion, replication of three food-intake factors using standardised information from a short FFQ and a DR gave similar factor loadings and factor scores. The correlations between the scores based on the two dietary assessment methods were consistent across strata of age, BMI, E intake, education, physical activity and smoking. Moreover, the same factor pattern was identified in two population groups, and the factors identified in one group could be applied in a different group using confirmatory factor analysis based on the same foods to compute scores. Hence, the factor scores were found to be acceptable and robust as markers of food intake pattern on group levels and may prove useful in studies of diet–disease relationships.

Acknowledgements

The authors thank the Research Unit for Dietary Studies at the Institute of Preventive Medicine (steering committee: Berit L. Heitmann, Lillian Mørch Jørgensen, Merete Osler, Agnes N. Pedersen and Marianne Schroll) and the Copenhagen County Research Centre for Prevention and Health for making data available. The establishment of the Research Unit for Dietary Studies was financed by the FREJA (Female Researchers in Joint Action) programme from the Danish Medical Research Council. The study was supported by The Danish Research Council FREJA programme, a grant from the University of Copenhagen, DK (j. nr. 301-116-5/99) and the Else and Mogens Wedell-Wedellsborgs Foundation (grant no. 664).

References

- Barker ME, McClean SI, Thompson KA & Reid NG (1990) Dietary behaviours and sociocultural demographics in Northern Ireland. *Br J Nutr* **64**, 319–329.
- Beaudry M, Galibois I & Chaumette P (1998) Dietary patterns of adults in Quebec and their nutritional adequacy. *Can J Public Health* **89**, 347–351.
- Fricker J, Baelde D, Igoin A, Huet JM & Apfelbaum M (1992) Underreporting of food intake in obese “small eaters”. *Appetite* **19**, 273–283.
- Gex-Fabry M, Raymond L & Jeanneret O (1988) Multivariate analysis of dietary patterns in 939 Swiss adults: sociodemographic parameters and alcohol consumption profiles. *Int J Epidemiol* **17**, 548–555.
- Heitmann BL (1993) The influence of fatness, weight change, slimming history and other lifestyle variables on diet reporting in Danish men and women aged 35–65 years. *Int J Obes Relat Metab Disord* **17**, 329–336.
- Hu FB (2002) Dietary pattern analysis: a new direction in nutritional epidemiology. *Curr Opin Lipidol* **13**, 3–9.
- Hu FB, Rimm E, Smith WS, *et al.* (1999) Reproducibility and validity of dietary patterns assessed with a food-frequency questionnaire. *Am J Clin Nutr* **69**, 243–249.
- Hu FB, Rimm EB, Stampfer MJ, Ascherio A, Spiegelman D & Willett WC (2000) Prospective study of major dietary patterns and risk of coronary heart disease in men. *Am J Clin Nutr* **72**, 912–921.
- Jacobson HN & Stanton JL (1986) Pattern analysis in nutrition. *Clin Nutr* **5**, 249–253.
- Jørgensen LM (1992) Who completes seven-day food records? *Eur J Clin Nutr* **46**, 735–741.
- Kant AK (1996) Indexes of overall diet quality: a review. *J Am Diet Assoc* **96**, 785–791.
- Kleinbaum DG, Kupper LL, Muller KE & Nizam A (1998) Testing for the equality of two correlations. In *Applied Regression Analysis and other Multivariable Methods*, 3rd ed., chapters 6 and 7, pp. 99–100. Pacific Grove, CA: Duxbury Press.
- Levnedsmiddelstyrelsen (1985) *Levnedsmiddeltabeller* (The Danish Food Composition Tables), 2nd ed. [A Møller, editor]. Frederikshavn, Denmark: Dafolo forlag, Dafolo A/S.
- Loehlin JC (1998) *Latent Variable Models: An Introduction to Factor, Path, and Structural Analysis*, 3rd ed. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- McCann SE, Weiner J, Graham S & Freudenheim JL (2001) Is principal components analysis necessary to characterise dietary behaviour in studies of diet and disease? *Public Health Nutr* **4**, 903–908.
- Martinez ME, Marshall JR & Sechrest L (1998) Invited commentary: Factor analysis and the search for objectivity (comment). *Am J Epidemiol* **148**, 17–19.
- Maskarinec G, Novotny R & Tasaki K (2000) Dietary patterns are associated with body mass index in multiethnic women. *J Nutr* **130**, 3068–3072.
- MONICA Data Centre (2000) The WHO MONICA project. <http://www.ktl.fi/Monica/>
- Mullen K, Williams R & Hunt K (2000) Irish descent, religion and food consumption in the west of Scotland. *Appetite* **34**, 47–54.
- Muthen LK & Muthen BO (2001) *Mplus User's Guide, Version 2*. Los Angeles, CA: Muthen & Muthen. <http://www.statmodel.com>
- Mørup I (1986) *Micro-CAMP®: Computer Assisted Menu Planning APS* (computer program; no longer available). Herlev, Denmark: Herlev University Hospital.
- Quatromoni PA, Copenhafer DL, Demissie S, *et al.* (2002) The internal validity of a dietary pattern analysis. The Framingham Nutrition Studies. *J Epidemiol Community Health* **56**, 381–388.
- Randall DE, Marshall JR, Brasure J & Graham S (1991) Patterns in food use and compliance with NCI dietary guidelines. *Nutr Cancer* **15**, 141–158.
- Schoeller DA (1995) Limitations in the assessment of dietary E intake by self-report. *Metabolism* **44**, 18–22.
- Schulze MB, Hoffmann K, Kroke A & Boeing H (2001) Dietary patterns and their association with food and nutrient intake in the European Prospective Investigation into Cancer and Nutrition (EPIC)–Potsdam study. *Br J Nutr* **85**, 363–373.

- Schulze MB, Hoffmann K, Kroke A & Boeing H (2003) An approach to construct simplified measures of dietary patterns from exploratory factor analysis. *Br J Nutr* **89**, 409–418.
- Slattery ML & Boucher KM (1998) The senior authors' response: Factor analysis as a tool for evaluating eating patterns (comment). *Am J Epidemiol* **148**, 20–21.
- Slattery ML, Boucher KM, Caan BJ, Potter JD & Ma KN (1998) Eating patterns and risk of colon cancer. *Am J Epidemiol* **148**, 4–16.
- Togo P, Osler M, Sørensen TIA & Heitmann BL (2001) Food intake patterns and body mass index in observational studies. *Int J Obes Relat Metab Disord* **25**, 1741–1751.
- Tseng M & DeVillis R (2000) Correlates of the “western” and “prudent” diet patterns in the US. *Ann Epidemiol* **10**, 481–482.
- Whicelow MJ & Prevost AT (1996) Dietary patterns and their associations with demographic, lifestyle and health variables in a random sample of British adults. *Br J Nutr* **76**, 17–30.
- Willett WC (1998) *Nutritional Epidemiology*. 2nd ed. *Monographs in Epidemiology and Biostatistics*, vol. 30. New York: Oxford University Press, Inc.
- Williams C, Wiseman M & Buttriss J (editors) (1999) Food-based dietary guidelines – a staged approach. Conclusions. *Br J Nutr* **81**, Suppl. 2, S153.
- World Health Organization (2000) Obesity: *Preventing and Managing the Global Epidemic*. Report of a WHO consultation, no. 894, p. i-253. Geneva, WHO.

Appendix

Estimation in SPSS (exploratory factor analysis with factor scores)

Principal component analysis on a correlation matrix was used for the extraction and Promax factor rotation (non-orthogonal) applied to allow correlated factors. Scores were computed using the regression method

option, allowing scores to be correlated even if factors are uncorrelated in the model.

Estimation in Mplus (confirmatory factor analysis with factor scores)

For the categorical indicator (food) variables the polychoric correlation coefficients were used. Polychoric correlation coefficients are based on an assumption that the categorical response variable observed is only an indicator of an underlying ‘true’ normally distributed food intake (the latent response variable), and thresholds in this latent variable determine the categorical response observed. In the confirmatory analysis, the weighted least square estimator with robust standard errors and mean- and variance-adjusted χ^2 test statistic was used. Individual factor scores were calculated by the use of minimisation techniques (Muthen & Muthen, 2001: pp. 385–386).

Comparison of correlations by Fisher transformation

Let x be the Pearson correlation coefficient, and z' the Fisher-transformed correlation coefficient, where $z' = \frac{1}{2} \ln \left(\frac{1+x}{1-x} \right)$, then the difference – variable (*Dif*) between two correlations z'_1 and z'_2 is given by the formula:

$$Dif = \frac{z'_1 - z'_2}{\sqrt{\frac{1}{n_1-3} + \frac{1}{n_2-3}}}$$

If *Dif* is outside the range -2.58 – 2.58 then the difference is significant at the 0.01 level. If *Dif* is outside the range -1.96 – 1.96 the difference is significant at the 0.05 level.