

## Identification and reproducibility of dietary patterns in a Danish cohort: the Inter99 study

Cathrine Lau<sup>1\*</sup>, Charlotte Glümer<sup>1,2</sup>, Ulla Toft<sup>2</sup>, Inge Tetens<sup>3</sup>, Bendix Carstensen<sup>1</sup>, Torben Jørgensen<sup>2</sup> and Knut Borch-Johnsen<sup>1</sup>

<sup>1</sup>Steno Diabetes Center, Niels Steensens Vej 2, 2820 Gentofte, Denmark

<sup>2</sup>Research Centre for Prevention and Health, Nordre Ringvej 57, 2600 Glostrup, Denmark

<sup>3</sup>Department of Nutrition, Danish Institute for Food and Veterinary Research, Mørkhøj Bygade 19, 2860 Søborg, Denmark

(Received 9 March 2007 – Revised 15 August 2007 – Accepted 30 August 2007 – First published online 22 October 2007)

We aimed to identify dietary patterns in a Danish adult population and assess the reproducibility of the dietary patterns identified. Baseline data of 3372 women and 3191 men (30–60 years old) from the population-based survey Inter99 was used. Food intake, assessed by a FFQ, was aggregated into thirty-four separate food groups. Dietary patterns were identified by principal component analysis. Confirmatory factor analysis and Bland Altman plots were used to assess the reproducibility of the dietary patterns identified. The Bland Altman plots were used as an alternative and new method. Two factors were retained for both women and men, which accounted for 15.1–17.4% of the total variation. The ‘Traditional’ pattern was characterised by high loadings ( $\geq 0.40$ ) on paté or high-fat meat for sandwiches, mayonnaise salads, red meat, potatoes, butter and lard, low-fat fish, low-fat meat for sandwiches, and sauces. The ‘Modern’ pattern was characterised by high loadings on vegetables, fruit, mixed vegetable dishes, vegetable oil and vinegar dressing, poultry, and pasta, rice and wheat kernels. Small differences were observed between patterns identified for women and men. The root mean square error approximation from the confirmatory factor analysis was 0.08. The variation observed from the Bland Altman plots of factors from explorative *v.* confirmatory analyses and explorative analyses from two sub-samples was between 18.8 and 47.7%. Pearson’s correlation was  $>0.89$  ( $P < 0.0001$ ). The reproducibility was better for women than for men. We conclude that the ‘Traditional’ and ‘Modern’ dietary patterns identified were reproducible.

### Dietary patterns: Reproducibility

Human beings do not eat isolated nutrients, but meals consisting of a variety of foods with complex combinations of nutrients. Hence, the single-nutrient approach may be inadequate to account for complicated interactions among nutrients in studies analysing diet–disease associations.

Previous analyses of dietary baseline data in the Inter99 study (a study aiming to prevent diabetes and CVD) have indicated that analyses of dietary patterns may give a more balanced description of data compared to analyses of single nutrients<sup>(1)</sup>. Use of explorative statistical methods is one way to examine dietary patterns in populations. This approach is based on empirical data and not on *a priori* hypothesis. Confirmatory analysis, by contrast, is based on *a priori* hypothesis. It may be guided both by results from an explorative analysis and by knowledge of nutritional behaviour. Confirmatory analysis can be used to assess the reproducibility of dietary patterns identified.

The dietary patterns identified in other studies using explorative principal component analysis (PCA) have been characterised by a relatively high variation in intake of one or more of the following food groups: vegetables, fruit, pasta, rice, whole grains, low-fat dairy products, fish and poultry<sup>(2–11)</sup> in one pattern; or red meat, potatoes, refined grains,

butter and lard, high-fat dairy products and confectionery in another pattern<sup>(2–7,9–11)</sup>. Generally two to four dietary patterns have been identified.

The reproducibility of the patterns is often questioned because most studies only use explorative analyses like the PCA. Few reproducibility studies on dietary patterns exist; some have combined explorative PCA and confirmatory factor analyses (CFA)<sup>(9,11,12)</sup> or applied the same procedure of PCA on two split-samples<sup>(9,13,14)</sup> or on the same cohort one or more years apart<sup>(3,10,12)</sup>. Both low and high reproducibility have been observed. The optimal method used to describe the reproducibility is still undetermined.

We aimed to identify dietary patterns in a Danish adult population (Inter99) and assess the reproducibility of the dietary patterns identified. Our first aim was to identify dietary patterns from explorative PCA. The second aim was to compare the results from the explorative analysis with the results from a CFA and an explorative analysis using a reduced number of food groups. The last aim was to compare results from explorative analyses between two split-samples. Bland Altman plots were used as an alternative and new method in the assessment of the reproducibility of identified dietary patterns.

**Abbreviations:** CFA, confirmatory factor analysis; PCA, principal component analysis; RV, relative variation.

\* **Corresponding author:** Cathrine Lau, fax +45 44 43 07 06, email cala@steno.dk

## Materials and methods

### Study population

This study used baseline data from the Danish population-based Inter99 study, which is an intervention study on diet, physical activity and smoking with the aim to decrease the incidence of CVD and diabetes. The overall aim, data collection methods, non-dietary and dietary baseline results have been reported elsewhere<sup>(15–17)</sup>.

The study base comprised 61 301 individuals born in 1939–40, 1944–45, 1949–50, 1954–55, 1959–60, 1964–65, and 1969–70, resident in eleven municipalities in the south-western part of Copenhagen County. An age- and sex-stratified random sample of 13 016 persons was drawn from the study base and 12 934 were eligible for further examination. All these individuals were invited for a health survey at the Research Centre for Prevention and Health in Glostrup. Baseline data were collected in 1999–2001, where 6784 (52.5 % of those invited) participated.

All participants gave written consent before taking part in the study. The protocol was in accordance with the Helsinki declaration and was approved by the local ethical committee (KA98155).

### Dietary data

The participants completed a self-administered 198-item FFQ, where they were asked to report their dietary intake during the previous month. Consumption of each food item in the FFQ was estimated in g/d. A detailed description of the questionnaire and estimation of the dietary intake in the population is published elsewhere<sup>(17)</sup>. However, in this study calculation of dietary intake was based on an updated version of the Danish Food Composition Data Bank<sup>(18)</sup>. Validation of the FFQ has shown that the FFQ provides a reasonable classification of individuals and a valid quantitative measurement of the dietary intake<sup>(19)</sup>.

We excluded individuals who had not filled in the FFQ, who had left five or more pages out of fourteen blank, or who had clearly misunderstood (including questionnaires where two or more answers per question line recurrently were ticked or questionnaires where only the highest response frequency was ticked repeatedly) the FFQ. A total of 6563 (96.7 %) individuals qualified for the present analyses.

Food intake assessed by questions in the FFQ was aggregated into thirty-four separate food groups (Appendix 1). The food groups were created based on food groups previously identified in Denmark<sup>(20)</sup>, the Danish food tradition, and the Danish dietary guidelines<sup>(21)</sup>. Furthermore, criteria for categorising food items into one group were that the items shared macronutrient composition (e.g. fat or fibre content) or use (e.g. paté and high-fat meat for sandwiches), and that each food group only reflected one type of eating behaviour (e.g. not mixing vegetarian food items with meaty food items).

### Statistical methods

Initially data were randomly split in two halves: sub-sample 1 and sub-sample 2 (Fig. 1). To identify and reproduce dietary patterns we applied PCA using the PROC FACTOR procedure and CFA using the PROC CALIS procedure in SAS version

8.2 (SAS Institute, Cary, NC, USA). Bland Altman plots were conducted to assess the agreement between the explorative analyses using R 2.1.0<sup>(22)</sup>.

### Identification of dietary patterns

Analyses of sub-sample 1 were conducted and decisions regarding the final factors were based on these results. We assumed, based on previous analyses<sup>(17)</sup>, that different dietary patterns would be identified for men and women. Hence, analyses were conducted separately for each gender, too.

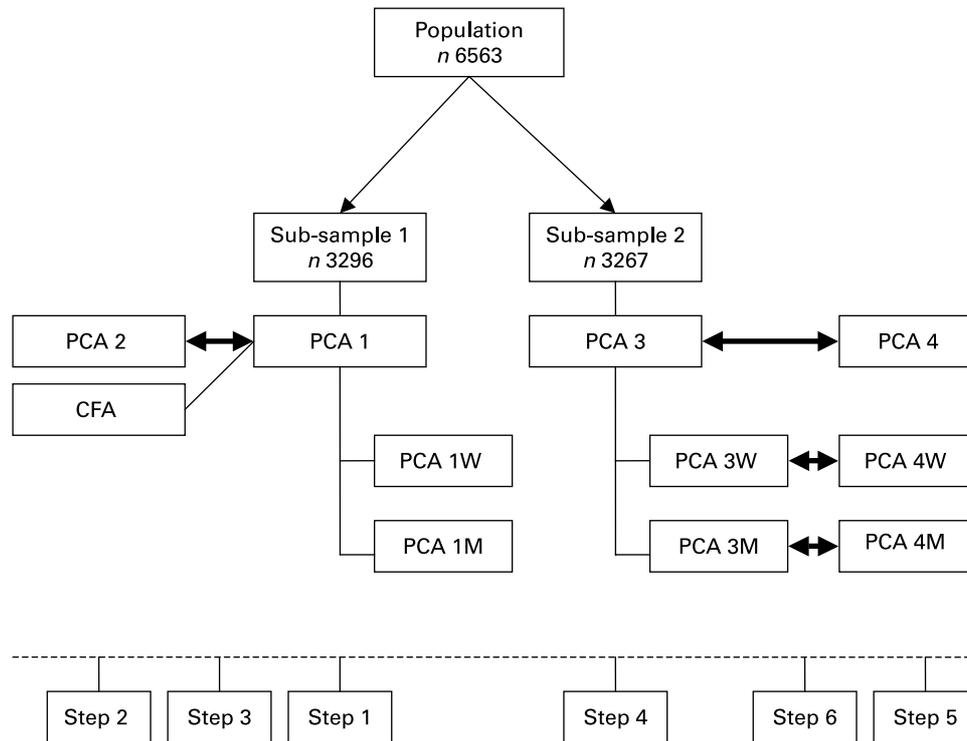
We conducted a PCA for sub-sample 1 (PCA 1; Fig. 1, step 1) using all the food groups (g/d) shown in Appendix 1. To identify the number of factors to be retained, we plotted the total variance associated with each factor (a Scree plot). After examination of the Scree plot we did an oblique transformation (promax rotation) to achieve a simpler structure with greater interpretability<sup>(23)</sup>. Food groups with absolute factor loadings  $\geq 0.40$  were considered as significantly contributing to a pattern. The larger the loading of a given food group to the factor, the greater the contribution of that food group to the factor. If at least three food groups had loadings  $\geq 0.40$  on one factor it was considered a potential pattern. The more food groups loading high on a factor the more reliable it will be<sup>(24)</sup>. We did not allow food groups to load  $\geq 0.40$  on more than one factor. PCA further provides factor scores that allow individuals to be ranked in terms of how closely they conform to the factor (dietary pattern).

### Reproducibility of dietary patterns

We conducted a CFA and a second PCA (PCA 2) on a reduced number of food groups in sub-sample 1 (Fig. 1, step 2). These analyses were both based on results from PCA 1 only including food groups with loadings  $\geq 0.40$  and on (*a priori*) knowledge regarding two food groups (whole-grain cereals and bread, and high-fat fish) that according to the Danish nutritional behaviour traditionally are important food items in the diet. The root mean square error approximation was used to assess goodness of fit in the CFA. It should be  $< 0.05$ – $0.10$  for a good fit<sup>(23)</sup>.

PCA 2 was conducted in order to estimate factor scores that could be compared with factor scores from PCA 1 in a Bland Altman plot (Fig. 1, step 3). In the Bland Altman plot the variation of the difference of the factor scores relative to the variation of average of them gives an impression of the reproducibility of the factor. The Bland Altman plot was used as an alternative and new method within this area. The degree of acceptable variation is a matter of judgement<sup>(25)</sup>, but a 'small' degree of variation is preferable.

To further assess the reproducibility of the factors identified in sub-sample 1 (PCA 1), we conducted analyses resulting in two sets of factor scores for sub-sample 2. First, a third PCA (PCA 3), following the same criteria as PCA 1, was conducted (Fig. 1, step 4) and factor scores of each factor were derived for each individual in sub-sample 2. Second, factor loadings on each factor estimated from PCA 1 were applied to the individuals in sub-sample 2, giving another set of factor scores in sub-sample 2. For simplicity this analysis is called PCA 4 (Fig. 1, step 5). The factor scores of each factor from PCA 3 and 4 were compared by Bland Altman plots



**Fig. 1.** Scheme of statistical analyses. PCA, principal component analysis; CFA, confirmatory factor analysis; W, women; M, men. PCA 2 and CFA included the following food groups: paté and high-fat meat for sandwiches, mayonnaise salads, butter and lard, low-fat meat for sandwiches, potatoes, whole-grain cereals and bread, red meat, sauces, low-fat fish, high-fat fish, vegetables, fruit, mixed vegetables dishes, vegetable oil and vinegar dressing, poultry, pasta, rice and wheat kernels. The factor scores from PCA 4 were based on factor loadings estimated in PCA 1 and food intake of individuals in sub-sample 2. The arrows (↔) indicate that factor scores are compared using Bland Altman plots.

(Fig. 1, step 6). These analyses were also done separately for men and women.

Pearson's correlation coefficient was calculated to assess the correlation between factor scores of similar factors from PCA 1 and 2, and PCA 3 and 4.

## Results

The study population comprised 3191 (48.6%) men and 3372 (51.4%) women with complete dietary data. Baseline characteristics and dietary intake are presented in Appendices 2 and 3. The two sub-samples were comparable.

### Identification of dietary patterns

For sub-sample 1 (PCA 1), two factors were identified whether the analyses were conducted on the entire sub-sample or for each gender (Table 1). We compared rotated factor solutions of two and four factors and found that the most meaningful factors were identified when only two factors were retained and rotated. We tested different rotation methods (varimax *v.* promax transformation) but did not find any significant differences in the final factors (data not shown). Hence, we allowed correlation between the factors using promax rotation.

The factors were either labelled 'Traditional' or 'Modern'. For the entire sub-sample 1, the factor 'Traditional' was characterised by high loadings ( $\geq 0.40$ ) on paté or high-fat meat for sandwiches,

mayonnaise salads, red meat, potatoes, butter and lard, low-fat fish, low-fat meat for sandwiches, and sauces. These food groups are traditionally eaten at lunchtime and in the evening in Denmark resulting in the label 'Traditional' for this factor. The 'Modern' factor was characterised by high loadings on vegetables, fruit, mixed vegetable dishes, vegetable oil and vinegar dressing, poultry, and pasta, rice and wheat kernels. These food groups are normally included in a modern and recommended diet, consequently the factor was labelled 'Modern'. Some differences were observed between loadings of food groups on factors for women and men given the same label as factors of the entire sub-sample 1. Small differences were observed between women and men (Table 1). Red meat had a loading of 0.38 on the 'Traditional' factor for women, but a loading of 0.42 (high) on the 'Modern' factor for men. Low-fat meat for sandwiches did not load high on any of the factors for women, but had a loading of 0.42 on the 'Traditional' factor for men. Poultry, and pasta, rice and wheat kernels had loadings above 0.40 on the 'Modern' factor for men, but not for women. Thus, the 'Modern' factor was characterised by more variation in meat intake for men than for women. The total variance explained by the factors in sub-sample 1 was 17.1% (15.4% for women, 17.0% for men).

From PCA 2 on sub-sample 1 we also identified two factors (a 'Traditional' and a 'Modern' factor) with very similar factor loadings to those identified from PCA 1 (difference in loadings  $< 0.07$ ). In sub-sample 2 (PCA 3) the same number of meaningful factors was identified also. The greatest difference observed between the factors in sub-sample 1 (PCA 1)

**Table 1.** Factor loading matrix of major factors (dietary patterns) identified using principal component analysis with promax rotation on food-frequency baseline data from sub-sample 1 in the Inter 99 study

Factor labels	All		Women		Men	
	Traditional	Modern	Traditional	Modern	Traditional	Modern
Food groups*						
Paté or high-fat meat for sandwiches	0.62	-0.24	0.60	-0.10	0.67	-0.03
Mayonnaise salads	0.50	-0.18	0.46	-0.14	0.54	0.01
Red meat	0.50	0.14	0.38	0.22	0.21	0.42
Potatoes	0.48	-0.04	0.46	0.13	0.42	0.09
Butter and lard	0.47	-0.20	0.45	-0.08	0.51	-0.03
Low-fat fish	0.44	0.30	0.23	0.41	0.29	0.42
Low-fat meat for sandwiches	0.42	0.01	0.34	0.14	0.42	0.11
Sauces	0.40	-0.25	0.43	-0.22	0.47	-0.05
Vegetables	0.02	0.62	-0.02	0.63	-0.04	0.58
Fruit	-0.10	0.58	-0.09	0.57	-0.11	0.47
Mixed vegetable dishes	0.04	0.58	-0.01	0.55	-0.19	0.56
Vegetable oil and vinegar dressing	0.17	0.51	-0.08	0.48	-0.08	0.57
Poultry	0.09	0.46	0.01	0.38	-0.16	0.53
Pasta, rice and wheat kernels	0.23	0.44	0.11	0.38	-0.09	0.56
Whole-grain cereals and bread	0.38	0.08	0.26	0.31	0.38	0.08
Eggs	0.38	0.08	0.35	0.04	0.35	0.27
High-fat fish	0.37	0.25	0.23	0.36	0.25	0.32
Refined grain cereals and bread	0.34	0.18	0.30	0.21	0.21	0.29
Fast food	0.34	-0.04	0.32	-0.09	0.14	0.18
Cakes	0.30	0.12	0.32	0.09	0.23	0.27
Coffee	0.18	-0.34	0.13	-0.33	0.27	-0.21
Tea	-0.12	0.31	-0.09	0.26	-0.17	0.26
Salty snacks	0.25	-0.01	0.37	-0.11	0.10	0.18
Margarine	0.28	-0.18	0.35	-0.18	0.32	-0.07
Dairy products	0.21	0.13	0.32	0.17	0.13	0.14
Nuts	0.17	0.29	0.03	0.26	0.09	0.38
Creamy dressing	0.13	0.25	0.14	0.21	0.09	0.30
Explained variance (%)	9.30	7.82	7.92	7.50	7.52	9.49

\* Sorted by factor loadings on 'Traditional' and 'Modern' factor for all in sub-sample 1. Food groups with loadings <0.30 for all the factors listed were excluded from the Table. These were: mix of butter, margarine and oil; cheese; ice cream and chocolate milk; confectioneries; sweet spread; alcohol; and miscellaneous foods.

and sub-sample 2 (PCA 3) regarded low-fat fish and margarine. These food groups had loadings of 0.44 and 0.28 on the 'Traditional' factor in sub-sample 1, but loadings of 0.17 and 0.42, respectively, on the 'Traditional' factor in sub-sample 2. Furthermore, low-fat fish had a loading of 0.40 on the 'Modern' factor in sub-sample 2. Apart from these differences loadings between food groups significantly contributing to the patterns differed <0.15.

#### Reproducibility of dietary patterns

The second aim of our study was to compare results from PCA 1 with those from a CFA in sub-sample 1. The loadings observed in the CFA were very similar to those observed in PCA 1 (difference in loadings <0.15). The CFA resulted in a root mean square error approximation value equal to 0.08 indicating a good fit.

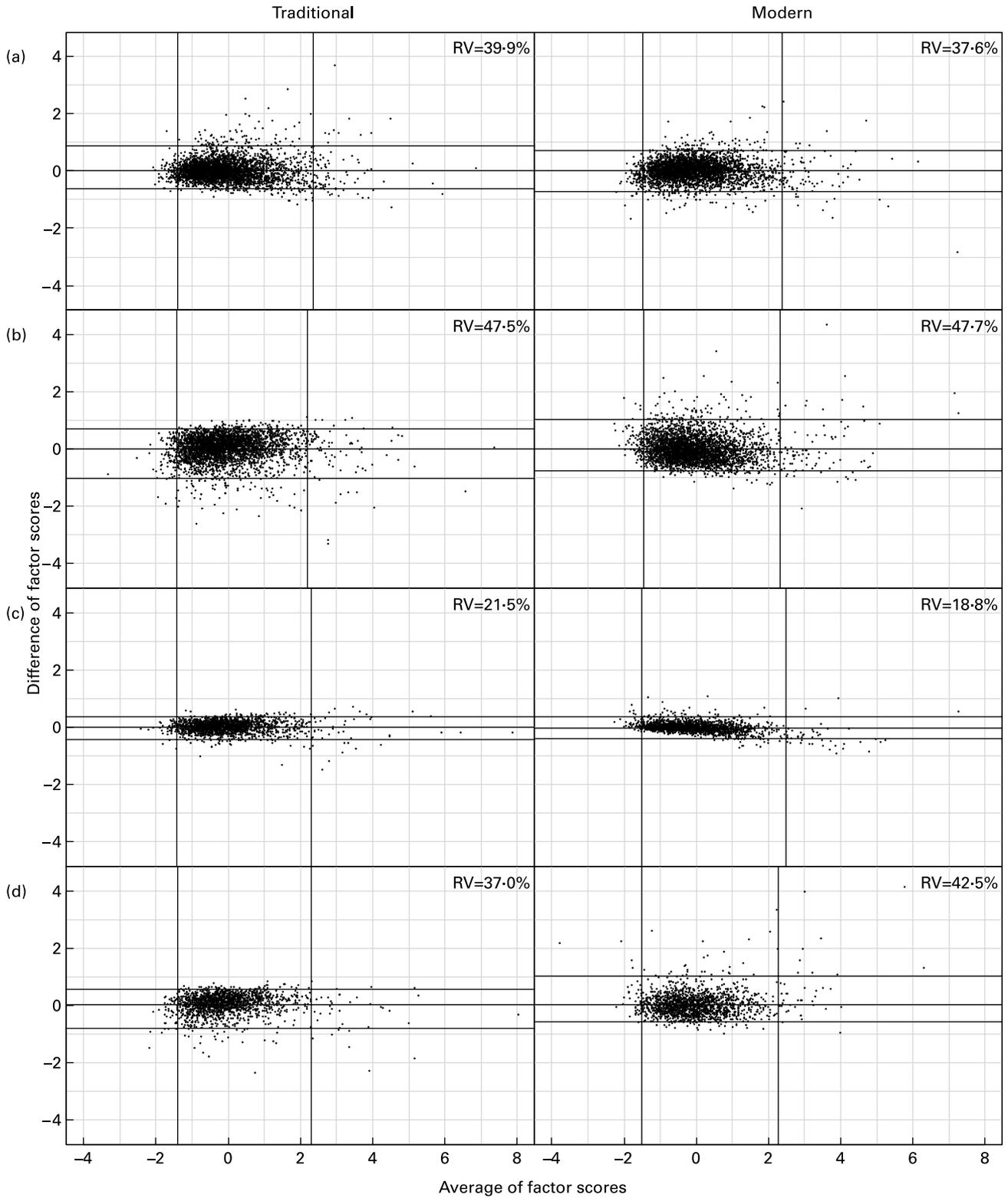
Furthermore, we aimed to assess agreement between PCA 1 and PCA 2 in sub-sample 1 using a Bland Altman plot (Fig. 2(a)). No systematic bias, across the range of average factor scores of similar factors from PCA 1 and 2, was identified. The variation in the difference between factor scores of the 'Traditional' factors from PCA 1 and 2 corresponded to 39.9% of the variation in the average of the factor scores from these analyses and for the 'Modern' factors this relative variation (RV) was 37.6%.

The reproducibility of the factors in the second sub-sample was then assessed. The variance in sub-sample 2 between the factors scores derived from factor loadings estimated in sub-sample 1 (PCA 4) and the factors scores derived within sub-sample 2 (PCA 3) are presented as Bland Altman plots in Fig. 2(b, c and d). No systematic bias, across the range of average factor scores of similar factors from PCA 3 and 4, was identified. For all in sub-sample 2, the RV regarding the 'Traditional' factors from PCA 3 and 4 corresponded to 47.5% and for the 'Modern' factors the RV was 47.7%. For women in sub-sample 2, the RV corresponded to 21.5 and 18.8% for the 'Traditional' and 'Modern' factor, respectively. For men in sub-sample 2, the RV corresponded to 37.0 and 42.5%, respectively.

Finally, Pearson's correlation was calculated. The correlation between the factor scores from PCA 1 and 2 was 0.93 ( $P < 0.0001$ ) for both the 'Traditional' and the 'Modern' factor. The correlations between the factor scores from PCA 3 and 4 were 0.89, 0.98 and 0.90 ( $P < 0.0001$ ) for the 'Traditional' factor in all, women and men, respectively, and 0.89, 0.99, 0.93 ( $P < 0.0001$ ) for the 'Modern' factor in all, women and men, respectively.

#### Discussion

Two major patterns were identified in the population: a 'Traditional' and a 'Modern' pattern. These patterns, with slight



**Fig. 2.** Bland Altman plots. Difference v. average of factor scores from (a) PCA 1 and 2 in sub-sample 1, and from PCA 3 and 4 in sub-sample 2 for (b) all, (c) women, and (d) men. Empirical 2.5 and 97.5% quantiles of the difference and average of the factor scores together with the mean of the difference are given in the plots. The relative variation (RV) is the variation of the difference of the factor scores relative to the variation of average of the factor scores ( $RV = 95\% \text{ CI of difference of factor scores divided by } 95\% \text{ CI of average of factors scores}$ ).

differences between food groups with high loadings, were also observed for each gender both in sub-sample 1 and 2.

Reproducibility of the patterns according to the CFA was acceptable. The RV and the correlations between factors from PCA 1 and 2, and PCA 3 and 4 on all individuals indicated that reproducibility of the patterns was moderate to high. The reproducibility of the patterns was better for women than for men, when analysed separately.

#### *Identification of dietary patterns*

Dietary patterns resulting from *a posteriori* analytical approaches across studies are complicated to compare because of differences in selected food groups, in assessment and analytical tools, and in decision criteria for the extracted factors. However, there seems to be some homogeneity of dietary patterns across populations. With the use of PCA, several studies have identified a 'Western/Traditional' pattern<sup>(2-4,6,7,9,10,12,26)</sup> and a 'Prudent/Healthy' pattern<sup>(2-4,6-10,12,26)</sup>, but minor differences in dietary patterns between studies are always observed. In our study, whole-grain cereals and bread did not load high on the patterns identified from the PCA, whereas whole grains in most other studies load high on the 'Prudent/Healthy' pattern<sup>(2,3,6,7,9,12,26)</sup>. It may be that the food group including whole grains in our study was more diverse as it included rye bread, whole-grain bread, oats and muesli. The variation in the intake may therefore have been too small to be detectable in the PCA analyses. Another example is fish, which was part of the 'Traditional' pattern in sub-sample 1 analyses. When the analyses were conducted for men and women, separately, or on sub-sample 2 low-fat fish had a high loading on the 'Modern' pattern. These latter findings are more congruent to most other studies<sup>(3,7,9,10,12)</sup>. Further, poultry contributed to the 'Modern' pattern in our study but not to the 'Prudent/Green' patterns in DAN-MONICA<sup>(2,9)</sup>. A possible explanation is that poultry was not a separate food group in DAN-MONICA, but was included in a larger meat group with both red and white meat. Difference in the grouping of poultry may simply be a result of the time difference between the studies (7–14 years between DAN-MONICA and Inter99). Poultry has only recently achieved status as healthy and modern meat to eat compared to red meat. Comparison with the Danish DAN-MONICA studies was obvious because differences caused by country-specific differences in dietary habits were eliminated. Altogether, our results are similar because we, like Osler and co-workers<sup>(2)</sup>, identified two dietary patterns of similar food and nutrient composition. Similar to the findings of Togo and colleagues<sup>(9)</sup>, we also identified small differences between the patterns for men and women, which agree with our hypothesis. That patterns might vary slightly between sexes has been observed also in other studies<sup>(27)</sup>. We labelled the dietary patterns of the entire sub-sample 1 and dietary patterns of women and men the same because most food groups contributing to the patterns for each gender were the same as those contributing to the patterns in the gender-combined sample.

The proportion of variance explained by the factors was between 15.4 and 17.1% in this study. Other studies that apply PCA to dietary data and identify two patterns usually find that 17–25% of total variance is explained<sup>(2,3,6,8,9,26)</sup>. Hence, variance explained by factors in our study was

relatively low. Findings from this and other studies<sup>(2,3,6,8,9,26)</sup> indicate that the number of food groups entering PCA affects the total variance explained by the factors. The larger the number of food groups entering the PCA, the lower is the proportion of total variance explained by the factors. Compared to studies where nearly 25% of the variance was explained<sup>(8,9)</sup> we included twelve to thirteen food groups more. Further, the lower the number of factors retained before rotation and score calculation or low variation in the diet may add to greater variation explained by the factors in the current studies.

In the identification of the dietary patterns we sought to use statistical methods as comparable as possible to methods used in other studies identifying dietary patterns. Furthermore, the number and/or nutrient composition of food groups entering the PCA was to a large extent similar to those observed in other studies<sup>(3,6,9,27,28)</sup>, which may make comparison of the final number and character of the factors with other studies more straightforward and reliable.

If the diet in itself is a potential risk factor for different diseases, then the dietary patterns identified will be more likely to include the risk factor than single nutrients or foods, as the dietary patterns explain more variation. However, this does not rule out the possibility that single nutrients or foods may be of great importance for disease development. But dietary pattern analysis is relevant when analysing the overall effect of the diet in population-based studies. The dietary patterns identified in this study enable future analyses on the effect of overall diet in relation to disease.

#### *Reproducibility of dietary patterns*

CFA was used to examine the robustness and goodness of fit of the factor structures derived from the PCA and our understanding of the Danish nutritional behaviour (e.g. that rye bread is traditionally eaten at lunchtime in Denmark (S. Fagt, personal communication)). CFA is an intuitively appealing method because it can be based in theory and also reduces some of the subjectivity involved in explorative procedures<sup>(29)</sup>. As in other studies using the same approach<sup>(9,11)</sup> we demonstrated according to the root mean square error approximation a reasonable value (<0.10) of goodness of fit. Unfortunately, CFA cannot confirm that this is the best fit from the infinity of possible models that might have been tried<sup>(30)</sup> and it cannot test the degree of model agreement. Consequently, the Bland Altman plots conducted in this study were used as an alternative and new method within assessment of reproducibility of identified dietary patterns. Bland Altman plots have to our knowledge not previously been used to assess the reproducibility of dietary patterns. Thus, we also estimated correlations as most studies assessing reproducibility of dietary patterns have used this method. We want to note, though, that the correlations measure the relation between two variables, not the agreement between them<sup>(25)</sup>. In principle, we do not favour this method in a setting comparing two methods.

We observed a high correlation between similar factors from PCA 1 and 2 ( $r$  0.93), but when the reproducibility was assessed by Bland Altman plots the result was less impressive (RV = 37–40%). Thus, data which seem to be in relatively poor agreement can produce relatively high correlations<sup>(25)</sup>. An explanation for some of the relatively large

variation between factors from PCA 1 and 2 may be that the number of food groups was greatly reduced in PCA 2, which affects estimates of factor loadings and hence factor scores. Others<sup>(9,12)</sup> using Pearson's correlation between factor scores on similar factors have concluded that no noteworthy differences in results from the explorative compared to the confirmative analyses were observed. The results from the Bland Altman plots blur the picture of our results because intuitively we do not find a RV of 37–40% reasonable. However, this degree of variation may be reasonable within nutritional epidemiology. We suggest that a variation of the size found in this study could also be identified by Togo *et al.*<sup>(9)</sup> and other studies examining dietary patterns. In summary, despite uncertainty regarding the Bland Altman plots, we find it reasonable to conclude that the two dietary patterns identified in sub-sample 1 were reproducible.

Next, we determined whether results from another sub-sample (2) conformed to what was expected from the findings in sub-sample 1. When data from both genders were analysed together the reproducibility assessed by Bland Altman plots was relatively poor (RV = 47%); possibly affected by sample size. Hence, we assume that the RV would have been lower if the sample had been larger. For women, the variation in the Bland Altman plot for each factor was acceptable (RV = 18–22%). For men, the variation in the plots was larger than for women (RV = 37–43%). As observed in other studies<sup>(9,14)</sup> we found a high correlation between similar factors identified in the two different sub-samples (all,  $r$  0.89; women,  $r \geq 0.98$ ; men,  $r \geq 0.90$ ). We and others<sup>(9)</sup> have further observed that the patterns identified between the two sub-samples were more homogeneous for women than men. Overall, our data indicate that the factors identified for the men were less reproducible. This finding may reflect that the dietary patterns are less structured between men than women. Women are generally more conscious about their dietary intake than men, which may generate less random variation.

#### General strengths and limitations of the study

Missing data or no intake with respect to a given food group for a great part of the population may be of concern working with PCA<sup>(23)</sup>. Missing data on the food groups was hence coded as an intake of 0 g/d. Furthermore, intake of each food group included in the PCA should exceed 0 g/d for > 50% of the population. Food groups unfulfilling this criterion were: mix of butter, margarine and oil, miscellaneous foods and tea. In- or exclusion of these food groups did not change the character of the final factors retained significantly (data not shown). Consequently, the food groups were included in the final analysis because exclusion would give less information.

We did not test whether food groups adjusted for energy derived different factors than food groups without adjustment. However, a previous study<sup>(31)</sup> has found that energy-adjusted food groups did not alter the patterns.

An important strength of this study was that we assessed the reproducibility of identified dietary patterns using a supplementary method: the Bland Altman plot. We suggest that Bland Altman plots are used in future studies assessing the reproducibility of identified dietary patterns because it questions findings from previously used methods.

#### Conclusion

Our study has described a 'Traditional' and 'Modern' pattern of Danes aged 30–60 years living in Copenhagen County. These patterns were in many ways similar to major patterns identified in other studies. Replication of the patterns within the same sub-sample was acceptable. Moreover, the same factor pattern was identified in two different sub-samples and the reproducibility of the patterns between the sub-samples was acceptable but better for women than for men. We introduced Bland Altman plots as an alternative method to assess the reproducibility of the dietary patterns identified and found that it questions results from previously used methods. We conclude that the identified patterns were robust as markers of food intake patterns on group level.

#### Acknowledgements

This study was supported by the Danish Medical Research Council, the Danish Centre for Evaluation and Health Technology Assessment, Novo Nordisk A/S, Copenhagen County, the Danish Heart Foundation, the Danish Diabetes Association, the Danish Pharmaceutical Association, the Augustinus Foundation, the Ib Henriksen Foundation, the Becket Foundation and the Danish Research Counselling System.

The authors thank all the participants who took part in the survey and the staff for their serious efforts that made this study possible.

K. B. J., T. J. and C. G. were responsible for the study design, data collection and obtaining funding. C. L. and B. C. were responsible for data analysis. C. L. was responsible for drafting the manuscript and interpretation of the results. C. G., I. T. and U. T. provided advice and consultation. All authors edited and reviewed the manuscript. The authors had no conflicts of interest.

#### References

1. Færch K, Lau C, Tetens I, Pedersen O, Jørgensen T, Borch-Johnsen K & Glümer C (2005) A statistical approach based on substitution of macronutrients provides additional information to models analyzing single dietary factors in relation to Type 2 Diabetes in Danish adults: the Inter99 study. *J Nutr* **135**, 1177–1182.
2. Osler M, Helms-Andreasen A, Heitmann B, Høidrup S, Gerdes U, Jørgensen LM & Schroll M (2002) Food intake patterns and risk of coronary heart disease: a prospective cohort study examining the use of traditional scoring techniques. *Eur J Clin Nutr* **56**, 568–574.
3. Hu FB, Rimm E, Smith-Warner SA, Feskanich D, Stampfer MJ, Ascherio A, Sampson L & Willett WC (1999) Reproducibility and validity of dietary patterns assessed with a food-frequency questionnaire. *Am J Clin Nutr* **69**, 243–249.
4. Montonen J, Knekt P, Härkänen T, Järvinen R, Heliövaara M, Aromaa A & Reunanen A (2005) Dietary patterns and the incidence of type 2 diabetes. *Am J Epidemiol* **161**, 219–227.
5. Engeset D, Alsaker E, Ciampi A & Lund E (2005) Dietary patterns and lifestyle factors in the Norwegian EPIC cohort: The Norwegian Women and Cancer (NOWAC) study. *Eur J Clin Nutr* **59**, 675–684.
6. Osler M, Heitmann BL, Gerdes LU, Jørgensen LM & Schroll M (2001) Dietary patterns and mortality in Danish men and women: a prospective observational study. *Br J Nutr* **85**, 219–225.

7. Fung TT, Willett WC, Stampfer MJ, Manson JE & Hu FB (2001) Dietary patterns and the risk of coronary heart disease in women. *Arch Intern Med* **161**, 1857–1862.
8. Bamia C, Orfanos P, Ferrari P, *et al.* (2005) Dietary patterns among older Europeans: the EPIC-Elderly study. *Br J Nutr* **94**, 100–113.
9. Togo P, Heitmann BL, Sørensen TI & Osler M (2003) Consistency of food intake factors by different dietary assessment methods and population groups. *Br J Nutr* **90**, 667–678.
10. Khani BR, Ye W, Terry P & Wolk A (2004) Reproducibility and validity of major dietary patterns among Swedish women assessed with a food-frequency questionnaire. *J Nutr* **134**, 1541–1545.
11. Maskarinec G, Novotny R & Tasaki K (2000) Dietary patterns are associated with body mass index in multiethnic women. *J Nutr* **130**, 3068–3072.
12. Newby PK, Weismayer C, Akesson A, Tucker KL & Wolk A (2006) Long-term stability of food patterns identified by use of factor analysis among Swedish women. *J Nutr* **136**, 626–633.
13. Newby PK, Muller D & Tucker KL (2004) Associations of empirically derived eating patterns with plasma lipid biomarkers: a comparison of factor and cluster analysis methods. *Am J Clin Nutr* **80**, 759–767.
14. Tseng M, Breslow RA, DeVellis RF & Ziegler RG (2004) Dietary patterns and prostate cancer risk in the National Health and Nutrition Examination Survey Epidemiological Follow-up Study cohort. *Cancer Epidemiol Biomarkers Prev* **13**, 71–77.
15. Glümer C, Jørgensen T & Borch-Johnsen K (2003) Prevalences of diabetes and impaired glucose regulation in a Danish Population: The Inter99 study. *Diabetes Care* **26**, 2335–2340.
16. Jørgensen T, Borch-Johnsen K, Thomsen TF, Ibsen H, Glümer C & Pisinger C (2003) A randomized non-pharmacological intervention study for prevention of ischaemic heart disease: baseline results Inter99 (1). *Eur J Cardiovasc Prevention Rehab* **10**, 377–386.
17. Lau C, Færch K, Glümer C, Toft U, Tetens I, Borch-Johnsen K & Jørgensen T (2004) Evaluation of dietary intake in a Danish population: the Inter99 study. *Scan J Nutr* **48**, 136–143.
18. Saxholt E, Møller A & Ygil KH, Danish Food Composition Databank Version 6.0 (Fødevaredatabanken Version 6.0). Available at <http://www.foodcomp.dk/> (accessed 29 June 2005) Søborg, Denmark: Department of Nutrition, The Danish Institute for Food and Veterinary Research.
19. Toft U, Kristoffersen L, Ladelund S, Bysted A, Jakobsen J, Lau C, Jørgensen T, Borch-Johnsen K & Ovesen L (2007) Relative validity of a food frequency questionnaire used in the Inter99 study. *Eur J Clin Nutr* 1–9 (Epublication 30 May 2007. DOI: 10.1038/SJ.EJCN.1602815).
20. Lyhne N, Christensen T, Groth M, *et al.* (2005) National Dietary Survey 2000–2002 (Danskernes Kostvaner 2000–2002. Hovedresultater). Søborg, Denmark: Department of Nutrition, The Danish Institute for Food and Veterinary Research.
21. The Danish Fitness and Nutrition Council, The Danish Dietary Recommendations 2005. Available at <http://www.meraadet.dk/> (accessed 4 July 2006).
22. R version 2.1.0, The R Foundation for Statistical Computing Version 2.1.0, Copyright 2005. Available at <http://www.r-project.org/> (accessed 18 April 2005).
23. Loehlin JC (2004) *Latent Variable Models: an introduction to factor, path, and structural analysis*, 4 ed., pp. 35–86–152–186, Mahwah, New Jersey, London: Lawrence Erlbaum Associates.
24. Hatcher L (1994) *A Step-by-Step Approach to Using SAS for Factor Analysis and Structural Equation Modelling*, pp. 12. Cary, NC, USA: SAS Institute Inc.
25. Bland JM & Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* **1**(8476), 307–310.
26. 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.
27. 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.
28. Newby PK, Muller D, Hallfrisch J, Qiao N, Andres RV & Tucker KL (2003) Dietary patterns and changes in body mass index and waist circumference in adults. *Am J Clin Nutr* **77**, 1417–1425.
29. Bollen KA (1989) *Structural Equation Modelling*, pp. 226–318. North Carolina: John Wiley & Sons.
30. Kline P (1994) *An Easy Guide to Factor Analysis*. New York: Routledge.
31. Balder HF, Virtanen M, Brants HA, *et al.* (2003) Common and country-specific dietary patterns in four European cohort studies. *J Nutr* **133**, 4246–4251.

**Appendix 1.** Food groups based on questions from FFQ, used in principal component analyses

Food group	Food items included from FFQ
Dairy products	Curdled milk products with and without fruit (0.1–3.5 % fat), whole milk, semi-skimmed milk and skimmed milk
Eggs	Eggs whole, egg white, omelette, gratin
Whole-grain cereals and bread	Rye bread, whole-grain bread, oats and muesli
Refined grain cereals and bread	Refined cereals and refined/white bread
Butter and lard	Butter and lard as spread and for preparation
Margarine	Margarine (40–80% vegetable fat) as spread and for preparation
Mix of butter, margarine and oil	Butter and oil mixed, margarine and oil mixed
Cheese	Hard and creamy cheeses (% fat in dry matter varies)
Paté or high-fat meat for sandwiches	Liver paste, paté, salami, pork sausages, roasted pork, hamburger steak and rissole
Low-fat meat for sandwiches	Ham, roast beef, smoked pork fillet, salted meat and liver ( $\leq 14\%$ fat)
Mayonnaise salads	Low and high-fat mayonnaise salads with meat or fish
High-fat fish	Mackerel, herring, tuna in oil, sardine, salmon, halibut and fish cake ( $> 7\%$ fat)
Low-fat fish	Tuna in tomato or water, canned cod roe, plaice, shrimps, shell fish, caviar, rainbow trout and garfish ( $\leq 7\%$ fat), and fish soup
Mixed vegetable dishes	Slices of tomato or potato, vegetable pie, vegetable beef, tomato soup, vegetable soup, vegetable sauce
Vegetables	Raw salad vegetables and cooked vegetables
Fruit	Fresh fruit, dried fruit, stewed fruit, fruit dessert
Red meat	Beef, pork, lamb, and dishes with these meats, minced or non-minced
Poultry	Poultry and dishes with poultry with or without skin
Potatoes	Potatoes; boiled, oven-baked, fried, mashed, cream baked and potato salad (except French fries and potato chips)
Pasta, rice and wheat kernels	Pasta, rice and wheat kernels
Fast food	Pre-made meals or fast food from grill bars, burger bars, pizzerias etc.
Salty snacks	Potato chips, popcorn, bacon chips and other salty snacks
Vegetable oil and vinegar dressing	Olive, rape seed, corn, grape, thistle or sunflower oil, and oil/vinegar dressings
Creamy dressing	Dressing based on cream, sour cream or curdled milk products
Sauces	Béarnaise sauce, hollandaise sauce, sauce with fat and pre-made sauces
Nuts	Walnuts, hazelnuts, almonds, cashew nuts, pine nuts, peanuts, pistachio nuts, peanut butter
Ice cream and chocolate milk	Cream based ice cream and chocolate milk
Confectioneries	Mixed sweets and chocolate
Sweet spread	Nutella, honey and marmalade
Cakes	Pastries, dry cakes, cream bun, biscuits and other cakes
Alcohol	Alcohol (grams of ethanol) from beer, wine and spirit
Miscellaneous foods	Soya milk
Coffee	Coffee
Tea	Tea

**Appendix 2.** Characteristics of the population

Characteristics	Sub-sample 1				Sub-sample 2				Total			
	<i>n</i>	%	Mean	SD	<i>n</i>	%	Mean	SD	<i>n</i>	%	Mean	SD
Men and women	3296				3267				6563			
Women	1688	51.2			1684	52.6			3372	51.4		
Age (years)			46.1	7.9			46.3	7.8			46.2	7.9
BMI* (kg/m <sup>2</sup> )			26.3	4.6			26.4	0.6			26.3	4.6
Physical activity during leisure												
Mainly sedentary	690	21			686	21			13	21		
Moderate activity	1985	62			1977	62			3962	62		
Regular exercise	509	16			498	16			1,007	16		
Heavy training	40	1			34	1			74	1		
Smoking												
Daily smoker	1144	35			1180	36			2324	36		
Occasional smoker	119	4			111	3			230	4		
Ex-smoker	809	24			843	26			1652	25		
Never-smoker	1198	37			1106	34			2304	35		
Educational level, years of vocational training												
None	504	16			529	18			1033	17		
Up to 1 year	145	5			132	4			277	4		
2–4 years	2010	66			1981	65			3991	66		
Over 4 years	388	13			385	13			773	13		

\* BMI was calculated from measured weight and height. Information on physical activity, smoking and education was obtained from a self-administered general questionnaire. Due to missing values for BMI (*n* 2), physical activity (*n* 144), smoking (*n* 53), and educational level (*n* 489) the numbers of participants vary in the Table.

**Appendix 3.** Dietary intake for the entire population  
(Medians and interquartile ranges (IQR))

Food group (g/d)	Sub-sample 1		Sub-sample 2		Total	
	Median	IQR	Median	IQR	Median	IQR
Dairy products	190.6	42.8–461.7	188.5	42.8–491.7	189.8	42.8–476
Eggs	12.3	4.9–24.3	12.4	4.9–24.7	12.3	4.9–24.5
Whole grain cereals and bread	140.0	87.5–218.8	140.0	87.5–210.2	140.0	87.5–215.5
Refined grain cereals and bread	61.5	31.1–122.2	62.3	32.2–121.0	61.7	31.5–121.7
Butter and lard	6.5	0.9–19.5	7.7	0.9–21.1	7.1	0.9–20.1
Margarine	1.3	0–8.3	1.5	0.0–8.5	1.4	0.0–8.4
Mix of butter, margarine and oil	0.0	0.0–3.7	0.0	0.0–3.4	0.0	0.0–3.5
Cheese	19.0	9.2–48.1	18.8	5.0–48.1	18.9	9.0–48.1
Paté or high-fat meat for sandwiches	12.6	4.9–26.3	12.4	5.0–25.6	12.6	5.0–26.0
Low-fat meat for sandwiches	5.8	2.0–13.2	5.5	1.7–12.2	5.6	1.8–12.7
Mayonnaise salads	0.9	0.0–6.1	0.9	0.0–6.1	0.9	0.0–6.1
High-fat fish	9.9	3.7–20.7	9.9	3.8–20.9	9.9	3.7–20.8
Low-fat fish	9.9	3.9–19.2	10.1	3.8–19.4	10.0	3.9–19.3
Mixed vegetable dishes	14.3	4.3–32.8	15.0	4.3–34	15.0	4.3–33.4
Vegetables	66.7	33.9–106.4	64.7	33.9–106.4	66.1	33.9–106.4
Fruit	119.6	37.5–314.7	116.9	40.5–312.5	118.8	38.4–313.1
Red meat	98.3	68.8–134.0	96.3	66.8–132.2	97.3	67.8–132.8
Poultry	14.7	6.8–27.2	14.3	6.7–26.2	14.6	6.7–26.8
Potatoes	127.4	79.9–207.3	126.8	80.0–205.6	127.1	80.0–206.0
Pasta, rice and wheat kernels	74.9	37.5–103.0	74.9	37.5–103.0	74.9	37.5–103.0
Fast food	24.3	10.0–43.7	23.8	10.0–43.7	24.2	10.0–43.7
Salty snacks	0.7	0.2–1.6	0.7	0.2–1.5	0.7	0.2–1.6
Vegetable oil and vinegar dressing	3.1	0.0–9.3	2.9	0.0–9.4	3.0	0.0–9.3
Creamy dressing	1.5	0.0–3.9	1.8	0.0–4.1	1.5	0.0–3.9
Sauces	11.7	5.8–23.1	11.7	5.8–22.0	11.7	5.8–23.1
Nuts	1.0	0.0–3.5	1.0	0.0–3.5	1.0	0.0–3.5
Ice cream and chocolate milk	4.9	2.0–12.6	4.9	2.0–12.6	4.9	2.0–12.6
Confectioneries	28.9	11.7–61.8	27.4	10.7–60.7	27.8	10.8–60.8
Sweet spread	2.6	0.1–8.6	2.6	0.0–9.1	2.6	0.0–9.0
Cakes	12.8	6.0–24.0	12.3	6.0–23.8	12.3	6.0–24.0
Alcohol	8.6	3.0–17.6	8.4	2.6–18.0	8.6	2.8–17.9
Miscellaneous foods	0	0–0	0.0	0.0–0.0	0.0	0.0–0.0
Coffee	800	400–1200	800	400–1200	800	400–1200
Tea	0	0–400	0	0–400	0	0–400