

Dietary patterns and their association with food and nutrient intake in the European Prospective Investigation into Cancer and Nutrition (EPIC)–Potsdam study

Matthias B. Schulze*, Kurt Hoffmann, Anja Kroke and Heiner Boeing

Department of Epidemiology, German Institute of Human Nutrition, Potsdam-Rehbruecke, Arthur-Scheunert-Allee 114-116, 14558 Bergholz-Rehbruecke, Germany

(Received 5 January 2000 – Revised 6 September 2000 – Accepted 21 September 2000)

Dietary pattern analysis has recently received growing attention, as it might be more appropriate in studies of diet–disease associations than the single food or nutrient approach that has dominated past epidemiological research. Factor analysis is a technique which is commonly used to identify dietary patterns within study populations. However, the ability of factor solutions to account for variance of food and nutrient intake has so far remained unclear. The present study therefore explored the statistical properties of dietary patterns with regard to the explained variance. Food intake of 8975 men and 13 379 women, assessed by a food-frequency questionnaire, was aggregated into forty-seven separate food groups. Dietary patterns were identified by principal component analysis and subsequent varimax rotation. Seven factors were retained for both men and women, which accounted for about 31 % of the total variance. The explained variance was relatively high (>40 %) for cooked vegetables, sauce, meat, dessert, cake, bread other than wholemeal, raw vegetables, processed meat, high-fat cheese, butter and margarine. Factor scores were used to investigate associations between the factors and nutrient intake. The patterns accounted for relatively large proportions of variance of energy and macronutrient intake, but for less variance of alcohol and micronutrient intake, especially of retinol, β -carotene, vitamin E, Ca and ascorbic acid. In addition, factors were related to age, BMI, physical activity, education, smoking and vitamin and mineral supplement use.

Dietary pattern: Factor analysis: Epidemiology: Food frequency

The intercorrelation of foods or nutrients often makes it impossible to examine separately their effects on disease risks. However, these intercorrelations are used in pattern analysis to characterise habitual food consumption. Dietary patterns might better reflect real-world conditions of food and nutrient intake (Sacks *et al.* 1995), and might have a greater effect on health than any single food item or nutrient (Appel *et al.* 1997). To identify dietary patterns within study populations, multivariate statistical techniques are applicable, with factor analysis being one technique predominantly used. Factor analysis characterises the covariance structure among variables in terms of a few underlying factors. Studies which applied factor analysis (Schwerin *et al.* 1981; Gex-Fabry *et al.* 1988; Barker *et al.* 1990; Randall *et al.* 1990; Wolff & Wolff, 1995; Whicelow & Prevost, 1996; Gittelsohn *et al.* 1998; Slattery *et al.* 1998; Hu *et al.* 1999; Williams *et al.* 2000) suggested that the extracted patterns

account for only a moderate part of the total variance of food intake. If patterns do not explain much of the total variance of the whole food intake, it is likely that they also do not explain much of the variance of single food and nutrient intake. This might dramatically restrict the usefulness of factor analysis in nutritional epidemiology. While Schwerin *et al.* (1981) reported estimates of the explained variance for single food intake which indicated large differences among foods, this issue was not addressed in other studies. In addition, no information is currently available with regard to the amount of variance of nutrient intake that is accounted for by patterns. The present paper, based on data from the German European Prospective Investigation into Cancer and Nutrition (EPIC) – Potsdam study, presents the relationship between dietary patterns and food and nutrient intake, with particular emphasis on the ability of patterns to explain variance.

Abbreviations: EPIC, European Prospective Investigation into Cancer and Nutrition.

* **Corresponding author:** Dr Matthias B. Schulze, fax +49 33200 88444, email mschulze@www.dife.de

Subjects and methods

Study population

The study population was selected from participants of the EPIC-Potsdam study (Boeing *et al.* 1999), which contributes a general population sample of 27 559 subjects to the EPIC multicentre cohort study (Riboli & Kaaks, 1997). A random sample of individuals meeting the age criteria, men aged 40–64 years and women 35–64 years, was provided by the registration offices of the selected municipalities in the Potsdam region. Individuals with missing information on dietary intake, smoking status, educational attainment, anthropometric measurements, and individuals reporting a change in their diet within the year before the assessment were excluded, leaving a total of 22 354 persons (8975 men and 13 379 women).

Data collection

Assessment of the study population was carried out between August 1994 and September 1998. Study participants filled out a self-administered food-frequency questionnaire and a lifestyle questionnaire at home, which were scanned by an optical reader at the beginning of the examination at the study centre. The examination consisted of a computer-guided interview, anthropometric measurements, and taking of blood samples.

The food-frequency questionnaire assessed the usual food and nutrient intake of individuals during the 12 months before the examination. Details on the validity and reproducibility of the questionnaire have been published previously (Boeing *et al.* 1997; Bohlscheid *et al.* 1997a,b; Kroke *et al.* 1999). The food-frequency questionnaire included 148 single food items and questions on specific dietary aspects, such as the fat content of dairy products and the types of fat used for food preparation. Photographs and, if available, standard portion sizes supported the estimation of portion sizes. The frequency of intake was measured using ten categories, ranging from 'never', 'one time per month or less' to 'five times per d or more'. The information on portion sizes and frequency of food intake was used to calculate the amount of each food item consumed on average per d. Daily nutrient intake was estimated from the consumed food items using the German Food Code (Federal Institute for Health Protection of Consumers and Veterinary Medicine, 1998). As nutrient intake is usually highly correlated with energy intake, we calculated energy-adjusted nutrient intakes also, using the regression residual method (Willett, 1998). Dietary patterns were explored with factor analysis, based on the food items of the food-frequency questionnaire, which were aggregated into forty-nine separate food groups (Table 1). The grouping scheme was generally based on culinary usage or nutrient profiles. However, grouping was also based on experiences of other studies (Slattery *et al.* 1998; Hu *et al.* 1999).

Smoking status, educational attainment, dietary changes during the previous year, physical activity, and vitamin and/or mineral supplement use were assessed through personal computer-guided interviews in the study centre. Smoking status was defined as current smoker and non-smoker. Educational attainment levels were defined as

vocational training or lower degree and trade school, technical school, or university degree. Users of vitamin or mineral supplements were defined as those individuals reporting a regular use within the 4 weeks before the examination. Physical activity level was determined as work metabolic rate: BMR. Energy costs of several leisure time activities and employed work, used to calculate the work metabolic rate, were based on Ainsworth *et al.* (1993). Estimation of BMR was based on Schofield *et al.* (1985). Frequency, duration and intensity of activities was assessed with the interview and the lifestyle questionnaire. All anthropometric measurements followed standardised procedures (Klipstein-Grobusch *et al.* 1997) and were performed with subjects wearing light underwear by trained and quality-monitored staff using instruments under permanent quality control. Body height was measured to the nearest 1 mm, body weight to the nearest 100 g. BMI was calculated as body weight (kg) divided by the square of body height (m²).

Statistical methods

All analyses were performed with the SAS System for Windows, release 6.12 (SAS Institute Inc., Cary, NC, USA), separately for men and women. Dietary patterns were identified with principal component analysis using the PROC FACTOR procedure. The factors are linear composites of the optimally-weighted observed variables. To identify the number of factors to be retained, we used the eigenvalue >1.0 criterion in a first step. This criterion is the most widely used in factor analysis and is based on the rationale that each factor retained should explain more variance than a single original variable in the data set. However, this procedure created sixteen independent factors for both men and women, a number too large for further analyses. A plot of the eigenvalues (scree test) indicated no clear break that could be used as a separate criterion to further reduce the number of factors. Thus, we used an eigenvalue >1.25 criterion in a second step, which is in agreement with Slattery *et al.* (1998), and finally retained seven factors for both men and women for further analyses. We allowed food items to load on more than one pattern simultaneously, which is in contrast with the interpretability criteria usually applied in factor analysis (Hatcher, 1994). After a varimax rotation of the factors, food items with absolute factor loadings >0.2 were considered as significantly contributing to a pattern. We excluded low-energy and high-energy soft drinks from the final analysis, because they did not load on any factor retained. The other food groups loaded high at least at one factor. Factor loadings can be interpreted as correlation coefficients between food items and patterns (Hatcher, 1994). The sum of the squares of a variable's factor loadings over all retained factors (communality) represents the variance of the variable that is explained by the final factors. Factor scores, i.e. the individual values of the factors, were saved for all seven factors for each study participant. They were computed by weighting each factor loading by the factor's eigenvalue, multiplying these weights with the subject's corresponding standardised food-group intake, and summing these products. The scores

Table 1. Food groupings used in the dietary pattern analysis

Foods or food groups	Food items
Cooked vegetables	Tomatoes, tomato sauce, sweet pepper, courgette, aubergine, spinach, carrots, asparagus, pea-carrot vegetable mix, leek, celery (all cooked)
Cabbage family	Broccoli, cauliflower, red and white cabbage, kohlrabi (all cooked)
Legumes	Green peas, green beans, pea-bean-lentil stew
Cooked potatoes	Salted potatoes, jacket potatoes, mashed potatoes, potato salad, dumplings
Mushrooms	Fresh mushrooms, mushroom dishes
Sauce	Ketchup, brown and white sauce, salad dressing, sauce for vegetables
Poultry	Fried, grilled or roasted chicken or turkey
Meat except fish and poultry	Pork, beef, hamburger, minced meat, liver, lamb, roast hare
Animal fat except butter	Animal fat used for food preparation
Dessert	Pudding, sweet soufflé
Cake, cookies	Cake, tart, cookies
Confectionery, ice cream	Chocolate, candy bars, pralines, sugar, ice-cream
Jam, honey, chocolate spread	Jam, honey, chocolate spread, peanut butter
Canned fruit	Canned fruit
Fruit juice	Citrus, apple, orange, grapefruit, grape, cherry, pineapple juice, multi-vitamin drinks
Tea	Black tea, green tea, fruit and herbal teas
Muesli	Whole-grain breakfast cereal, muesli
Cornflakes	Cornflakes, other refined grain-based breakfast cereal
Pasta, rice	Cooked pasta, cooked rice
Pizza	Pizza, quiche
Vegetarian dishes	Vegetarian dishes
Garlic	Raw, fried or cooked garlic
Wholemeal bread	Wholemeal bread, dark and wholemeal rolls
Other bread	Rye bread, wheat bread, mixed bread, pale rolls, crispbread, croissants
Olive oil	Olive oil used for food preparation
Fresh fruit	Apple, pear, peach, cherry, grape, strawberry, blackberry, raspberry, kiwi, pineapple, mango, banana
Raw vegetables	Cucumber, carrot, sprouts, paprika, tomato, onion, radish
Other vegetable oils and fats	Vegetable fat used for food preparation (frying, dressing etc.)
Water	Tap water, mineral water
Fish	Fish, canned fish, smoked fish
Nuts	Nuts
Chips, salt sticks	Chips, salt sticks, cracker
Fried potatoes	French fries, potato fritters, fried potatoes
Beer	Beer
Spirits	Spirits
Wine	Wine, fruit wine, champagne
Other alcoholic beverages	Dessert wine, liqueur, aperitif
Eggs	Boiled eggs, fried eggs, omelette
Coffee	Coffee
Soup	Vegetable or potato stew, vegetable soup, meat or fish soup, broth, thickened soup
Processed meat	Salami, cold-cut sausage, ham, fried sausage
Low-fat dairy products	Milk or yoghurt (≤ 1.5 % fat), soured milk, low-fat curd cheese
High-fat dairy products	Other milk or yoghurt, curd cheese, cream
Low-fat cheese	Low-fat cheese
High-fat cheese	Other cheese
Butter	Butter as spread and for food preparation
Margarine	Margarine as spread and for food preparation

represent standardised variables with mean = 0 and standard deviation = 1. They were used to study associations between the factors and nutrient intake. Pearson's correlation coefficients were calculated between factors and absolute as well as energy-adjusted nutrient intake. As the principal component analysis and the subsequent orthogonal rotation of all factors guarantee that all factors are uncorrelated, the sum of the squared correlations between absolute nutrient intake and factor scores can be interpreted as the proportion of the variance of nutrient intake that is explained by the factors. Pattern associations with age, anthropometric and lifestyle variables were assessed using multiple linear regression with a stepwise variable selection procedure. The regression analysis was restricted to 7644 men and 11 900 women because of missing data for supplement use.

Results

The factor analysis identified seven major factors according to the eigenvalue >1.25 criterion applied. These factors accounted for 30.9 % of the variance in food intake in men and 30.8 % in women. The factor loading matrices are presented in Tables 2 and 3. High loadings indicate strong associations between the corresponding observed variables and patterns. While many food groups have only one high loading and can therefore be assigned to a specific pattern, quite a few food groups loaded on different patterns simultaneously, e.g. tea, wholemeal bread, other bread, beer, wine, soup, processed meat and margarine; some food groups even had comparable loadings. The patterns were labelled according to the food groups with high loadings. In general, the patterns separately derived for men and women

Table 2. Factor loading matrix for men in the European Prospective Investigation into Cancer and Nutrition (EPIC)–Potsdam study (*n* 8975)*†

	Dietary pattern							Percentage of variance explained
	Plain cooking	Sweets	Cereals	Fruit and vegetables	Alcohol	High-fat dairy	Bread and sausage	
Cooked vegetables	0.66							49.9
Cabbage family	0.52				−0.22			34.0
Legumes	0.38					−0.22		24.3
Cooked potatoes	0.55		−0.25					39.4
Mushrooms	0.38							18.7
Sauce	0.67							49.6
Poultry	0.52							32.3
Meat except fish and poultry	0.74							61.2
Animal fat except butter	0.23							7.4
Dessert		0.60						41.0
Cake, cookies		0.62						41.8
Confectionery, ice cream		0.48			0.31			39.0
Jam, honey, chocolate spread		0.47						25.3
Canned fruit		0.51						35.7
Fruit juice		0.23						8.8
Tea		0.21	0.23	0.21				15.0
Muesli			0.39					22.8
Cornflakes			0.27					9.7
Pasta, rice			0.55					34.8
Pizza			0.57					38.0
Vegetarian dishes			0.47					24.3
Garlic			0.25		0.21			16.6
Wholemeal bread			0.36	0.25		−0.20		27.9
Other bread			−0.32			0.25	0.33	53.9
Olive oil			0.29	0.46				33.1
Fresh fruit				0.61				41.3
Raw vegetables				0.75				57.2
Other vegetable oils and fats				0.53				31.3
Water				0.26				13.9
Fish				0.27	0.33			20.6
Nuts					0.37			17.8
Chips, salt sticks					0.38			18.4
Fried potatoes	0.24				0.36			20.9
Beer		−0.38	−0.23		0.37			35.5
Spirits		−0.26	0.31		0.34			21.2
Wine			0.34		0.34			27.3
Other alcoholic beverages					0.41			17.1
Eggs					0.45			26.4
Coffee					0.24			11.2
Soup	0.21	0.21			0.31			25.5
Processed meat			−0.25		0.29		0.53	43.6
Low-fat dairy products						−0.44		21.5
High-fat dairy products		0.29				0.29		21.7
Low-fat cheese						−0.58		37.0
High-fat cheese						0.52	0.30	40.0
Butter						0.59	−0.33	54.8
Margarine			−0.22			−0.36	0.69	61.0
Percentage of variability	7.2	5.4	4.6	3.9	3.6	3.2	3.0	Σ = 30.9

* Factor loadings <0.21 are not shown.

† For details of procedures, see p. 364.

were very similar. Thus, some patterns were identically labelled for both sexes ('plain cooking', 'sweets', 'fruit and vegetables' and 'alcohol'), although their order of importance and the corresponding food groups were somewhat different. The communalities, representing the variability in food intake explained by the patterns, were relatively high (>40 %) for cooked vegetables, sauce, meat, dessert, cake, bread other than wholemeal, raw vegetables, processed meat, high-fat cheese, butter and margarine (Tables 2 and 3). Relatively low communalities (<20 %) were observed for mushrooms, animal fat, garlic, nuts, chips, fruit juice, cornflakes, tea, water, coffee and alcoholic beverages other than beer, wine and spirits.

The retained patterns accounted for a relatively large proportion of variance (>50 %) of nutrient intake with the exception of retinol, β -carotene, vitamin E, and Ca in both men and women, and alcohol in men and ascorbic acid in women (Tables 4 and 5). Most patterns were positively correlated with total energy intake; however, the 'cereals' pattern in men was nearly uncorrelated and the 'low-fat dairy' pattern in women was negatively correlated. Patterns positively associated with total energy consequently showed positive associations with the absolute intake of the energy-providing nutrients protein, carbohydrates and fat. However, as Tables 4 and 5 show, the correlation coefficients between pattern scores and the energy-adjusted

Table 3. Factor loading matrix for women in the European Prospective Investigation into Cancer and Nutrition (EPIC)–Potsdam study (*n* 13 379)*†

	Dietary pattern							Percentage of variance explained
	Plain cooking	Bread and sausage	Sweets	Fruit and vegetables	Low-fat dairy	Alcohol	Cereals and meat	
Cooked vegetables	0.64							47.7
Cabbage family	0.48							27.9
Legumes	0.52							27.8
Cooked potatoes	0.61							42.7
Mushrooms	0.35							21.3
Sauce	0.54						0.26	43.2
Poultry	0.49						0.26	32.2
Meat except fish and poultry	0.66	0.27					0.21	55.9
Animal fat except butter	0.22							5.6
Dessert			0.59					37.9
Cake, cookies			0.63					41.4
Confectionery, ice cream			0.59					37.9
Jam, honey, chocolate spread			0.42				−0.27	27.2
Canned fruit	0.23		0.45					30.4
Fruit juice			0.25					9.4
Tea		−0.24		0.23				16.0
Muesli		−0.43						23.6
Cornflakes			0.22					8.6
Pasta, rice		−0.32					0.49	39.0
Pizza		−0.25					0.56	40.1
Vegetarian dishes		−0.45						25.9
Garlic		−0.28						13.3
Wholemeal bread		−0.43						26.0
Other bread		0.62			−0.24			47.1
Olive oil				0.43				33.3
Fresh fruit				0.63				43.9
Raw vegetables				0.80				64.4
Other vegetable oils and fats				0.59				37.7
Water		−0.22		0.20	0.23			17.2
Fish	0.20		0.21			0.22		20.1
Nuts			0.25			0.20		12.1
Chips, salt sticks						0.21	0.24	15.1
Fried potatoes	0.28						0.28	24.8
Beer						0.33		16.7
Spirits						0.73		20.0
Wine						0.41		57.6
Other alcoholic beverages						0.61		38.0
Eggs						0.24		13.9
Coffee						0.21		10.3
Soup	0.32		0.21					17.8
Processed meat		0.52					0.22	39.5
Low-fat dairy products					0.49			28.1
High-fat dairy products			0.28		−0.32			19.9
Low-fat cheese					0.62			41.9
High-fat cheese				0.26	−0.49		0.27	42.7
Butter					−0.60		−0.32	52.4
Margarine		0.41			0.37		0.41	50.7
Percentage of variability	7.2	5.6	4.4	4.0	3.5	3.3	2.8	Σ = 30.8

* Factor loadings <0.21 are not shown.

† For details of procedures, see p. 364.

nutrient intakes differed in many cases to those for the absolute nutrient intake. For example, after controlling for total energy, the ‘sweets’ pattern showed only a strong positive association with carbohydrate intake, but not with fat and protein intake. The ‘alcohol’ pattern was no longer positively correlated with the intake of protein, carbohydrates and fat, but strongly with alcohol. Furthermore, relatively strong positive correlations (>0.4) were observed in both men and women between the ‘fruit and vegetables’ pattern and fibre, Fe and ascorbic acid intake; in men between the ‘bread and sausage’ pattern and salt; in women between the ‘low-fat dairy’ pattern and polyunsaturated:

saturated fatty acids. Strong negative correlations were observed in men between the ‘sweets’ pattern and alcohol intake, and in men between the ‘high-fat dairy’ pattern and polyunsaturated:saturated fatty acids.

Multiple linear regression analysis revealed that pattern scores were differently associated with age, BMI, physical activity level, smoking, educational attainment, and supplement use (Tables 6 and 7). Relative strong positive associations were found for age and ‘sweets’ in men and ‘plain cooking’ patterns in women, while in men ‘cereals’, ‘alcohol’, ‘high-fat dairy’, and ‘bread and sausage’ patterns and in women ‘cereals and meat’ pattern showed negative

Table 4. Pearson correlation coefficients* between dietary pattern score and total energy, absolute and energy-adjusted nutrient intake, and percentage of explained variance of energy and nutrient intake in men in the European Prospective Investigation into Cancer and Nutrition (EPIC)–Potsdam study (*n* 8975)†

	Dietary pattern														Percentage of variance explained
	Plain cooking		Sweets		Cereals		Fruit and vegetables		Alcohol		High-fat dairy		Bread and sausage		
	Absolute	Energy-adjusted	Absolute	Energy-adjusted	Absolute	Energy-adjusted	Absolute	Energy-adjusted	Absolute	Energy-adjusted	Absolute	Energy-adjusted	Absolute	Energy-adjusted	
Total energy (KJ/d)	0.26		0.44		-0.01		0.23		0.49		0.23		0.38		75.2
Protein (g/d)	0.37	0.28	0.34	-0.08	0.03	0.06	0.27	0.15	0.40	-0.06	0.09	-0.22	0.43	0.20	67.9
Carbohydrates (g/d)	0.13	-0.19	0.64	0.52	0.04	0.08	0.21	0.02	0.27	-0.32	0.13	-0.13	0.30	-0.07	65.2
Fat (g/d)	0.27	0.09	0.44	0.11	-0.01	-0.01	0.17	-0.07	0.42	-0.05	0.30	0.22	0.42	0.19	73.8
P:S	-0.03	-0.03	-0.22	-0.22	0.06	0.06	0.13	0.13	-0.05	-0.05	-0.62	-0.62	0.32	0.32	55.9
Cholesterol (mg/d)	0.33	0.20	0.39	0.06	-0.07	-0.12	0.10	-0.14	0.50	0.18	0.35	0.28	0.26	-0.07	71.6
Fibre (g/d)	0.17	0.01	0.44	0.21	0.05	0.06	0.48	0.43	0.06	-0.32	-0.02	-0.21	0.26	0.03	52.7
Alcohol (g/d)	0.08	0.01	-0.42	-0.58	-0.11	-0.11	0.04	-0.04	0.52	0.38	0.10	0.03	-0.05	-0.18	42.2
Retinol (mg/d)	0.28	0.17	0.16	-0.09	-0.12	-0.14	0.07	-0.07	0.25	-0.02	0.24	0.13	0.31	0.13	34.0
β-Carotene (mg/d)	0.11	0.05	0.22	0.12	0.05	0.05	0.34	0.29	0.03	-0.08	-0.07	-0.13	0.04	-0.05	18.6
Vitamin E mg/d	0.13	-0.01	0.33	0.12	0.09	0.10	0.28	0.19	0.19	-0.07	-0.11	-0.26	0.33	0.16	39.1
Ascorbic acid (mg/d)	0.19	0.11	0.31	0.17	0.03	0.03	0.50	0.45	0.04	-0.14	-0.06	-0.15	0.05	-0.09	67.9
Ca (g/d)	0.09	-0.10	0.40	0.15	0.19	0.25	0.28	0.17	0.19	-0.16	0.02	-0.16	0.16	-0.11	34.5
Fe (mg/d)	0.32	0.18	0.39	0.06	0.10	0.16	0.40	0.37	0.33	-0.10	0.06	-0.21	0.37	0.11	66.8
Table salt (g/d)	0.23	0.04	0.22	-0.21	-0.04	-0.06	0.28	0.16	0.35	-0.06	0.11	-0.11	0.57	0.43	63.0

P:S, polyunsaturated:saturated fatty acids.

* Correlation coefficients >0.02 are significantly different from 0 (*P* = 0.05).

† For details of procedures, see p. 364.

Table 5. Pearson correlation coefficients* between dietary pattern score and total energy, absolute and energy-adjusted nutrient intake, and percentage of explained variance of energy and nutrient intake in women in the European Prospective Investigation into Cancer and Nutrition (EPIC)–Potsdam study (n 13 379)†

	Dietary pattern														Percentage of explained variance
	Plain cooking		Bread and sausage		Sweets		Fruit and vegetables		Low-fat dairy		Alcohol		Cereals and meat		
	Absolute	Energy-adjusted	Absolute	Energy-adjusted	Absolute	Energy-adjusted	Absolute	Energy-adjusted	Absolute	Energy-adjusted	Absolute	Energy-adjusted	Absolute	Energy-adjusted	
Total energy (KJ/d)	0.24		0.19		0.72		0.30		-0.15		0.18		0.20		79.7
Protein (g/d)	0.34	0.27	0.20	0.08	0.52	-0.20	0.32	0.14	0.04	0.34	0.12	-0.06	0.23	0.11	59.7
Carbohydrates (g/d)	0.15	-0.16	0.09	-0.19	0.77	0.30	0.24	-0.05	-0.08	0.12	0.01	-0.34	0.07	-0.27	69.3
Fat (g/d)	0.25	0.08	0.28	0.27	0.61	-0.08	0.26	-0.02	-0.25	-0.27	0.13	-0.08	0.26	0.19	72.8
P:S	-0.06	-0.06	0.07	0.07	-0.20	-0.20	0.22	0.22	0.58	0.58	-0.07	-0.07	0.23	0.23	49.1
Cholesterol (mg/d)	0.34	0.25	0.26	0.19	0.56	0.01	0.08	-0.23	-0.26	-0.23	0.20	0.10	0.19	0.06	64.7
Fibre (g/d)	0.22	0.08	-0.01	-0.17	0.45	-0.05	0.53	0.45	-0.01	0.13	-0.05	-0.23	0.01	-0.17	53.5
Alcohol (g/d)	0.04	-0.01	-0.07	-0.11	-0.11	-0.25	0.04	-0.01	-0.08	-0.06	0.75	0.73	0.10	0.06	59.9
Retinol (mg/d)	0.26	0.16	0.33	0.27	0.29	-0.13	0.11	-0.06	-0.18	-0.11	0.09	-0.01	0.18	0.08	34.6
β-Carotene (mg/d)	0.10	0.03	-0.01	-0.07	0.22	0.01	0.39	0.31	0.02	0.06	-0.01	-0.06	-0.01	-0.07	21.1
Vitamin E (mg/d)	0.10	-0.04	0.12	0.02	0.40	0.02	0.35	0.22	0.05	0.16	0.02	-0.09	0.20	0.10	35.0
Ascorbic acid (mg/d)	0.17	0.07	0.02	-0.07	0.30	-0.01	0.55	0.47	0.03	0.10	0.01	-0.08	-0.03	-0.13	42.4
Ca (g/d)	0.07	-0.12	-0.07	-0.27	0.47	-0.01	0.33	0.18	0.09	0.25	0.06	-0.08	0.04	-0.12	35.3
Fe (mg/d)	0.31	0.20	-0.01	-0.23	0.50	-0.05	0.46	0.37	0.02	0.20	0.12	-0.01	0.15	0.01	59.5
Table salt (g/d)	0.24	0.09	0.34	0.31	0.38	-0.27	0.34	0.17	-0.01	0.17	0.11	-0.04	0.31	0.25	54.2

P:S, polyunsaturated:saturated fatty acids.

* Correlation coefficients >0.02 are significantly different from 0 (P = 0.05).

† For details of procedures, see p. 364.

Table 6. Regression parameters* (β); men in the European Prospective Investigation into Cancer and Nutrition (EPIC)–Potsdam study (n 7644)†

	Dietary pattern						
	Plain cooking	Sweets	Cereals	Fruit and vegetables	Alcohol	High-fat dairy products	Bread and sausage
Age (y)	0.007	0.014	-0.032	0.006	-0.020	-0.015	-0.030
BMI (kg/m ²)	0.012	-0.045	-0.019	0.012	0.023	-0.037	0.008
PAL	-0.103	0.138	NS‡	0.182	0.080	0.126	0.099
Education	-0.090	-0.058	0.349	0.064	NS	0.083	-0.168
Current smoking	0.114	-0.093	-0.264	-0.188	0.314	0.218	-0.045
Supplement use	NS	NS	0.273	0.184	NS	NS	NS

PAL, physical activity level (work metabolic rate:BMR).

* Parameter estimates of multiple linear regression analysis for the dependent variable 'pattern score' with stepwise selection of the independent variables age, BMI, PAL, education ('vocational training or lower degree' or 'trade school, technical school, or university degree'), current smoking, and supplement use (regular use of mineral or vitamin supplements within previous 4 weeks).

† For details of procedures, see p. 364.

‡ Not significantly different from 0 in the regression model.

associations. BMI showed positive associations, particularly with 'alcohol' in men and 'plain cooking', 'bread and sausage' and 'low-fat dairy' patterns in women. In contrast, 'sweets', 'cereals' and 'high-fat dairy' patterns in men were negatively associated with BMI. Physical activity level was positively associated with the 'fruit and vegetables' pattern in men and negatively with the 'bread and sausage' pattern in women; other patterns were only weakly associated. The 'cereals' pattern in men was positively associated with the educational attainment level, while the opposite was found for the 'bread and sausage' pattern in women. Smoking showed a strong positive association, particularly with the 'alcohol' pattern in both men and women. Supplement use was not related to the scores for many patterns. Only 'cereals' and 'fruit and vegetables' patterns in men showed relatively strong associations.

Discussion

This analysis explored dietary patterns in a large population-based cohort in relation to nutrient intake, age, anthropometry and lifestyle characteristics. Of seven patterns identified in both men and women, four patterns were relatively consistent across sexes. Here, food-group loadings and correlations with nutrient intake were similar. That patterns might vary between sexes has been suggested by previous studies, where women tended to have generally

higher loadings on healthfulness patterns (Gex-Fabry *et al.* 1988; Barker *et al.* 1990; Williams *et al.* 2000). Food choices may be determined by gender-related health, pleasure and convenience attributes and cognition (Rappoport *et al.* 1993), with healthy meals appealing more to women than to men. However, it is not possible to investigate whether there are major differences between men and women within the four relatively similar patterns in our study, as the separated extraction of patterns does not allow a direct comparison.

Our observation that different dietary patterns are differently related to age and anthropometric variables is consistent with earlier studies (Gex-Fabry *et al.* 1988; Barker *et al.* 1990; Slattery *et al.* 1998; Williams *et al.* 2000). Furthermore, the hypothesis that patterns might be governed by a number of socio-demographic factors, including socio-economic status, marital status and religion, has been suggested by Barker *et al.* (1990). We have observed that patterns were differently related to educational attainment level, where, for example, individuals with higher scores for the 'fruit and vegetables' pattern tended to be more educated than those scoring lower. Furthermore, the finding that the 'fruit and vegetable' pattern was positively associated with physical activity but negatively associated with smoking is in agreement with relationships of patterns strongly associated with fruit and vegetable intake observed in other studies (Slattery *et al.*

Table 7. Regression parameters* (β); women in the European Prospective Investigation into Cancer and Nutrition (EPIC)–Potsdam study (n 11 900)†

	Dietary pattern						
	Plain cooking	Bread and sausage	Sweets	Fruit and vegetables	Low-fat dairy products	Alcohol	Cereals and meat
Age (y)	0.015	0.002	-0.003	NS‡	0.005	-0.004	-0.046
BMI (kg/m ²)	0.021	0.017	-0.007	0.004	0.042	0.003	0.013
PAL	-0.060	-0.267	0.160	0.140	NS	NS	-0.156
Education	-0.047	-0.254	-0.042	0.146	-0.059	NS	0.076
Current smoking	0.079	0.184	-0.121	-0.225	-0.035	0.321	NS
Supplement use	NS	NS	NS	0.087	0.123	0.034	0.042

PAL, physical activity level (work metabolic rate:BMR).

* Parameter estimates of multiple linear regression analysis for the dependent variable 'pattern score' with stepwise selection of the independent variables age, BMI, PAL, education ('vocational training or lower degree' or 'trade school, technical school, or university degree'), current smoking, and supplement use (regular use of mineral or vitamin supplements within previous 4 weeks).

† For details of procedures, see p. 364.

‡ Not significantly different from 0 in the regression model.

1998; Williams *et al.* 2000). In addition, patterns associated with alcohol intake have been previously reported to be linked to cigarette smoking (Slattery *et al.* 1998), an observation also made in our study. That dietary patterns might interact with other lifestyle behaviours or might rather be part of broader lifestyles than specific dietary patterns has been argued previously (Barker *et al.* 1990; Martinez *et al.* 1998; Williams *et al.* 2000). While these interrelationships might increase the confidence that the statistical solution of factor analysis is in fact meaningful (Slattery *et al.* 1998), it might not be possible to separate the effect of dietary patterns from the effect of other lifestyle behaviours in risk assessments, as has been shown recently by Williams *et al.* (2000).

The present study, in a German population, identified patterns which are to some extent similar to those found in previous studies. For example, meat and vegetables loaded on our 'plain cooking' pattern and the 'meat and two vegetables' pattern in the study of Barker *et al.* (1990) as well as the pattern '2' in the study of Whichelow & Prevost (1996). However, there are quite important differences. Barker *et al.* (1990) reported high negative loadings within the 'meat and two vegetables' pattern for nuts, cider, patés, cheese and confectionery, with nuts having the highest absolute loading of all food groups. Whichelow & Prevost (1996) reported high loadings for confectionery, biscuits, cake and pudding within the pattern '2'. This is in contrast to our 'plain cooking' pattern. Since patterns should be judged as comparable only if the food groups significantly contributing are similar and their factor loadings are furthermore comparable in their magnitude, we conclude that the patterns extracted in our study are quite different from those retained from British (Barker *et al.* 1990; Whichelow & Prevost, 1996; Williams *et al.* 2000), Swiss (Gex-Fabry *et al.* 1988) or US (Schwerin *et al.* 1981; Randall *et al.* 1990; Wolff & Wolff, 1995; Gittelsohn *et al.* 1998; Slattery *et al.* 1998; Hu *et al.* 1999) study populations. This finding is not surprising, because patterns are retained in an explorative way rather than established *a priori*, and are therefore unlikely to be reproducible in populations with different dietary habits. However, it emphasises the possible inability to reproduce risk estimates, which are estimated for dietary patterns retained with exploratory approaches, across populations.

Randall *et al.* (1990) suggested that a link between patterns and disease risk is most probably identifiable among those patterns contributing most to the variance in dietary intake. In our study, the seven strongest patterns accounted for only about 31 % of the total variance, which is in agreement to some previous studies (Gex-Fabry *et al.* 1988; Barker *et al.* 1990; Randall *et al.* 1990; Whichelow & Prevost, 1996; Slattery *et al.* 1998; Hu *et al.* 1999). It contrasts the results of Schwerin *et al.* (1981), who reported a relatively high percentage of explained variance (55.3 %). However, their study was based on fifteen food groups only, from which seven factors were retained. The factor solution of our study, using forty-seven original food groups, represents a much higher variable reduction. Since the explained variance depends largely on the degree to which the variables are reduced, it cannot be concluded that our study is less precise than the factor solution of the study

by Schwerin *et al.* (1981). Moreover, a similar reduction (from forty-seven food groups to twenty-three factors) would have led to about 65 % explained total variance in our study. Communalities observed in the study by Schwerin *et al.* (1981) varied between single food groups, despite their relative high values ranging from 0.30 to 0.77. This finding supports our finding that patterns quite differently account for variances of single food-group intake.

Low explained variance of food intake might further lead to inappropriately measured nutrient intake. For example, only moderate communalities were found for alcoholic beverages, vegetable oils and dairy products and cheese. This finding might explain why patterns accounted for only small percentages of variance of alcohol, vitamin E and Ca intake. However, such a link between variance of food intake and variance of nutrient intake will only be present if the aggregated food-group variable sufficiently discriminates individuals with regard to the nutrient (Willett, 1998). This situation might not be the case in our study for the food group meat and retinol, even though the food group contains liver as the food item believed to contribute most to the inter-individual variation of retinol intake. The variance of meat intake was relatively well explained, but the variance of retinol intake was not. In a re-analysis of our data, now using liver as a single food group instead of including it in the meat food group, the amount of explained variance of retinol intake did not change (data not shown). Liver loaded only moderately at the 'plain cooking' pattern, resulting in a low communality estimate (<0.2). This finding indicates that even if the food item is seen to discriminate individuals with regard to nutrient intake, the explained variance of nutrient intake still depends on how well the variance of food intake is explained by the extracted patterns. Whether a more nutrient-oriented food grouping with isolated key food items might be a strategy to improve the explained variance of nutrients has not yet been investigated in detail. None of the previous studies using factor analysis addressed the issue of explained variance. Randall *et al.* (1990) and Hu *et al.* (1999) reported correlation coefficients between factor scores and nutrient intake. However, estimates of the explained variance were not reported and cannot be estimated from the reported correlation coefficients.

As a consequence of our observation that factors might explain nutrient intake quite differently, the usefulness of factor analysis might be limited. In cases where specific nutrients are of interest, but are not well represented by the extracted patterns, factor analysis might be disadvantageous compared with more specific methods which allow selection of those single food items explaining the largest part of the variance of specific nutrients, such as Max_r (Mark *et al.* 1996; Thomas & Mark, 1997). Max_r allows selection of a subset of food items for which the unweighted sum of the intake of a specific nutrient shows the highest correlation with the nutrient intake estimated over all food items. This method has been shown to be effective even for total energy and nutrients contained in a wide range of foods, such as protein (Mark *et al.* 1996). However, the seven factors in our study explained a large percentage of variance of total energy as well as protein,

carbohydrates and fat. It is unlikely that a subset of food items selected with Max_r, comparable in their number with the extracted patterns of a factor analysis, account simultaneously for a comparable percentage of variance of these nutrients. If all these nutrients are of primary interest, factor analysis might therefore be an alternative. However, this premise will only be true for the whole factor solution, but not for single factors, as single factors seem to contribute differently to the explained variance of nutrients. The use of single factors in risk assessments instead of the whole factor solution might therefore be an inappropriate strategy.

As the study by Slattery *et al.* (1998) suggested, dietary patterns describe disease associations beyond that described by single food items. Dietary patterns allow examination not only of single nutrients and foods but also of the effect of the overall diet. Patterns can furthermore be used as covariates when examining a specific nutrient, in order to establish whether the nutrient-related effect is independent of overall dietary patterns (Hu *et al.* 1999). Factor analysis might not, however, be appropriate in situations where specific foods or nutrients of interest are not well represented by the patterns extracted. The arbitrary nature of factor analysis as well as the inability to reproduce results across different populations are further pitfalls of this approach, and might restrict its usefulness in investigations of diet-disease relationships in epidemiological studies.

Acknowledgements

We wish to thank all study participants for their cooperation and all interviewers, programmers and documentarists who devoted their energy to collect and process all the data. The recruitment phase of the EPIC-Potsdam study was mainly supported by the Federal Ministry of Science, Germany (grant no. 01 EA 9401). Further financial support was given by the 'Europe against Cancer' programme of the European Community (grant no. SOC 95 201408 05F02). The EPIC-Potsdam study is now supported by the Deutsche Krebshilfe (grant no. 70-2488-Ha I) and the European Community (grant no. SOC 98 200769 05F02). M. B. Schulze was furthermore financially supported by the German Research Foundation (grant no. BO 807/6-1).

References

- Ainsworth BE, Haskell WL, Leon AS, Jacobs DR Jr, Montoye HJ, Sallis JF & Paffenbarger RS Jr (1993) Compendium of physical activities: classification of energy costs of human physical activities. *Medicine and Science in Sports and Exercise* **25**, 71–80.
- Appel LJ, Moore TJ, Obarzanek E, Vollmer WM, Svetkey LP, Sacks FM, Bray GA, Vogt TM, Cutler JA, Windhauser MM, Lin PH & Karanja N (1997) A clinical trial of the effects of dietary patterns on blood pressure. *New England Journal of Medicine* **336**, 1117–1124.
- Barker ME, McClean SI, Thompson KA & Reid NG (1990) Dietary behaviours and sociocultural demographics in Northern Ireland. *British Journal of Nutrition* **64**, 319–329.
- Boeing H, Bohlscheid-Thomas S, Voss S, Schneeweiß S & Wahrendorf J (1997) The relative validity of vitamin intakes derived from a food frequency questionnaire compared to 24-hour recalls and biological measurements: results from the EPIC pilot study in Germany. *International Journal of Epidemiology* **26**, Suppl. 1, S82–S90.
- Boeing H, Wahrendorf J & Becker N (1999) EPIC-Germany – a source for studies into diet and risk of chronic diseases. *Annals of Nutrition and Metabolism* **43**, 195–204.
- Bohlscheid-Thomas S, Hoting I, Boeing H & Wahrendorf J (1997a) Reproducibility and relative validity of group intake in a food frequency questionnaire developed for the German part of the EPIC project. *International Journal of Epidemiology* **26**, Suppl. 1, S59–S70.
- Bohlscheid-Thomas S, Hoting I, Boeing H & Wahrendorf J (1997b) Reproducibility and relative validity of food energy and macronutrient intake of a food frequency questionnaire developed for the German part of the EPIC project. *International Journal of Epidemiology* **26**, Suppl. 1, S71–S81.
- Federal Institute for Health Protection of Consumers and Veterinary Medicine (1998) *The German Food Code and Nutrient Database, BLS 2.1*. Berlin: BgVV-Journal 2.
- Gex-Fabry M, Raymond L & Jeanneret O (1988) Multivariate analysis of dietary patterns in 939 Swiss adults: sociodemographic parameters and alcohol consumption profiles. *International Journal of Epidemiology* **17**, 548–555.
- Gittelsohn J, Wolever TM, Harris SB, Harris-Giraldo R, Hanley AJ & Zinman B (1998) Specific patterns of food consumption and preparation are associated with diabetes and obesity in a native Canadian community. *Journal of Nutrition* **128**, 541–547.
- Hatcher L (1994) *An Easy Guide to Use the SAS System for Factor Analysis and Structural Equation Modeling*. Cary, NC: SAS Institute Inc.
- 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. *American Journal of Clinical Nutrition* **69**, 243–249.
- Klipstein-Grobusch K, Georg T & Boeing H (1997) Interviewer variability in anthropometric measurements and estimates of body composition. *International Journal of Epidemiology* **26**, Suppl. 1, S174–S180.
- Kroke A, Klipstein-Grobusch K, Voss S, Moseneder J, Thielecke F, Noack R & Boeing H (1999) Validation of a self-administered food-frequency questionnaire administered in the European Prospective Investigation into Cancer and Nutrition (EPIC) Study: comparison of energy, protein, and macronutrient intakes estimated with the doubly labeled water, urinary nitrogen, and repeated 24-h dietary recall methods. *American Journal of Clinical Nutrition* **70**, 439–447.
- Mark SD, Thomas DG & Decarli A (1996) Measurement of exposure to nutrients: an approach to the selection of informative foods. *American Journal of Epidemiology* **143**, 514–521.
- Martinez ME, Marshall JR & Sechrest L (1998) Invited commentary: factor analysis and the search for objectivity. *American Journal of Epidemiology* **148**, 17–19.
- Randall E, Marshall JR, Graham S & Brasure J (1990) Patterns in food use and their associations with nutrient intakes. *American Journal of Clinical Nutrition* **52**, 739–745.
- Rappoport L, Peters GR, Downey R, McCann T & Huff-Corzine L (1993) Gender and age differences in food cognition. *Appetite* **20**, 33–52.
- Riboli E & Kaaks R (1997) The EPIC Project: rationale and study design. *International Journal of Epidemiology* **26**, Suppl. 1, S6–S14.
- Sacks FM, Obarzanek E & Windhauser NM (1995) Rationale and design of the Dietary Approaches to Stop Hypertension Trial (DASH). A multicenter controlled-feeding study of dietary

- patterns to lower blood pressure. *Annals of Epidemiology* **5**, 108–118.
- Schofield WN, Schofield C & James WPT (1985) Basal metabolic rate – review and prediction. *Human Nutrition Clinical Nutrition* **39**, Suppl. 1, 1–96.
- Schwerin HS, Stanton JL, Riley AM Jr, Schaefer AE, Leveille GA, Elliott JG, Warwick KM & Brett BE (1981) Food eating patterns and health: a reexamination of the Ten-State and HANES I surveys. *American Journal of Clinical Nutrition* **34**, 568–580.
- Slattery ML, Boucher KM, Caan BJ, Potter JD & Ma KN (1998) Eating patterns and risk of colon cancer. *American Journal of Epidemiology* **148**, 4–16.
- Thomas DG & Mark SD (1997) Max_r: an optimal method for the selection of subsets of foods for the measurement of specific nutrient exposures. *Computer Methods and Programs in Biomedicine* **54**, 151–156.
- Whichelow MJ & Prevost AT (1996) Dietary patterns and their association with demographic, lifestyle and health variables in a random sample of British adults. *British Journal of Nutrition* **76**, 17–30.
- Willett WC (1998) *Nutritional Epidemiology*, 2nd ed. New York: Oxford University Press.
- Williams DEM, Prevost TP, Whichelow MJ, Cox BD, Day NE & Wareham NJ (2000) A cross-sectional study of dietary patterns with glucose intolerance and other features of the metabolic syndrome. *British Journal of Nutrition* **83**, 257–266.
- Wolff CB & Wolff HK (1995) Maternal eating patterns and birth weight of Mexican American infants. *Nutrition and Health* **10**, 121–134.