British Journal of Nutrition (2005), **93**, 709–716
DOI: 10.1079/BJN20051399

Comparison of two statistical approaches to predict all-cause mortality by dietary patterns in German elderly subjects

Kurt Hoffmann[1]*, Heiner Boeing[1], Paolo Boffetta[2], Gabriele Nagel[2], Philippos Orfanos[3], Pietro Ferrari[4] and Christina Bamia[3]

[1]*Department of Epidemiology, German Institute of Human Nutrition, Arthur-Scheunert-Allee 114-116, 14558 Nuthetal, Germany*
[2]*Division of Clinical Epidemiology, German Cancer Research Centre, Heidelberg, Germany*
[3]*Department of Hygiene and Epidemiology, University of Athens Medical School, Athens, Greece*
[4]*Unit of Nutrition and Cancer, International Agency for Research on Cancer, Lyon, France*

(Received 30 June 2004 – Revised 4 November 2004 – Accepted 21 December 2004)

Dietary patterns are comprehensive variables of dietary intake appropriate to model the complex exposure in nutritional research. The objectives of this study were to identify dietary patterns by applying two statistical methods, principal component analysis (PCA) and reduced rank regression (RRR), and to assess their ability to predict all-cause mortality. Motivated by previous studies we chose percentages of energy from different macronutrients as response variables in the RRR analysis. We used data from 9356 German elderly subject enrolled in the European Prospective Investigation into Cancer and Nutrition study. The first RRR pattern, subjects which explained 30·8 % of variation in energy sources and especially much variation in intake of saturated fat, monounsaturated fat and carbohydrates was a significant predictor of all-cause mortality. The pattern score had high positive loadings in all types of meat, butter, sauces and eggs, and was inversely associated with bread and fruits. After adjustment for other known risk factors, the relative risks from the lowest to highest quintiles of the first RRR pattern score were 1·0, 1·01, 0·96, 1·32, 1·61 (*P* for trend: 0·0004). In contrast, the first two PCA patterns explaining 19·7 % of food intake variation but only 7·0 % of variation in energy sources were not related to mortality. These results suggest that variation in macronutrients is meaningful for mortality and that the RRR method is more appropriate than the classic PCA method to identify dietary patterns relevant to mortality.

Nutrition: Mortality: Dietary patterns: Statistical methods

In older age, mortality is prevailingly related to chronic diseases such as cancer and CVD. Because of the growing evidence that diet plays an important role in developing chronic diseases (World Cancer Research Fund, 1997; Willett, 2000*a*; Hu & Willett, 2002), dietary factors are also expected to affect mortality rates. Recent results from cohort studies of elderly people corroborate this hypothesis Lasheras *et al.* 2000; (Kumagai *et al.* 1999; Fortes *et al.* 2000; Strandhagen *et al.* 2000; Trichopoulou *et al.* 2003). An important consequence is that changes in dietary behaviour can avoid diseases or postpone their onset with the subsequent effect of prolonged life expectancy.

To study the effect on mortality of a specific food or food group consumed regularly is neither promising nor scientifically justified because foods are consumed in combination and intake data are highly correlated. Adjustment for all other foods would result in a vast number of regression parameters to estimate. As a consequence, multicollinearity of the food intakes and the restricted power of a study may lead to unstable parameter estimates with wide confidence intervals. A possible approach is to consider diet-quality scores based on recommended diets or dietary guidelines (Patterson *et al.* 1994; Kennedy *et al.* 1995; Trichopoulou *et al.* 1995, 2003; Huijbregts *et al.* 1997; Osler & Schroll, 1997; Haines *et al.* 1999; Kant *et al.* 2000; McCullough *et al.* 2002; Seymour *et al.* 2003). However, diet-quality scores are based on selected dietary components and do not allow for the intake of some food groups.

An alternative approach is to construct dietary patterns and to study the combined effect of foods differently weighted in pattern scores (Kumagai *et al.* 1999; Osler *et al.* 2001). Principal component analysis (PCA) is the most widely applied statistical method to derive dietary patterns in nutritional science (Randall *et al.* 1992; Slattery *et al.* 1998; Hu *et al.* 2000; Osler *et al.* 2001; Schulze *et al.* 2001; Hu, 2002; Schulze & Hu, 2002; Costacou *et al.* 2003; Van Dam *et al.* 2003). PCA, or the very similar factor analysis, aims to construct linear combinations of food intakes, which explain a high proportion of the variation in food intakes. If the objective of the study is to describe typical eating patterns in the study population without reference to health outcomes, PCA is a useful tool. However, explaining much variation in food intakes by PCA does not necessarily mean that also much variation in macronutrients such as fatty acids and carbohydrates will be explained. From a theoretical point of view, dietary patterns that focus on the variation in selected nutrients and especially on changes of the mix of energy sources could be

Abbreviations: AIC, Akaike's Information Criterion; EPIC, European Prospective Investigation into Cancer and Nutrition; PCA, principal component analysis; RRR, reduced rank regression.

* **Corresponding author:** Dr Kurt Hoffmann, fax + 49 33200 88 721, email khoff@mail.dife.de

more useful in examining the effects of diet on disease incidence or mortality than the classic PCA patterns.

The focus on variation in biologically relevant nutrients was the reason to introduce reduced rank regression (RRR) in nutritional epidemiology (Hoffmann *et al.* 2004*a*). RRR is a statistical method that is more flexible than PCA because it works with two sets of variables. It aims to construct linear combinations of variables belonging to one set by maximizing the explained variation in variables of the other set. Defining the second set by a few variables that are expected to be related to the health outcome of the study should identify dietary patterns that are possibly associated with the outcome. In nutritional epidemiology, many previous studies have focused on major energy sources because they are quantitatively important in our diets and differences in the mix of energy sources among human populations are correlated with striking variation in rates of many diseases (Willett, 1998). Although large prospective studies on an individual level have not supported an important role of energy sources on cancer (Willett, 2000*b*) and stroke (He *et al.* 2003), energy intake from specific types of fat appear to be related to the risk of CVD (Jakobsen *et al.* 2004; Tanasescu *et al.* 2004) and type 2 diabetes (Salmeron *et al.* 2001) that may have small effects on total mortality (Hooper *et al.* 2001). Therefore, choosing percentages of total energy intake from different macronutrients as a second set of variables in the RRR analysis could be promising in identifying dietary patterns that are meaningful for the development of chronic diseases and for mortality.

In the present study, we applied the RRR method to food consumption data collected by means of a food frequency questionnaire and focused on the variation in percentages of energy from saturated fat, monounsaturated fat, polyunsaturated fat, protein and carbohydrates. We also applied the classic PCA method to the same data and compared the performance of the two approaches to predict mortality. We used data from the European Prospective Investigation into Cancer and Nutrition (EPIC) study. This work was involved in the EPIC-Elderly project and considered only dietary patterns in the German EPIC cohorts.

Subjects and methods

Study population

Study participants were volunteers from the two German cohorts in Potsdam and Heidelberg of the EPIC study who were aged 60 years or older at baseline. Participants were recruited between 1994 and 1998. The response rate was 22·7 % in Potsdam and 38·3 % in Heidelberg (Boeing *et al.* 1999*a*). Subjects were only included in this analysis if they had known vital status and known length of follow-up; 2·2 % of subjects were lost to follow-up. The study sample comprised 4990 elderly subjects from Potsdam (2605 women, 2385 men) and 4366 elderly subjects from Heidelberg (2088 women, 2278 men) giving 9356 subjects in total. Vital status was determined by active follow-up to June 2003. Four hundred and four study participants (221 from Potsdam, 183 from Heidelberg) died during the follow-up time, which varied between 4 and 8 years. Because of the high number of 287 diseased who had self-reported prevalent chronic diseases at enrolment we did not exclude these persons from the analysis, but we adjusted for prevalence of chronic diseases in the respective analysis.

Data collection

Dietary intake information was collected by a self-administered scanner-readable food frequency questionnaire, which included questions on the frequency and portion size of 148 food items eaten during the year preceding enrolment (Schulze *et al.* 1999). Photographs supported the estimation of portion sizes. The frequency of intake was measured using ten categories, ranging from 'never' to 'five times per d or more' (Boeing *et al.* 1999*b*). Foods were classified into twenty-three food groups based on nutrient profiles or culinary usage according to the common classification of the EPIC project (Slimani *et al.* 2002). Nutrient intake was calculated using data from the German Food Code BLSII.3 that is a slight modification of BLSII.2 (Dehne *et al.* 1999). Macronutrient intake estimated from the food frequency questionnaire was validated by the mean macronutrient intake obtained from twelve 24 h dietary recalls. The correlation coefficient between both intakes adjusted for energy varied between 0·58 for carbohydrates and 0·84 for protein (Kroke *et al.* 1999*b*). The percentage of energy from a specific macronutrient was determined by multiplying the intake of that macronutrient (in g) by the energy contained in 1 g, dividing this product by the total energy intake, and finally multiplying this quotient by 100 %.

Prevalent diseases, smoking status, educational level and physical activity were assessed through personal computer-guided interviews in the study centres at baseline. Smoking status was defined as current smoker, former smoker or non-smoker. Educational level was considered a categorical variable with the levels 'no education or primary school', 'technical school' and 'university degree'. Physical activity was differentiated between activity at work and during leisure time. Physical activity at work was categorized as unemployed, sedentary occupation, standing occupation and manual work, whereas physical activity at leisure time was assessed on an ordinal scale (minimal, moderate and heavy) according to a standardized procedure regulated in EPIC. Questionnaire data for assessing physical activity in the EPIC study were intensively described in a previous paper (Haftenberger *et al.* 2002) and party validated by use of heart-rate monitoring (Wareham *et al.* 2003). Anthropometric measurements of body height, body weight, waist girth and hip circumference were taken at baseline by trained personnel, with subjects wearing light underwear (Kroke *et al.* 1999*a*). BMI was calculated as weight (kg) divided by squared height (m^2).

Statistical methods

All analyses were performed with the SAS System® for Windows™, release 8·02 (SAS Institute Inc., Cary, NC, USA). To derive dietary patterns we applied the two methods, PCA and RRR. RRR is a statistical dimension-reduction technique recently introduced in epidemiology (Hoffmann *et al.* 2004*a*,*b*), which is similar to PCA. However, the two methods differ in their aim. RRR works with two different sets of variables called predictors and responses whereas PCA works only with predictors. Formally, PCA can be considered as a special case of RRR in which the predictors will be chosen as responses. The objective of PCA is to determine linear functions of predictors by maximizing the explained variation in all predictor variables. In contrast, RRR identifies linear functions of predictors, which explain as much response variation as possible. The calculations of pattern scores by using the PCA and RRR methods are based

on the determination of eigenvalues and corresponding eigenvectors of the covariance matrix of predictors and responses, respectively. Since the eigenvalue reflects the proportion of variation that is explained by the corresponding score, only the first pattern scores with the highest eigenvalues are of importance. To ensure comparability of the methods we chose the same number of selected patterns. A plot of the eigenvalues of the corresponding covariance matrix (scree test) indicated a break between the second and third eigenvalue for both methods. Thus, we computed only the first two pattern scores of either method.

In the subsequent RRR analysis we chose twenty-three food groups as predictors and the percentages of energy from saturated fat, monounsaturated fat, polyunsaturated fat, protein and carbohydrates as response variables. Thus, RRR focused on variation in five variables instead of the original twenty-three predictor variables. This reduction of dimension can also be considered as lessening the number of possible directions of variation prior to the statistical analysis. A more detailed representation of RRR and its application to problems of nutritional epidemiology can be found elsewhere (Hoffmann *et al.* 2004*a*,*b*). RRR and PCA are implemented in the special procedure PLS of SAS System® for Windows™, release 8·02. Patterns cannot be rotated in this procedure. Simplified patterns were determined applying the method of Schulze *et al.* (2003).

Relative risks (hazard ratios) for total mortality were calculated by Cox's proportional hazards model using the counting process style of input (SAS Institute Inc., 1999, p. 2595) with age as time variable varying across the study period. Because the proportional hazards assumption may not be realistic for all data, we stratified the regression analysis by centre using the STRATA statement of the PHREG procedure. Adjustment for sex, prevalent cancer, CHD, diabetes and hypertension, BMI, waist:hip ratio, smoking status, education level, physical activity at work, physical activity at leisure time and total energy intake was performed by addition of covariates to the model equation.

Results

The factor loadings of the first two patterns obtained by PCA and RRR are presented in Table 1. A high positive loading indicates a strong direct association between the food group and the pattern, whereas a high negative loading reflects a strong inverse association. To make visible what the important food groups for each pattern were we omitted all factor loadings between −0·2 and +0·2. The major contributors to the first PCA pattern were potatoes, vegetables, legumes, bread, all types of meat, eggs, sauces and soups which were all positively correlated with the pattern score. These foods are often eaten together at lunch or at supper in German households. The second PCA pattern was characterized by high positive loadings of vegetables, fruits, dairy products, other cereals, vegetable oils and non-alcoholic beverages, as well as by a negative loading of alcoholic beverages other than wine. This pattern reflects prudent nutrition consistent with dietary recommendations. The first RRR pattern had high positive loadings in all types of meat, butter, sauces and eggs, and was inversely associated with bread and fruits. It represents dietary habits expected to be detrimental for human health. Finally, legumes, poultry, fish and margarine were directly associated and butter, sugar and cakes were inversely associated with the second RRR pattern score.

Table 1. Factor loadings* of the dietary patterns obtained by different statistical methods in the German cohorts of the European Prospective Investigation into Cancer and Nutrition Elderly study (*n* 9356)

	Principal component analysis		Reduced rank regression	
Food groups	Pattern 1	Pattern 2	Pattern 1	Pattern 2
Potatoes and other tubers	0·25			
Vegetables	0·25	0·44		
Legumes	0·23			0·20
Fruits		0·36	−0·20	
Dairy products		0·26		
Pasta, rice and other grain				
Bread	0·22		−0·28	
Other cereals		0·27		
Red meat	0·42		0·35	
Poultry	0·25		0·21	0·22
Processed meat	0·35		0·51	
Fish and shellfish				0·20
Eggs and egg products	0·21		0·20	
Vegetable oils		0·43		
Butter			0·31	−0·64
Margarine				0·37
Sugar and confectioneary				−0·36
Cakes				−0·31
Non-alcoholic beverages		0·31		
Wine				
Other alcoholic beverages		−0·31		
Sauces and condiments	0·38		0·26	
Soups and bouillons	0·21			

* Factor loadings that are between −0·2 and +0·2 are not shown.

The variation in food groups and energy sources explained by each of the four dietary patterns is summarized in Table 2. The two PCA patterns explained 11·3 and 8·4 % of food intake variation, respectively. In contrast, the first and second RRR patterns explained only 4·4 and 6·1 % of the total variation in all twenty-three food groups. However, the variation in energy sources accounting for the two RRR patterns were 30·8 and 15·9 % and as expected much higher than the proportion of energy sources variation explained by the two PCA patterns (together 7 %). Table 2 also gives a more detailed picture of how much variation in every energy source can be explained by the patterns. As indicated, RRR patterns performed better than PCA patterns in explaining the variation in all energy sources. For example, the first RRR pattern alone accounted for 39·1 % of variation in energy from saturated fat and simultaneously for much of the variation in monounsaturated fat and carbohydrates.

As can be seen in Table 3, mean percentages of energy from different sources varied across quintiles of the two PCA and two RRR dietary pattern scores. Because of the high number of subjects all trends were statistically significant ($P<0·0001$). However, the most striking variation was that of energy from saturated fat, monounsaturated fat and carbohydrates across quintiles of the first RRR pattern. A high score of the first RRR pattern reflected a diet rich in saturated and monounsaturated fat and poor in carbohydrates. Energy from total fat increased from 25·9 % in the lowest quintile to 37·2 % in the highest quintile of the pattern score, whereas the proportion of energy from carbohydrates simultaneously decreased from 48·0 to 37·6 %. The energy composition in the highest quintile did not meet current dietary recommendations, which advise an intake of less than 35 % of energy from fat and more than 45 % of energy from carbohydrates (Institute of Medicine of the National Academies, 2002).

Table 2. Explained variation (%) of food groups and energy sources by dietary patterns obtained by different statistical methods in the German cohorts of the European Prospective Investigation into Cancer and Nutrition Elderly study (*n* 9356)

	Principal component analysis			Reduced rank regression		
	Pattern 1	Pattern 2	Total*	Pattern 1	Pattern 2	Total*
Explained variation (%) of:						
All twenty-three food groups	11·3	8·4	19·7	4·4	6·1	10·5
All five energy sources	4·4	2·6	7·0	30·8	15·9	46·7
Energy from saturated fat (%)	1·5	1·2	2·7	39·1	20·4	59·5
Energy from monounsaturated fat (%)	8·0	0·2	8·2	57·7	0·9	58·6
Energy from polyunsaturated fat (%)	1·0	2·9	3·9	12·2	29·2	41·4
Energy from protein (%)	1·5	0·3	1·8	12·0	28·7	40·7
Energy from carbohydrates (%)	10·0	8·4	18·4	33·2	0·1	33·3

* Proportion of variation explained by pattern 1 and pattern 2 together.

The relative risks of all-cause mortality associated with an increase in each pattern score by one standard deviation are presented in Table 4. The relative risks adjusted for centre and sex were significantly different from 1·0 for the first PCA pattern (relative risk 1·12; 95 % CI 1·01, 1·24) and for the first RRR pattern (relative risk 1·25; 95 % CI 1·14, 1·40). After additional adjustment for prevalent cancer, prevalent CHD, prevalent diabetes, prevalent hypertension, BMI and waist:hip ratio, the first PCA pattern was no more significant. In contrast, the first RRR pattern remained significant even after further adjustment for smoking status, educational level, physical activity at work, physical activity at leisure time and total energy intake (relative risk 1·20; 95 % CI 1·09, 1·31). Comparing the goodness of fit of the fully adjusted models by Akaike's Information Criterion (AIC) suggests the use of RRR patterns (AIC = 5580) instead of PCA patterns (AIC = 5592). Also a fully adjusted model with percentages of energy from macronutrients chosen as predictors had a lower goodness of fit (AIC = 5594). Actually, no single energy source had a significant effect on mortality. Moreover, substituting the RRR patterns by the Mediterranean diet-quality score (Trichopoulou *et al.* 1995, 2003) was associated with a decrease of model fit and a loss of significance (data not shown).

Table 3. Mean percentages of energy from different sources according to quintiles of dietary pattern scores in the German cohorts of the European Prospective Investigation into Cancer and Nutrition Elderly study (*n* 9356)

	Quintiles of dietary pattern scores					
Percentage of energy from:	1	2	3	4	5	Trend*
PCA† Pattern 1						
Saturated fat	13·4	13·6	13·7	14·0	14·4	→
Monounsaturated fat	10·9	11·3	11·6	11·9	12·5	→
Polyunsaturated fat	5·6	5·9	5·9	6·0	6·1	→
Total fat	29·9	30·8	31·2	31·9	33·0	→
Protein	13·4	13·6	13·7	13·8	14·1	→
Carbohydrates	45·8	44·6	43·6	42·6	40·2	←
PCA Pattern 2						
Saturated fat	14·1	14·3	13·9	13·6	13·2	←
Monounsaturated fat	11·8	11·9	11·6	11·6	11·4	←
Polyunsaturated fat	5·6	5·8	5·9	6·0	6·2	→
Total fat	31·5	32·0	31·4	31·2	30·8	←
Protein	13·4	13·7	13·8	13·9	13·8	→
Carbohydrates	39·7	42·7	43·9	44·7	45·5	→
RRR‡ Pattern 1						
Saturated fat	11·3	12·7	13·7	14·8	16·5	→
Monounsaturated fat	9·5	10·6	11·6	12·5	14·0	→
Polyunsaturated fat	5·1	5·6	5·9	6·3	6·7	→
Total fat	25·9	28·9	31·2	33·6	37·2	→
Protein	12·6	13·3	13·7	14·1	14·8	→
Carbohydrates	48·0	45·8	43·6	41·5	37·6	←
RRR Pattern 2						
Saturated fat	16·1	14·4	13·3	12·7	12·6	←
Monounsaturated fat	12·2	11·7	11·3	11·3	11·7	←
Polyunsaturated fat	4·7	5·3	5·8	6·4	7·4	→
Total fat	33·0	31·4	30·4	30·4	31·7	←
Protein	12·0	13·1	13·7	14·4	15·4	→
Carbohydrates	42·1	43·5	44·3	44·2	42·5	→

PCA, principal component analysis; RRR, reduced rank regression.
* All trends across quintiles of dietary patterns were significant ($P<0·0001$).
† PCA and RRR were applied to twenty-three food groups defined in Table 1.
‡ In the RRR analysis the percentages of energy from saturated fat, monounsaturated fat, polyunsaturated fat, protein and carbohydrates were used as response variables.

Table 5 shows the relative risks of death according to quintiles of dietary pattern scores by taking into account all confounders as before. Again, the first RRR pattern was the only significant predictor of total mortality. The relative risks across increasing quintiles of this pattern score were 1·0, 1·01, 0·96, 1·32 and 1·61 (95 % CI 1·17, 2·21; *P* for trend: 0·0004). Interpreting this relationship as effect of energy composition means that only a simultaneous high proportion of energy from total fat and protein and low proportion of energy from carbohydrates was associated with increased mortality risk.

In Table 6, means and percentages of different baseline characteristics according to quintiles of the first RRR pattern are summarized. Individuals with high scores had a higher BMI, were more likely to come from the Potsdam centre, tended to smoke, had a lower education level and were more likely to be unemployed. Confounding effects on relative risk estimates due to these associations were taken into account as before.

To examine the robustness of our findings and to reduce the dependency of the dietary pattern from the data we simplified the first RRR pattern score. At first, we shortened the score by neglecting all food groups with loadings between −0·2 and +0·2. Then we formed a simplified pattern (Schulze *et al.* 2003) by defining the coefficients of the remaining food groups as +1 or as −1 depending on the sign of the factor loading. The resulting simplified pattern score is simply the sum of the standardized food groups of red meat, poultry, processed meat, butter, sauces and eggs, minus the sum of the standardized food groups of bread and fruits. The shortened and the simplified versions of the first RRR pattern were separately used as a predictor of mortality (see Table 5). The relative risks for the shortened and simplified patterns were attenuated as compared with the full pattern and the trend across increasing

Table 4. Relative risk and 95 % CI of mortality according to standardized* increase of pattern scores obtained by different statistical methods in the German cohorts of the European Prospective Investigation into Cancer and Nutrition Elderly study (*n* 9356)

	Principal component analysis (PCA)†				Reduced rank regression (RRR)‡			
	Pattern 1		Pattern 2		Pattern 1		Pattern 2	
	Relative risk	95 % CI	Relative risk	95 % CI	Relative risk	95 % CI	Relative risk	95 % CI
Model 1§	1·12	1·01, 1·24	0·91	0·82, 1·01	1·25	1·14, 1·40	0·98	0·89, 1·07
Model 2	1·11	0·98, 1·23	0·92	0·82, 1·02	1·23	1·12, 1·34	0·94	0·86, 1·04
Model 3	1·10	0·96, 1·28	0·99	0·89, 1·10	1·20	1·09, 1·31	0·96	0·87, 1·06

* A unit increase of the pattern score is an increment of SD.
† PCA and RRR were applied to twenty-three food groups defined in Table 1.
‡ In the RRR analysis the percentages of energy from saturated fat, monounsaturated fat, polyunsaturated fat, protein, and carbohydrates were used as response variables.
§ Model 1: adjusted for centre and sex. Model 2: adjusted for centre, sex, prevalent cancer, CHD, diabetes and hypertension, BMI and waist:hip ratio. Model 3: adjusted for all variables included in model 2 and additionally for smoking status, education level, physical activity at work, physical activity at leisure time and total energy intake.

quintiles of the score was only borderline significant in both cases ($P=0{\cdot}062$ and $P=0{\cdot}061$).

As further sensitivity analysis we excluded all participants with prevalent chronic diseases (cancer, CHD, diabetes) and prevalent hypertension. The first RRR pattern only slightly changed after exclusion. The factor loadings were almost the same as before with differences of corresponding loadings varying between $-0{\cdot}01$ and $+0{\cdot}01$. The proportion of variation explained by the first RRR pattern was 30·7 %, which represented a decrease of only 0·1 %. The adjusted relative risk associated with an increase of the first RRR score by SD was 1·24 (95 % CI 1·05, 1·48) and therefore similar to the result obtained with the total sample, although it was less precise because of the exclusion of a relatively large proportion of deaths.

Discussion

In the German cohorts of the EPIC-Elderly study we found a significant association between a specific dietary pattern and total mortality. This pattern was derived by the statistical RRR method recently introduced in nutritional epidemiology (Hoffmann *et al.* 2004*a*). The RRR pattern score adequately reflects variation in energy sources and corroborates the hypothesis that certain combinations of energy sources affect human health. A high RRR pattern score, which was associated with high intake of fat and protein and low intake of carbohydrates, increased the risk of death. Subjects with a pattern score belonging to the highest quintile obtained on average 37·2 % of their energy from fat and 37·6 % from carbohydrates and thus did not meet current dietary recommendations (Institute of Medicine of the National Academies, 2002). Food groups that contributed to this unfavourable pattern of energy sources were red meat, poultry, processed meat, butter, sauces and eggs, whereas a high intake of bread and fruits decreased the pattern score. These food groups did not have simultaneously high loadings in one of the first two PCA patterns that reflect a typical German diet and a diet supposed to be prudent.

The most appealing advantage of the new RRR method compared with PCA or traditional factor analysis is the capability to incorporate prior information. The type of prior information needed is the knowledge of diet-related variables that are expected to be predictive for the health outcome under study. Previous correlation and epidemiological studies indicated that the mix of energy sources in diet could be important for health (Hu *et al.* 1997; Willett, 1998; Liu & Manson, 2001; Salmeron *et al.* 2001; Mann, 2002). Evidence clearly suggests that dietary fat affects the lipid and lipoprotein risk profile (Kris-Etherton *et al.* 2002; Lichtenstein, 2003; Wolfram, 2003; Pelkman *et al.* 2004). Because the number of energy sources is markedly smaller than the number of foods, focusing on the set of energy sources as done in the present RRR analysis results in a reduction of dimension. This reduction can be considered gain by prior knowledge since explaining a high proportion of

Table 5. Relative risk* and 95 % CI of mortality according to quintiles of pattern scores in the German cohorts of the European Prospective Investigation into Cancer and Nutrition Elderly study (*n* 9356)

	Quintiles of dietary patterns									
	1		2		2		4		5	
Dietary patterns	Relative risk	Relative risk	95 % CI	Relative risk	95 % CI	Relative risk	95 % CI	Relative risk	95 % CI	*P* for trend
PCA pattern 1	1·0	0·83	0·57, 1·22	1·00	0·70, 1·45	1·03	0·70, 1·51	1·06	0·68, 1·65	0·50
PCA pattern 2	1·0	0·91	0·68, 1·22	0·90	0·66, 1·23	1·10	0·81, 1·51	0·80	0·55, 1·15	0·61
RRR pattern 1	1·0	1·01	0·70, 1·46	0·96	0·66, 1·38	1·32	0·95, 1·85	1·61	1·17, 2·21	0·0004
RRR pattern 2	1·0	0·87	0·63, 1·21	0·81	0·57, 1·13	1·07	0·78, 1·48	0·96	0·70, 1·33	0·74
Shortened† RRR 1	1·0	1·07	0·73, 1·58	1·25	0·87, 1·80	1·35	0·94, 1·94	1·34	0·93, 1·94	0·062
Simplified‡ RRR 1	1·0	0·99	0·67, 1·45	1·24	0·86, 1·77	1·26	0·88, 1·79	1·31	0·91, 1·87	0·061

PCA, principal component analysis; RRR, reduced rank regression.
* Adjusted for centre, sex, prevalent cancer, CHD, diabetes and hypertension, BMI, waist:hip ratio, smoking status, education level, physical activity at work, physical activity at leisure time and total energy intake.
† Derived from the original pattern by neglecting all food groups with factor loadings between $-0{\cdot}2$ and $+0{\cdot}2$.
‡ Derived from the shortened pattern by setting all positive factor loadings equal to $+1$ and all negative loadings equal to -1.

Table 6. Baseline characteristics according to quintiles of the first reduced rank regression (RRR) pattern score in the German cohorts of the European Prospective Investigation into Cancer and Nutrition Elderly study (*n* 9356)

	Quintiles of the first RRR pattern					
Characteristics	1	2	3	4	5	*P* for trend
Mean age (years)	62·5	62·5	62·7	62·7	62·6	0·02
Mean BMI (kg/m^2)	27·1	27·3	27·7	27·7	28·2	<0·0001
Mean waist:hip ratio	0·90	0·89	0·89	0·90	0·91	0·0007
Potsdam (%)	0·46	0·53	0·57	0·59	0·57	<0·0001
Male (%)	0·43	0·55	0·56	0·51	0·46	0·59
Current smokers (%)	0·11	0·11	0·12	0·16	0·20	<0·0001
Former smokers (%)	0·42	0·36	0·36	0·37	0·38	0·08
Education level (%)						
None/primary school	0·36	0·40	0·41	0·42	0·43	<0·0001
Technical school	0·30	0·31	0·31	0·30	0·30	0·53
University degree	0·34	0·29	0·28	0·28	0·27	<0·0001
Physical activity at work (%)						
Unemployed	0·74	0·79	0·81	0·80	0·79	<0·0001
Sedentary occupation	0·16	0·13	0·12	0·11	0·12	<0·0001
Standing occupation	0·07	0·07	0·06	0·07	0·07	0·56
Manual work	0·02	0·01	0·01	0·02	0·02	0·79
Physical activity at leisure time (%)						
Minimal	0·23	0·22	0·19	0·21	0·22	0·32
Moderate	0·32	0·32	0·31	0·33	0·31	0·56
Heavy	0·45	0·46	0·50	0·46	0·47	0·17

variation in a smaller set of variables is much easier than in a larger set of variables. The result of the present study, that in contrast to PCA patterns the first RRR pattern was significantly associated with all-cause mortality, confirms prior knowledge and corroborates the hypothesis that variation in the mix of energy sources is meaningful for mortality. However, whether the macronutrient composition or other dietary components highly correlated with some macronutrients are linked to disease prevention and higher life expectancy remains unclear (Sacks & Katan, 2002). Randomized clinical trials and intervention studies could help to find out the crucial components and aspects of diet that could be the response variables in future RRR analyses.

In the current study several limitations should be considered. First, dietary data were collected by food frequency questionnaire and, therefore, were subject to considerable measurement error. Because dietary patterns use intakes of all foods the effect of measurement error on the pattern score will be high but in general cannot be quantified.

Second, dietary intake assessed by a single food frequency questionnaire referred to a period of 1 year and cannot be considered lifetime exposure. Diet may change over lifetime and these changes in dietary habits may have an impact on mortality. Repeated measurements of dietary intake over many years would be necessary to model diet history and to explore the short- and long-term effects on life expectancy.

Third, the definition and number of food groups may have an effect on the results of a pattern analysis. Because of adherence to the food group classification of the EPIC project and the necessary consistency and comparability of dietary data from both German EPIC cohorts, we chose fewer food groups in the present study than in previous studies (Hoffmann *et al.* 2004*a*,*b*). A drawback of a small number of food groups is that foods with potential different health effects are possibly combined in the same group and important variation in dietary intake can possibly not be reflected by patterns derived subsequently.

Fourth, residual confounding due to insufficient consideration of socio-demographic, occupational and other environmental factors may have caused bias in relative risk estimation. For example, bias may probably be attributed to modelling the important exposures, smoking and physical activity, only by categorical variables with few categories. Interactions between nutrition, physical activity and obesity were not considered in the present study although these risk factors could be components of the same sufficient condition for mortality following the concept of Rothman (1976).

Fifth, the low response rate of the cohort study calls into question whether the cohort is representative for the German population. Distinctions in dietary habits may lead to patterns derived in the study population that are quite different from those of the target population.

Sixth, the findings of an RRR analysis are not completely reproducible by other studies because they depend on the data at hand. The construction of simplified patterns may be helpful to find more robust results and possibly form a basis for dietary recommendations.

Seventh, the choice of response variables is not unique and seems to be somewhat arbitrary. Two RRR analyses with different sets of responses will, in general lead to different patterns.

However, the non-uniqueness of responses can also be considered a strength of the RRR method because RRR can allow for new findings in research and additional information sources by modifying the set of response variables in future studies. For example, in the case of available blood samples, biomarkers may be more informative than nutrients because they can provide the link between diet and the health outcome. Thus, choosing appropriate biomarkers as responses in the RRR analysis is a promising way to quantify indirect effects of diet on chronic diseases and mortality by specifying possible pathways (Hoffmann *et al.* 2004*b*).

In conclusion, the RRR method is a powerful tool to derive dietary patterns in nutritional epidemiology. It is a flexible method that combines the strength of PCA to consider inter-correlations of dietary intakes and the advantage of diet-quality scores to allow for use of biologically plausible prior information.

The application of RRR in the present study to evaluate the impact of diet on mortality suggests that a diet rich in fat and poor in carbohydrates decreases life expectancy.

Acknowledgements

The present study was supported by the Quality of Life and Management ot Living Resources Programme of the European Commission (DG Research, contact no. QLK6-CT-2001-00241) for the project EPIC-Elderly coordinated by the Department of Hygiene and Epidemiology, University of Athens Medical School; the Europe against Cancer Programme of the European Commission (DG SANCO) for the project EPIC coordinated by the International Agency for Research on Cancer (WHO); and by grants of the Deutsche krebshilfe.

References

Boeing H, Korfmann A & Bergmann MM (1999) Recruitment procedures of EPIC-Germany. *Ann Nutr Metab* **43**, 205–215.

Boeing H, Wahrendorf J & Becker N (1999) EPIC-Germany – a source for studies into diet and risk of chronic diseases. *Ann Nutr Metab* **43**, 195–204.

Costacou T, Bamia C, Ferrari P, Riboli E, Trichopoulos D & Trichopoulou A (2003) Tracing the Mediterranean diet through principal components and cluster analyses in the Greek population. *Eur J Clin Nutr* **57**, 1378–1385.

Dehne LI, Klemm C, Henseler G & Herrmann-Kunz E (1999) The German Food Code and Nutrient Data Base (BLSII.2). *Eur J Epidemiol* **15**, 255–259.

Fortes C, Forastiere F, Farchi S, Rapiti E, Pastori G & Perucci CA (2000) Diet and overall survival in a cohort of very elderly people. *Epidemiology* **11**, 440–445.

Haftenberger M, Schuit AJ, Tormo MJ, *et al.* (2002) Physical activity of subjects aged 50–64 years involved in the European Prospective Investigation into Cancer and Nutrition (EPIC). *Public Health Nutr* **5**, 1163–1176.

Haines PS, Siega-Riz AM & Popkin BM (1999) The Diet Quality Index revised: a measurement instrument for populations. *J Am Diet Assoc* **99**, 697–704.

He K, Merchant A, Rimm EB, Rosner BA, Stampfer MJ, Willett WC & Ascherio A (2003) Dietary fat intake and risk of stroke in male US healthcare professionals: 14 year prospective cohort study. *Br Med J* **327**, 777–782.

Hoffmann K, Schulze MB, Schienkiewitz A, Nöthlings U & Boeing H (2004*a*) Application of a new statistical method to derive dietary patterns in nutritional epidemiology. *Am J Epidemiol* **159**, 935–944.

Hoffmann K, Zyriax BC, Boeing H & Windler E (2004*b*) A dietary pattern derived to explain biomarker variation is strongly associated with risk of coronary artery disease. *Am J Clin Nutr* **80**, 633–640.

Hooper L, Summerbell CD, Higgins JP, Thompson RL, Capps NE, Smith GD, Riemersma RA & Ebrahim S (2001) Dietary fat intake and prevention of cardiovascular disease: systematic review. *Br Med J* **322**, 757–763.

Hu FB (2002) Dietary pattern analysis: a new direction in nutritional epidemiology. *Curr Opin Lipidol* **13**, 3–9.

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.

Hu FB, Stampfer MJ, Manson JE, Rimm E, Colditz GA, Rosner BA, Hennekens CH & Willett WC (1997) Dietary fat intake and the risk of coronary heart disease in women. *N Engl J Med* **337**, 1491–1499.

Hu FB & Willett WC (2002) Optimal diets for prevention of coronary heart disease. *JAMA* **288**, 2569–2578.

Huijbregts P, Feskens E, Räsänen L, Fidanza F, Nissinen A, Menotti A & Kromhout D (1997) Dietary pattern and 20 year mortality in elderly men in Finland, Italy, and the Netherlands: longitudinal cohort study. *Br Med J* **315**, 13–17.

Institute of Medicine of the National Academies, Panel on Dietary Reference Intakes for Macronutrients (2002) *Dietary Reference Intakes for Energy, Carbohydrate, Fiber, Fat, Fatty Acids, Cholesterol, Protein, and Amino Acids*. Washington, DC: National Academy Press.

Jakobsen MU, Overvad K, Dyerberg J, Schroll M & Heitmann BL (2004) Dietary fat and risk of coronary heart disease: possible effect modification by gender and age. *Am J Epidemiol* **160**, 141–149.

Kant AK, Schatzkin A, Graubard BI & Schairer C (2000) A prospective study of diet quality and mortality in women. *JAMA* **283**, 2109–2115.

Kennedy ET, Ohls J, Carlson S & Fleming K (1995) The Healthy Eating Index: design and applications. *J Am Diet Assoc* **95**, 1103–1108.

Kris-Etherton PM, Kris-Etherton PM, Binkoski AE, Zhao G, Coval SM, Clemmer KF, Hecker KD, Jacques H & Etherton TD (2002) Dietary fat: assessing the evidence in support of a moderate-fat diet; the benchmark based on lipoprotein metabolism. *Proc Nutr Soc* **61**, 287–298.

Kroke A, Bergmann MM, Lotze G, Jeckel A, Klipstein-Grobusch K & Boeing H (1999*a*) Measures of quality control in the German component of the EPIC Study. *Ann Nutr Metab* **43**, 216–224.

Kroke A, Klipstein-Grobusch K, Voss S, Möseneder J, Thielecke F, Noack R & Boeing H (1999*b*) 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. *Am J Clin Nutr* **70**, 439–447.

Kumagai S, Shibata H, Watanabe S, Suzuki T & Haga H (1999) Effect of food intake pattern on all-cause mortality in the community elderly: a 7-year longitudinal study. *J Nutr Health Aging* **3**, 29–33.

Lasheras C, Fernandez S & Patterson AM (2000) Mediterranean diet and age with respect to overall survival in institutionalised, non-smoking elderly people. *Am J Clin Nutr* **71**, 987–992.

Lichtenstein AH (2003) Dietary fat and cardiovascular disease risk: quantity or quality? *J Womens Health* **12**, 109–114.

Liu S & Manson JE (2001) Dietary carbohydrates, physical inactivity, obesity, and the 'metabolic syndrome' as predictors of coronary heart disease. *Curr Opin Lipidol* **12**, 395–404.

Mann JL (2002) Diet and risk of coronary heart disease and type 2 diabetes. *Lancet* **360**, 783–789.

McCullough ML, Feskanich D, Stampfer MJ, Giovannucci EL, Rimm EB, Hu FB, Spiegelman D, Hunter DJ, Colditz GA & Willett WC (2002) Diet quality and major chronic disease risk in men and women: moving toward improved dietary guidance. *Am J Clin Nutr* **76**, 1261–1271.

Osler M, Heitmann BL, Gerdes LU, Jorgensen LM & Schroll M (2001) Dietary patterns and mortality in Danish men and women: a prospective observational study. *Br J Nutr* **85**, 219–225.

Osler M & Schroll M (1997) Diet and mortality in a cohort of elderly people in a North European community. *Int J Epidemiol* **26**, 155–159.

Patterson RE, Haines PS & Popkin BM (1994) Diet quality index: capturing a multidimensional behavior. *J Am Diet Assoc* **94**, 57–64.

Pelkman CL, Fishell VK, Maddox DH, Pearson TA, Mauger DT & Kris-Etherton PM (2004) Effects of moderate-fat (from monounsaturated fat) and low-fat weight-loss diets on the serum lipid profile in overweight and obese men and women. *Am J Clin Nutr* **79**, 204–212.

Randall E, Marshall JR, Brasure J & Graham S (1992) Dietary patterns and colon cancer in western New York. *Nutr Cancer* **18**, 265–276.

Rothman KJ (1976) Causes. *Am J Epidemiol* **104**, 587–592.

Sacks FM & Katan M (2002) Randomized clinical trials on the effects of dietary fat and carbohydrate on plasma lipoproteins and cardiovascular disease. *Am J Med* Suppl., **9B**, 13S–24S.

Salmeron J, Hu FB, Manson JE, Stampfer MJ, Colditz GA, Rimm EB & Willett WC (2001) Dietary fat intake and risk of type 2 diabetes in women. *Am J Clin Nutr* **73**, 1019–1026.

SAS Institute Inc. (1999) *SAS/STAT User's Guide*. Cary, NC: SAS Institute Inc.

Schulze MB, Brandstetter BR, Kroke A, Wahrendorf J & Boeing H (1999) Quantitative food intake in the EPIC-Germany cohorts. *Ann Nutr Metab* **43**, 235–245.

Schulze MB, Hoffmann K, Kroke A & Boeing H (2001) Dietary patterns and their association with food and nutrient intake in the European Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam study. *Br J Nutr* **85**, 363–373.

Schulze MB, Hoffmann K, Kroke A & Boeing H (2003) An approach to construct simplified measures of dietary patterns from exploratory factor analysis. *Br J Nutr* **89**, 409–418.

Schulze MB & Hu FB (2002) Dietary patterns and risk of hypertension, type 2 diabetes mellitus, and coronary heart disease. *Curr Atherosclerosis Reports* **4**, 462–467.

Seymour JD, Calle EE, Flagg EW, Coates RJ, Ford ES, Thun MJ & American Cancer Society (2003) Diet quality index as a predictor of short-term mortality in the American Cancer Society Cancer Prevention Study II Nutrition Cohort. *Am J Epidemiol* **157**, 980–988.

Slattery ML, Boucher KM, Caan BJ, Potter JP & Ma K (1998) Eating patterns and risk of colon cancer. *Am J Epidemiol* **148**, 4–16.

Slimani N, Fahey M, Welch AA, *et al.* (2002) Diversity of dietary patterns observed in the European Prospective Investigation into Cancer and Nutrition (EPIC) project. *Public Health Nutr* **5**, 1311–1328.

Strandhagen E, Hansson PO, Bosaeus I, Isaksson B & Eriksson H (2000) High fruit intake may reduce mortality among middle-aged and elderly men. The study of men born in 1913. *Eur J Clin Nutr* **54**, 337–341.

Tanasescu M, Cho E, Manson JE & Hu FB (2004) Dietary fat and cholesterol and the risk of cardiovascular disease among women with type 2 diabetes. *Am J Clin Nutr* **79**, 999–1005.

Trichopoulou A, Costacou T, Bamia C & Trichopoulos D (2003) Adherence to a Mediterranean diet and survival in a Greek population. *N Engl J Med* **348**, 2599–2608.

Trichopoulou A, Kouris-Blazos A, Wahlqvist ML, Gnardellis C, Lagiou P, Polychronopoulos E, Vassilakou T, Lipworth L & Trichopoulos D (1995) Diet and overall survival in elderly. *Br Med J* **311**, 1457–1460.

Van Dam RM, Grievink L, Ocké MC & Feskens EJM (2003) Patterns of food consumption and risk factors for cardiovascular disease in the general Dutch population. *Am J Clin Nutr* **77**, 1156–1163.

Wareham NJ, Jakes RW, Rennie KL, Schuit J, Mitchell J, Hennings S & Day NE (2003) Validity and repeatability of a simple index derived from the short physical activity questionnaire used in the European Prospective Investigation into Cancer and Nutrition, (EPIC) study. *Public Health Nutr* **6**, 407–413.

Willett W (1998) *Nutritional Epidemiology*. New York: Oxford University Press.

Willett WC (2000) Nutritional epidemiology issues in chronic disease at the turn of the century. *Epidem Rev* **22**, 82–86.

Willett WC (2000) Diet and cancer. *Oncologist* **5**, 393–404.

Wolfram G (2003) Dietary fatty acids and coronary heart disease. *Eur J Med Res* **8**, 321–324.

World Cancer Research Fund (1997) *Food, Nutrition, and the Prevention of Cancer: A Global Perspective*. Washington, DC: World Cancer Research Fund.