European Journal of Clinical Nutrition (2003) 57, 1378–1385

0954-3007/03 $25.00
www.nature.com/ejcn

ORIGINAL COMMUNICATION

Tracing the Mediterranean diet through principal components and cluster analyses in the Greek population

T Costacou[1], C Bamia[1], P Ferrari[2], E Riboli[2], D Trichopoulos[1,3] and A Trichopoulou[1]*

[1]*Department of Hygiene and Epidemiology, University of Athens Medical School, Athens, Greece;* [2]*Unit of Nutrition and Cancer, International Agency for Research on Cancer, Lyon, France; and* [3]*Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA*

Objective: To identify dietary patterns, and their socio-demographic and lifestyle correlates in a large sample of Greek adults, and assess their adherence to the traditional Mediterranean diet.
Design: Principal component (PC) analysis was used to identify dietary patterns among 28 034 participants of the Greek branch of the European Prospective Investigation into Cancer and Nutrition. Dietary information was collected through a validated, semiquantitative, food-frequency questionnaire. The extracted PCs were subsequently regressed on sociodemographic and lifestyle variables. Analyses were also performed to classify individuals with similar dietary behavior into clusters.
Results: Four PCs were identified: PC1 resembled the Mediterranean diet, PC2 approximated a vegetarian diet with emphasis on seed oils, PC3 reflected a preference for sweets, and PC4 reflected a Western diet. PC1 and PC2 were positively associated with age, education, physical activity, and nonsmoking status. Females, in comparison to males, scored higher on PC1 but lower on PC2. Males, younger, more educated individuals, nonsmokers and residents of Greater Athens (Attica) scored higher on PC3. PC4 was associated with younger age, less education, and current smoking. In cluster analyses, cluster A contrasted clusters B and C in having much higher mean PC1- and PC2-scores and substantially lower PC3- and PC4-scores. PC1 and PC4 were, respectively, positively and inversely correlated with an *a priori* Mediterranean-diet score; PC2 and PC3 were unrelated to it.
Conclusion: The Mediterranean-like PC1-score as well as the vegetarian-like PC2 were higher among older, more educated people, and were associated with a healthier lifestyle than PC4, which reflected a Western-type diet. PC1 was strongly positively associated with an *a priori* Mediterranean-diet score.
Sponsorship: The European Prospective Investigation into Cancer and Nutrition (EPIC) is coordinated by the International Agency for Research on Cancer and supported by the Europe Against Cancer Programme of the European Commission. The Greek segment of the EPIC study is also supported by the Greek Ministry of Health and the Greek Ministry of Education.
European Journal of Clinical Nutrition (2003) **57,** 1378–1385. doi:10.1038/sj.ejcn.1601699

Keywords: dietary patterns; principal-component analysis; cluster analysis; Mediterranean diet; Mediterranean-diet score

Introduction

The role of diet in the causation of many diseases including coronary heart disease and several forms of cancer has been extensively studied and many food groups, foods, or nutrients with beneficial or detrimental effects have been identified (World Cancer Research Fund and American Institute for Cancer Prevention, 1997; Willett, 1998; Lagiou *et al*, 2002). Fewer investigations have focused on the role that particular dietary patterns may play in health and disease (Slattery *et al*, 1998; Hu *et al*, 2000; Fung *et al*, 2001a, b; Osler *et al*, 2001; Terry *et al*, 2001), possibly because it is difficult to assess or even define a dietary exposure pattern. Yet, dietary patterns have the ability to integrate complex or subtle interactive effects of many dietary exposures and bypass problems generated by multiple testing and the high correlations that may exist among these exposures. Indicators of dietary patterns may be

*Correspondence: A Trichopoulou, Department of Hygiene and Epidemiology, University of Athens Medical School, Athens, Greece.
E-mail: antonia@nut.uoa.gr
Guarantor: A Trichopoulou.
Contributors: TC had the primary responsibility for writing the manuscript and performing the statistical analysis. CB and PF participated in the statistical analysis, discussion, and interpretation of results. AT, DT, and ER contributed to the design of the study. All contributors read and commented on the manuscript.
Received 23 September 2002; revised 22 October 2002; accepted 7 November 2002

developed *a priori* on the basis of previous knowledge concerning the favorable or adverse health effects of various dietary constituents (Trichopoulos & Lagiou, 2001), although from a quantitative point of view they can be thought of as arbitrary or subjective. Dietary patterns may also be described *a posteriori* on the basis of existing data. Principal component (PC) and factor analyses have both been popular statistical techniques for the *a posteriori* identification of dietary patterns. The two approaches are closely related, their main difference being that factor analysis assumes a statistical model for the existing data set, whereas PC analysis is mainly a mathematical method, and does not rely on statistical assumptions (Chatfield & Collins, 1995).

From another perspective, the diet of the Mediterranean people has received increased attention in recent years because of ecological and analytical evidence, suggesting that it is inversely associated with coronary heart disease and several forms of cancer (Willett & Sacks, 1991; Serra Majem *et al*, 1993; Willett, 1994; Trichopoulou *et al*, 1995b, Kouris-Blazos *et al*, 1999; Trichopoulou *et al*, 2000b; Kris-Etherton *et al*, 2001). Sanchez-Villegas *et al* (2002) have successfully used factor analysis to extract two major factors, corresponding to a 'Western' dietary pattern and to a 'Spanish-Mediterranean' dietary pattern, using dietary intake data reported by some 3800 alumni of the University of Navarra, Spain. We have used PC analysis to identify dietary patterns among Greek participants of the European Prospective Study into Cancer and Nutrition (EPIC) and we have evaluated, through multiple regression analyses, the dependence of these dietary patterns (scores of PCs) on several socio-demographic and lifestyle variables. Moreover, we have used cluster analysis to identify groups of individuals with similar dietary behavior and examine whether the resulting clusters can be characterized on the basis of the dietary patterns described through the extracted PCs.

Materials and methods

The study sample consisted of a total of 28 034 volunteers 20–86 y (mean = 53, s.d. = 12.7), recruited from all regions of Greece in order to participate in the Greek component of the EPIC study. EPIC is a multicountry, prospective cohort study examining the role of dietary, lifestyle, and environmental factors in the etiology of cancer and other chronic diseases. Details on the design and methods of the EPIC study and the Greek cohort were previously described in detail (Riboli & Kaaks, 1997; Trichopoulou *et al*, 2001, 2002). All participants signed an informed consent form before enrollment.

Standard interviewing procedures were used to assess demographic and lifestyle characteristics, such as age, educational attainment, smoking status, and permanent residence. All anthropometric measurements also followed standard procedures and were taken with subjects wearing light clothing and without any restrictive underwear or shoes. Body weight was measured using a digital scale to the nearest 100 g, and height to the nearest 0.1 cm. Body mass index (BMI) was calculated as weight (in kg) over height squared (in m^2). Waist and hip circumferences were measured using an inelastic tape, and were recorded to the nearest 0.1 cm.

Usual dietary intake over the past year was assessed by a validated, semiquantitative, interviewer-administered food frequency questionnaire (FFQ) (Gnardellis *et al*, 1995; Katsouyianni *et al*, 1997), including approximately 150 food items and beverages, as well as questions on habitual cooking methods, type of lipids used in cooking, etc. Standard portion sizes were used for the estimation of consumed quantities (Trichopoulou, 1992; Gnardellis *et al*, 1995), and nutrient intakes were calculated using a food composition database modified to accommodate the particularities of the Greek diet (Trichopoulou, 1992). Intakes were expressed as grams per day, accounting for seasonal variation (Katsouyianni *et al*, 1997). PC and cluster analyses were based on food items aggregated into 17 food groups, as shown in Table 1.

The traditional Mediterranean diet is characterized by high consumption of olive oil, high intake of vegetables, fruits, legumes, and cereals (and thus high intake of complex carbohydrates), moderately high fish intake, regular but modest ethanol intake, and low consumption of meat and dairy products. Trichopoulou *et al* (1995b) constructed a study-specific gradient of adherence to the traditional Greek–Mediterranean diet, the Mediterranean-diet score, by assigning values of 1 or 0 to each of the indicated components using the respective median as cutoff. A modification of this score, to include fish intake, was used in this study to examine the relationship between the *a priori* Mediterranean-diet score and data-derived dietary patterns. The score takes values from 9 (highest adherence) to 0 (lowest adherence).

Professional and leisure time physical activity were assessed by a special section of the lifestyle EPIC questionnaire (Trichopoulou *et al*, 2000a, b). Briefly, the average time per day spent on household, professional, sporting and other activities, as well as sleep, was calculated. A metabolic equivalent index was computed by assigning a multiple of resting metabolic rate (Ainsworth *et al*, 1993) to each activity (MET value). Time spent on each of the above activities was multiplied by the MET value of the activity, and all MET-hour products were summed to give a total daily MET-score, representing the amount of energy per kilogram body weight expended during an average day.

Statistical analysis

PC and cluster analyses were performed for the total cohort, as well as separately for male ($n = 11\,558$) and female (n = 16 476) participants, using the SAS statistical software (SAS Institute Inc., 1999). The original dietary variables were considered by means of their residuals from the regression of

Table 1 Food groups and food items included in EPIC

Food groups	*Food items included*
Potatoes	Potatoes, other tubers
Vegetables	Leafy, fruiting, stalk, root, grain, and pod vegetables, cabbages, mushrooms, onion, garlic, sprouts, mixed salad
Legumes	All types of legumes
Fruit	Fruits, nuts and seeds, mixed fruits, olives
Cheese	All types of cheese included
Other dairy products	Milk, yogurt, beverages made with milk, creams
Cereal	Flour, flakes, starches, pasta, rice, other grain, bread, crisp bread, rusks, breakfast cereal, salty/aperitif biscuits, dough, and pastry, etc.
Meat	Fresh meat (beef, veal, pork, mutton/lamb, goat), poultry (chicken, hen, turkey, duck, goose, rabbit), game, processed meat, entrails
Fish	Fish, crustaceans, mollusks, fish products, fish in crumbs
Eggs	Eggs and egg products
Olive oil	All types of olive oil
Vegetable seed oils	All types of vegetable seed oils
Butter	All types of butter
Margarine	All types of margarine
Sugar	Sugar, honey, jam, chocolate and products, candy bars, confectionery nonchocolate, syrup, ice cream, sorbet, etc.
Nonalcoholic beverages	Fruit and vegetable juices, carbonated/soft/isotonic drinks, coffee, tea, chicory
Ethanol	Ethanol

each food group on total energy intake (Willett, 1998) (and gender for analyses combining male and female participants), rather than the actual values. This approach was adopted in order to control for the role of energy intake (and gender, when required).

PCs analysis: Dietary patterns were extracted with PCs analysis (using the PRINCOMP procedure in SAS) using the correlation matrix. To identify the number of PCs to be retained, we used the eigenvalue of >1.0 criterion. The rationale for choosing this value is that a PC must have a variance at least as large as that of a single standardized original variable. The number of PCs described was determined by the Scree test (SAS Institute Inc., 1999), and the interpretability of each component. Food groups (residuals) with absolute scoring coefficients >0.2 were considered important contributors to a pattern; scoring coefficients can be regarded as correlation coefficients between the original variables (food groups) and the components (dietary patterns) extracted. Scores were calculated for each of the components retained by summing the standardized values of the food groups weighted by their scoring coefficients. Thus, each individual received a score for each dietary pattern retained. Multiple regression models were fitted for each of the retained components on a series of sociodemographic and lifestyle characteristics (age, grouped into five categories: <35, 35–44, 45–54, 55–64, ≥ 65 y; BMI, in 3 kg/m^2 increments; height, in 10 cm increments; WHR, in 0.5 units increments; total energy intake, in 300 kcal (1257 kJ) increments; energy expenditure in 5 MET-hours per day increments; smoking status in three categories: never, past, current smokers; educational attainment level grouped as ≤ 5, 6–11, 12, ≥ 13 y of schooling completed; and permanent residence, in five major geographic regions of Greece).

Cluster analysis: The cluster procedure in SAS was used to hierarchically cluster subjects into a smaller number of mutually exclusive groups according to their reported daily dietary intake (g/day, unstandardized). Clusters were identified according to the same list of 17 food groups with adjustment for total energy intake and gender by the residual method, as previously indicated. Cluster analysis is sensitive to outliers and thus, clusters with less than five observations were removed. The ability of the clusters to segregate study subjects was assessed by pseudo-F and pseudo-t^2 statistics, as well as by tree plots (SAS Institute Inc., 1999).

Results

Among men in our study, energy intake was derived 38% from carbohydrates, 14% from protein, 13% from saturated lipids, 22% from monounsaturated lipids, 7% from polyunsaturated lipids, and 5% from ethanol. Among women, the corresponding percentages were 40, 14, 13, 23, 7, and 1%. Mean energy intake was 9958 kJ for men and 7937 kJ for women. Trends with age were minor and inconsistent.

The PCs extracted and clusters identified were almost identical for male and female participants; thus, findings from PC and cluster analyses on the total sample are reported. Seven dietary patterns emerged from 17 food groups, explaining approximately 61% of the total variance in food intake. The number of PCs described here was determined by the Scree test (a plot of the total variance associated with each component), and the interpretability of each component. The plot of total variance associated with each component indicated that a four-factor model explained an acceptable fraction of the total variation (40%). This, together with the lack of any meaningful interpretation

Table 2 Principal components and corresponding scoring coefficients as extracted from 17 food groups consumed by participants in the Greek EPIC study

	Positive scoring coefficients	*Negative scoring coefficients*	*Variance explained (%)*
Principal component 1	Vegetables (0.41)	Cereal (−0.22)	
	Legumes (0.25)	Vegetable seed oils (−0.32)	15
(Mediterranean-type diet)	Fruit (0.22)	Margarine (−0.30)	
	Fish (0.24)	Sugar (−0.24)	
	Olive oil (0.54)		
Principal component 2	Vegetables (0.31)	Potatoes (−0.37)	
	Fruit (0.29)	Meat (−0.30)	10
(Vegeterian diet)	Other dairy (0.31)	Butter (−0.24)	
	Vegetable seed oils (0.44)	Olive oil (−0.24)	
	Margarine (0.20)		
Principal component 3	Other dairy (0.32)	Meat (−0.25)	
	Butter (0.46)	Vegetable seed oils (−0.29)	8
(Sweet-based diet)	Sugar (0.49)	Ethanol (−0.35)	
	Nonalcoholic beverages (0.22)		
Principal component 4	Potatoes (0.29)	Cereal (−0.52)	
	Vegetables (0.24)	Ethanol (−0.20)	7
(Western-type diet)	Fruit (0.20)		
	Meat (0.41)		
	Eggs (0.32)		
	Vegetable seed oil (0.23)		
	Margarine (0.26)		
	Nonalcoholic beverages (0.24)		

of the three remaining components, led us to use only the first four PCs.

Table 2 presents the retained PCs after adjustment for total energy intake and gender. Principal component 1 (PC 1) closely resembled the Mediterranean diet, with high consumption of vegetables, legumes, fruit, fish, and olive oil. Although high consumption of cereal is part of a traditional Mediterranean diet, the low consumption of this food group in PC 1 can be explained by the fact that flour is an important component of many pastries. The second principal component (PC 2) resembled a vegetarian-type diet, with high consumption of vegetables and fruit, as well as dairy products, vegetable seed oils, and margarine. Principal component 3 (PC 3) reflected an identifiable dietary pattern, where the emphasis was on high consumption of sweets (other dairy, sugar, butter, nonalcoholic beverages). Principal component 4 (PC 4) approximated a Western-type diet (potatoes, vegetables, meat, eggs, vegetable seed oils, margarine, nonalcoholic beverages) although vegetable and legume intake were still important positive contributors, indicating the strong preference of Greek people for vegetable and legume intake, even when they tend to adopt a Western-type diet. The 10th and 90th centiles of the four components were −1.23 and 1.14 for PC 1, −1.14 and 1.27 for PC 2, −1.08 and 1.12 for PC 3, and −1.16 and 1.19 for PC 4.

Table 3 shows multiple regression-derived, mutually adjusted, partial regression coefficients and corresponding standard errors and *P*-values of PCs 1, 2, 3, and 4 regressed, in turn, on sociodemographic and lifestyle variables. The ‘Mediterranean’ principal component (PC 1) was positively associated with female gender, age, educational level, BMI, height, and total physical activity, but inversely associated with waist-to-hip ratio and current smoking. As expected, PC 1 was positively associated with residence in the islands of Greece and inversely associated with residence in the northern parts of the country. The interpretation of the partial regression coefficient should be made in conjunction with the range of this component. Thus, the regression coefficient of 0.37 contrasting the more to the less-educated people implies a noticeable increase in PC 1 of about 16% of the difference between the 90th and the 10th centiles, whereas the association of BMI with PC 1 is weak (an increment of $3\,kg/m^2$ of BMI corresponds to an increase in PC 1 of about 0.4% of the difference between the 90th and the 10th centiles).

Individuals scoring high on PC 2 were more likely to be older, more educated, males, not currently smoking, residents of Northern Greece, with a slightly lower BMI and a noticeably lower waist-to-hip ratio. Similarly to PC 1, PC 2 also was positively associated with height, MET-score and past smoking status, but contrary to the ‘Mediterranean’ PC, PC 2 was also positively associated with total energy intake.

Male participants, as well as younger, more educated individuals, and residents of Attica generally scored higher on PC 3 (the ‘Sweets-based’ dietary pattern). PC 3 was inversely associated with total physical activity, total energy intake, BMI, waist-to-hip ratio, and past or current smoking of cigarettes.

The ‘Western’ principal component, PC 4, was strongly inversely associated with age and educational level. In contrast to PC 1 and PC 2, PC 4 was positively associated

Table 3 Multiple regression-derived coefficients β (and standard error, s.e.) linking, alternatively, the four major principal components summarizing dietary intakes in the Greek-EPIC population to specified predictors

Predictor characteristics	*Component 1 'Mediterranean-type diet'* β *(s.e.)*	*P-value*	*Component 2 'Vegetarian-type diet'* β *(s.e.)*	*P-value*	*Component 3 'Sweets-focused diet'* β *(s.e.)*	*P-value*	*Component 4 'Western-type diet'* β *(s.e.)*	*P-value*
Gender								
Males	Referent		Referent		Referent		Referent	
Females	0.10 (0.02)	<0.0001	−0.04 (0.02)	0.045	−0.06 (0.02)	0.010	0.03 (0.02)	0.149
Age (y)								
<35	Referent		Referent		Referent		Referent	
35–44	0.20 (0.02)	<0.0001	0.21 (0.02)	<0.0001	−0.16 (0.03)	<0.0001	−0.19 (0.03)	<0.0001
45–54	0.33 (0.02)	<0.0001	0.45 (0.02)	<0.0001	−0.17 (0.03)	<0.0001	−0.37 (0.03)	<0.0001
55–64	0.47 (0.03)	<0.0001	0.62 (0.03)	<0.0001	−0.17 (0.03)	<0.0001	−0.51 (0.03)	<0.0001
≥65	0.48 (0.03)	<0.0001	0.69 (0.03)	<0.0001	−0.03 (0.03)	0.287	−0.60 (0.03)	<0.0001
Years of schooling								
≤5	Referent		Referent		Referent		Referent	
6–11	0.10 (0.02)	<0.0001	0.06 (0.02)	0.0003	0.09 (0.02)	<0.0001	−0.08 (0.02)	<0.0001
12	0.30 (0.02)	<0.0001	0.24 (0.02)	<0.0001	0.29 (0.02)	<0.0001	−0.08 (0.03)	0.002
≥13	0.37 (0.02)	<0.0001	0.37 (0.02)	<0.0001	0.32 (0.02)	<0.0001	−0.04 (0.02)	0.093
BMI (per 3 kg/m^2	0.01 (0.004)	<0.0001	−0.02 (0.004)	<0.0001	−0.04 (0.004)	<0.0001	0.03 (0.004)	<0.0001
Height (per 10 cm)	0.04 (0.009)	<0.0001	0.02 (0.009)	0.044	0.01 (0.01)	0.224	0.04 (0.01)	<0.0001
Waist-to-hip ratio (per 0.5 units)	−0.09 (0.04)	0.017	−0.42 (0.04)	<0.0001	−0.39 (0.04)	<0.0001	0.05 (0.04)	0.208
Energy intake (per 300 kcal)	0.0008 (0.003)	0.764	0.007 (0.003)	0.011	−0.01 (0.003)	<0.0001	−0.04 (0.003)	<0.0001
MET-score (per 5 MET-h/day)	0.04 (0.005)	<0.0001	0.03 (0.005)	<0.0001	−0.02 (0.006)	0.008	0.005 (0.006)	0.361
Smoking status								
Never smokers	Referent		Referent		Referent		Referent	
Past smokers	0.07 (0.02)	<0.0001	0.13 (0.02)	<0.0001	−0.04 (0.02)	0.019	−0.01 (0.02)	0.425
Current smokers	−0.15 (0.01)	<0.0001	−0.16 (0.01)	<0.0001	−0.08 (0.02)	<0.0001	0.06 (0.02)	0.002
Residence								
Attica, including greater Athens	Referent		Referent		Referent		Referent	
Central Greece	−0.03 (0.02)	0.062	−0.35 (0.02)	<0.0001	−0.06 (0.02)	0.001	−0.20 (0.02)	<0.0001
North Greece	−1.00 (0.02)	<0.0001	0.46 (0.02)	<0.0001	−0.35 (0.02)	<0.0001	0.08 (0.02)	<0.0001
South Greece	−0.04 (0.02)	0.042	−0.68 (0.02)	<0.0001	−0.22 (0.02)	<0.0001	−0.38 (0.02)	<0.0001
Islands	0.11 (0.02)	<0.0001	−0.35 (0.02)	<0.0001	−0.10 (0.02)	<0.0001	−0.28 (0.02)	<0.0001

with current smoking. PC 4 was more strongly associated with BMI than PC 1. There was a positive association between the 'Western'-type diet and residence in Northern Greece, whereas residents of South Greece had the lowest scores on PC 4.

Three major clusters were identified in cluster analysis, using the criteria indicated in the Materials and methods section, and are depicted in Figure 1 Cluster A was characterized by high consumption of vegetables, legumes, fruit, fish, and olive oil, and low consumption of meat, dairy products, cereals (as defined in this analysis), eggs, vegetable seed oils, butter, margarine, sugar, nonalcoholic beverages, and ethanol. The other two clusters, B and C, were similar in that they were both characterized by low consumption of vegetables, legumes, fish, olive oil, and high consumption of meat, eggs, butter, margarine, and sugar products. What distinguishes clusters B and C was that consumption of cereal and alcoholic beverages was considerably lower in cluster B than in cluster C and, in particular, consumption of fruit, sugar, and nonalcoholic beverages (essentially sugar) was higher in cluster B than in cluster C. The actual size of the three clusters was 12 724, 5615, and 9695 for clusters A, B, and C, respectively. In subsequent analyses, we combined clusters B and C (hereafter referred to as cluster BC) in order to compare them with the distinct cluster A.

In Figure 2, the means (and standard errors of the means) of PCs 1, 2, 3, and 4 in cluster A and cluster BC are shown. There were striking differences with respect to the means of PC 1 and PC 2 between clusters A and BC. Differences with respect to PC 3 and PC 4 were much less striking. Thus, the odds of an individual to belong to the highest compared with the lowest quintile of PC 1 and PC 2, respectively, were approximately 24 and 5 times higher if the individual belonged to cluster A compared with cluster BC. Similar sharp contrasts were not evident for PC 3 or PC 4.

Since an *a priori* Mediterranean-diet score has been used with some success in previous studies of dietary patterns in relation to mortality, we calculated simple correlation coefficients of that score with PC 1, 2, 3, and 4. Pearson correlation coefficients were, respectively, 0.50, 0.02, −0.06, and −0.10. It is inferred that PC 1 approximates the Mediterranean-diet and that the Mediterranean-diet score can be empirically documented. Indeed, the odds of an individual to belong to the highest compared to the lowest quintile of PC 1 were 19 times higher if the individual had a Mediterranean-diet score of 5 or higher compared

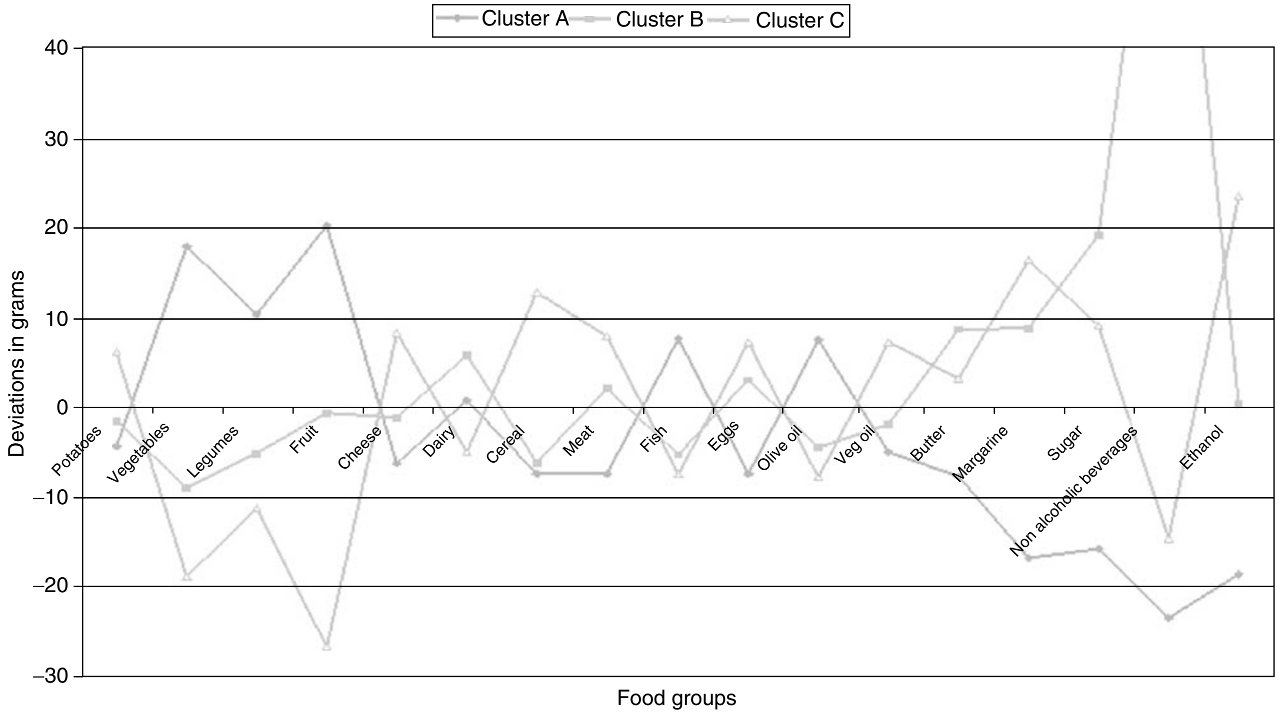

Figure 1 Three clusters of the participants in the Greek EPIC-study on the basis of their dietary intakes. The vertical axis indicates deviation of the within cluster mean dietary intake of each food group from the corresponding overall mean intake.

to a Mediterranean-diet score of less than 5. Similarly, the proportion of participants in cluster A with a Mediterranean-diet score of 5 or higher was 58%, whereas the same proportion in clusters B and C was 38 and 30%, respectively.

Discussion

We have identified four major PCs in the Greek population, one of them approximating the traditional Mediterranean diet (Trichopoulou *et al*, 1995a), another denoting a vegetarian diet, with emphasis on vegetable seed oils rather than olive oil, a third reflecting a preference for sweets (Schulze *et al*, 2001), and a fourth reflecting a Western-type diet (Slattery *et al*, 1998; Hu *et al*, 2000; Fung *et al*, 2001a, b; Osler *et al*, 2001; Terry *et al*, 2001). The Mediterranean PC was positively associated with female gender, age, educational level, physical activity, and nonsmoking status, indicating that this diet represents the choice of the educated and health conscious people. High scores of the Mediterranean PC were found in the islands, which are essentially rural, whereas lower scores were evident in Attica and Northern Greece which included the large cities of Athens and Thessaloniki. The second component (vegetarian) had similarities with the first one, but was associated with male gender, lower BMI and lower waist-to-hip ratio. This component is reminiscent of the diet frequently advocated for middle-aged men by general practitioners and cardiologists for the prevention of coronary heart disease (Willett, 1998). The third PC, with emphasis on sweets, reflected an identifiable pattern particularly in respect to its inverse association with physical activity and its association with young residents of the greater Athens area (Attica). The Western PC was strongly inversely associated with age, and level of education, and positively associated with current smoking. It appears that the Mediterranean and the Western dietary patterns have distinct sociodemographic and lifestyle profiles, as it has also been observed in Spain (Sanchez-Villegas *et al*, 2002).

We have also identified clusters of individuals that share similar dietary behavior and show a strong preference to the Mediterranean PC and the closely related vegetarian-like component. Individuals in this cluster are distinct from individuals outside this cluster both in terms of their preference towards the Mediterranean and vegetarian PCs and their relative antipathy towards the two other dietary patterns reflected in the corresponding components PC 3 and PC 4. Lastly, we have found that the *a priori* defined Mediterranean-diet score was strongly positively associated with PC 1, essentially unrelated to PC 2 and PC 3, and negatively associated with PC 4, a fact that increases confidence in the interpretability of both PC 1 and the Mediterranean-diet score. We have not attempted to identify clusters on the basis of macro- and micronutrients because diet (and thus dietary clusters) largely reflects peoples' choices which are focused on foods rather than nutrients.

The clinical implications of this study stem from the fact that high scores in the Mediterranean and vegetarian PCs are

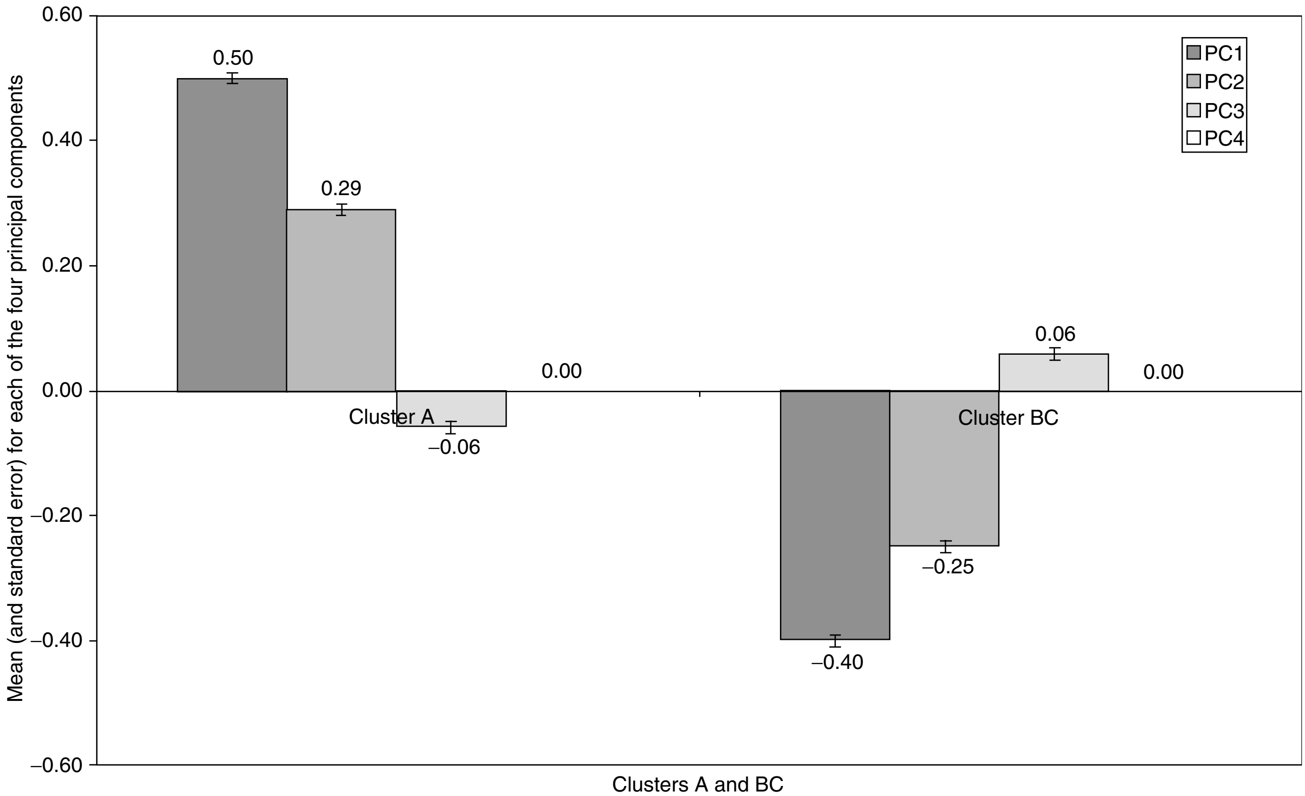

Figure 2 Mean (and standard error) of the four principal components in clusters of individuals with similar dietary intakes in the Greek population (cluster A versus clusters B and C combined).

likely to be associated with lower incidence of, and mortality from, several chronic diseases. The EPIC study will allow a prospective evaluation of the related hypotheses. An important methodological issue raises from the positive association of the Mediterranean and the vegetarian PCs with educational level (Table 3), and the well-known inverse association of socioeconomic status with morbidity and mortality from several causes in most populations. Control for mutual confounding of dietary patterns and socioeconomic status with respect to incidence of, and mortality from, the relevant diseases is crucial in the investigation of the role of dietary patterns in health and disease.

Among the advantages of the present study are the large size, the reliance on sampling from the general adult population with considerable variability on several characteristics including age and educational level, the use of a validated, semiquantitative FFQ, and the standardized measurement of all examined variables. Disadvantages of the study are the inherent problems in dietary assessment, the questionable robustness of cluster analysis, and the suboptimal correspondence between PC 4 and the Western diet.

In conclusion, we have identified in a large sample of the general Greek population a PC that closely approximates the Mediterranean diet, another that resembles a vegetarian-type diet, a third reflecting a preference for sweets, and a fourth that approximates the Western diet. Moreover, we have determined the lifestyle and sociodemographic predictors of these components. We have also identified clusters of individuals whose dietary behavior is fairly similar within clusters and fairly different between clusters and we have found that in one of these clusters PC 1 and the Mediterranean-diet scores were generally high. The current study contributes to the development of the foundations for investigating the health implications of dietary patterns during the on-going follow-up, since EPIC is a prospective cohort study. It is also possible to use more targeted measures and health education messages in order to promote a healthier diet in the population.

References

Ainsworth B, Haskell WL & Leon AS (1993): Compendium of physical activities: classification of energy costs of human physical activities. *Med. Sci. Sports Exerc.* **25**, 71–80.

Chatfield C & Collins AJ (1995): In *Introduction to Multivariate Analysis*, eds C Chatfield & JV Zidek. London: Chapman & Hall.

Kris-Etherton P, Eckel RH, Howard BV, St Jeor S, Bazzarre TL & Nutrition Committee Population Science Committee and Clinical Science Committee of the American Heart Association (2001): AHA Science Advisory: Lyon Diet Heart Study. Benefits of a Mediterranean-style, National Cholesterol Education Program/American Heart Association Step I Dietary Pattern on Cardiovascular Disease. *Circulation* **103**, 1823–1825.

Fung TT, Rimm EB, Spiegelman D, Rifai N, Tofler GH, Willett WC & Hu FB (2001a): Association between dietary patterns and plasma biomarkers of obesity and cardiovascular disease risk. *Am. J. Clin. Nutr.* **73**, 61–67.
Fung TT, Willett WC, Stampfer MJ, Manson JA & Hu FB (2001b): Dietary patterns and the risk of coronary heart disease in women. *Arch. Intern. Med.* **161**, 1857–1862.
Gnardellis C, Trichopoulou A, Katsouyianni K, Polychronopoulos E, Rimm EB & Trichopoulos D (1995): Reproducibility and validity of an extensive semi-quantitative Food Frequency Questionnaire among Greek school teachers. *Epidemiology* **6**, 74–77.
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.
Katsouyianni K, Rimm EB, Gnardellis C, Trichopoulos D, Polychronopoulos E & Trichopoulou A (1997): Reproducibility and relative validity of an extensive semi-quantitative Food Frequency Questionnaire using dietary records and chemical markers among Greek school teachers. *Int. J. Epidemiol.* **26**(Suppl 1), S118–S127.
Kouris-Blazos A, Gnardellis C, Wahlqvist ML, Trichopoulos D, Lukito W & Trichopoulou A (1999): Are the advantages of the Mediterranean diet transferable to other populations? A cohort study in Melbourne, Australia. *Br. J. Nutr.* **82**, 57–61.
Lagiou P, Trichopoulou A & Trichopoulos D (2002): Nutritional epidemiology of cancer: accomplishments and prospects. *Proc. Nutr. Soc.* **61**, 217–222.
Osler M, Heitmann BL, Gerdes LU, Jørgensen LM & Schroll M (2001): Dietary patterns and mortality in Danish men and women: a prospective observational study. *Br. J. Nutr.* **85**, 219–225.
Riboli E & Kaaks R (1997): The EPIC Project: rationale and study design. *Int. J. Epidemiol* **26**(Suppl 1), S6–S14.
Sánchez-Villegas A, Delgado-Rodríguez M, Martínez-González MA, de Irala-Estévez J & for the SUN group (2002): Gender, age, socio-demographic and lifestyle factors associated with major dietary patterns in the Spanish Project SUN (Suguimiento Universidad de Navarra). *Eur. J. Clin. Nutr.* **56**, 1–8.
SAS Institute Inc (1999): *SAS/STAT User's Guide*, Version 8, Cary, NC: SAS Institute Inc.
Schulze MB, Hoffmann K, Kroke A & Boeing H (2001): Dietary patterns and their association with food and nutrient intake in the prospective investigation into cancer and nutrition (EPIC)-Post-dam Study. *Br. J. Nutr.* **85**, 363–373.
Serra-Majem L, La Vecchia C, Ribas-Barba L, Prieto-Ramos F, Lucchini F, Ramon JM & Salleras L (1993): Changes in diet and mortality from selected cancers in southern Mediterranean countries, 1960–1989. *Eur. J. Clin. Nutr.* **47**(Suppl 1), S25–S34.
Slattery ML, Boucher KM, Caan BJ, Potter JD & Ma KN (1998): Eating patterns and risk of colon cancer. *Am. J. Epidemiol.* **148**, 4–16.
Terry P, Hu FB, Hansen H & Wolk A (2001): Prospective study of major dietary patterns and colorectal cancer risk in women. *Am. J. Epidemiol.* **154**, 1143–1149.
Trichopoulos D & Lagiou P (2001): Dietary patterns and mortality. *Br. J. Nutr.* **85**, 133–134.
Trichopoulou A (1992): *Composition of Greek Foods and Dishes* (in Greek and English), Athens: Athens School of Public Health.
Trichopoulou A, Gnardellis C, Benetou V, Lagiou P, Bamia C & Trichopoulos D (2002): Lipid, protein and carbohydrate intake in relation to body mass index. *Eur. J. Clin. Nutr.* **56**, 37–43.
Trichopoulou A, Gnardellis C, Lagiou P, Benetou V, Naska A & Trichopoulos D (2001): Physical activity and energy intake selectively predict the waist-to-hip ratio in men but not in women. *Am. J. Clin. Nutr.* **74**, 574–578.
Trichopoulou A, Gnardellis C, Lagiou P, Benetou V & Trichopoulos D (2000): Body mass index in relation to energy intake and expenditure among adults in Greece. *Epidemiology* **11**, 333–336.
Trichopoulou A, Kouris-Blazos A, Vassilakou T, Gnardellis C, Polychronopoulos E, Venizelos M, Lagiou P, Wahlqvist ML & Trichopoulos D (1995a): Diet and survival of elderly Greeks: a link to the past. *Am. J. Clin. Nutr.* **61**(Suppl 6), S1346–S1350.
Trichopoulou A, Kouris-Blazos A, Wahlqvist ML, Gnardellis C, Lagiou P, Polychronopoulos E, Vassilakou T, Lipworth L & Trichopoulos D (1995b): Diet and overall survival in the elderly. *BMJ* **311**, 1457–1460.
Trichopoulou A, Lagiou P, Kuper H & Trichopoulos D (2000b): Cancer and Mediterranean dietary traditions. *Cancer Epidemiol. Biomarkers Prev.* **9**, 869–873.
Willett WC (1998): Nutritional Epidemiology, 2nd edition, New York: Oxford University Press.
Willett WC (1994): Diet and health: what should we eat? *Science* **264**, 532–537.
Willett W & Sacks FM (1991): Chewing the fat—how much and what kind? *N. Engl. J. Med.* **324**, 121–123.
World Cancer Research Fund and American Institute for Cancer Prevention (1997): Food. In: *Nutrition and the Prevention of Cancer: A Global Perspective*. Washington, DC: World Cancer Research Fund and American Institute for Cancer Prevention.