Dietary patterns in middle-aged Irish men and women defined by cluster analysis

R Villegas, A Salim, MM Collins, A Flynn and IJ Perry*
Department of Epidemiology and Public Health, Distillery House, North Mall, University College Cork, Cork, Ireland

Abstract

Objectives: To identify and characterise dietary patterns in a middle-aged Irish population sample and study associations between these patterns, sociodemographic and anthropometric variables and major risk factors for cardiovascular disease.
Design: A cross-sectional study.
Subjects and methods: A group of 1473 men and women were sampled from 17 general practice lists in the South of Ireland. A total of 1018 attended for screening, with a response rate of 69%. Participants completed a detailed health and lifestyle questionnaire and provided a fasting blood sample for glucose, lipids and homocysteine. Dietary intake was assessed using a standard food-frequency questionnaire adapted for use in the Irish population. The food-frequency questionnaire was a modification of that used in the UK arm of the European Prospective Investigation into Cancer study, which was based on that used in the US Nurses' Health Study. Dietary patterns were assessed primarily by K-means cluster analysis, following initial principal components analysis to identify the seeds.
Results: Three dietary patterns were identified. These clusters corresponded to a traditional Irish diet, a prudent diet and a diet characterised by high consumption of alcoholic drinks and convenience foods. Cluster 1 (Traditional Diet) had the highest intakes of saturated fat (SFA), monounsaturated fat (MUFA) and percentage of total energy from fat, and the lowest polyunsaturated fat (PUFA) intake and ratio of polyunsaturated to saturated fat (P:S). Cluster 2 (Prudent Diet) was characterised by significantly higher intakes of fibre, PUFA, P:S ratio and antioxidant vitamins (vitamins C and E), and lower intakes of total fat, MUFA, SFA and cholesterol. Cluster 3 (Alcohol & Convenience Foods) had the highest intakes of alcohol, protein, cholesterol, vitamin B_{12}, vitamin B_6, folate, iron, phosphorus, selenium and zinc, and the lowest intakes of PUFA, vitamin A and antioxidant vitamins (vitamins C and E). There were significant differences between clusters in gender distribution, smoking status, physical activity, body mass index, waist circumference and serum homocysteine concentrations.
Conclusion: In this general population sample, cluster analysis methods yielded two major dietary patterns: prudent and traditional. The prudent dietary pattern is associated with other health-seeking behaviours. Study of dietary patterns will help elucidate links between diet and disease and contribute to the development of healthy eating guidelines for health promotion.

Diet is an important risk factor for chronic diseases. The conventional approach in nutrition epidemiology has focused on the relationship between specific food items and nutrients and chronic disease. However, as foods and nutrients are consumed in combination, their joint effects may be investigated by looking at dietary patterns. Furthermore, work on diet–disease relationships based on dietary patterns is of particular value in addressing issues of collinearity of nutrient intake, nutrient interactions and confounding in epidemiological studies[1]. Thus, in recent years there has been increasing interest in the identification of dietary patterns as consumed by populations[2,3]. Knowledge of specific food patterns is also important for relating diet to nutritional status and for the identification of groups at risk of under- or over-consumption of specific nutrients[4]. Understanding the patterns of core staples around which diets are formed is important for meal planning and nutritional counselling.

A number of different approaches to uncovering patterns of food intake have been developed, based on either an *a priori* or an *a posteriori* approach[1]. The dietary indexes approach is *a priori* because the pattern scores are created on the basis of current knowledge of a 'healthy diet', whereas quantitative approaches are considered

Corresponding author: Email i.perry@ucc.ie

a posteriori because the dietary patterns are derived through statistical modelling of the data[1]. Quantitative approaches include principal components analysis[5−7], factor analysis[3,8−10] and cluster analysis[4,11−17]. Cluster analysis offers advantages over the alternative quantitative approaches as it aims to identify distinct, relatively homogeneous groups based upon selected attributes (the dietary variables)[1].

The aim of the present study was to identify dietary patterns within a general population sample of middle-aged Irish men and women. We also report on associations between dietary patterns, sociodemographic and anthropometric variables, major risk factors for cardiovascular disease (CVD), prevalent CVD and glucose intolerance.

Methods

Design, subjects and methods of data collection

We performed a cross-sectional study based in primary care: The Cork and Kerry Diabetes and Heart Disease Study[18]. The overall aim of the study was to formally estimate, using standardised methods, the prevalence of glucose intolerance, including type 2 diabetes and associated heart disease risk factors, in an Irish general population sample. A detailed self-completed questionnaire data, physical measurements and fasting blood samples were obtained from a group of 1018 men and women, randomly sampled from 17 general practices in Cork and Kerry between March and August 1998. Details of the sampling methods have been reported elsewhere[18]. Subjects with CVD, known diabetes mellitus or other disease, or those receiving medication, were *included* where identified by the sampling process. A total of 1473 potential participants were identified as eligible for inclusion, of whom 1018 attended for the assessment, a response rate of 69.1%. Allowing for those who could not attend by reason of being hospitalised ($n = 5$), out of the country ($n = 5$), no longer alive ($n = 2$), outside the target age group ($n = 2$), too confused ($n = 1$) and untraceable ($n = 2$), the effective response was 69.9%. Details of the questionnaire including questions on smoking and physical activity and details of physical measurements, including measurement of body mass index (BMI), waist/hip ratio (WHR), fasting blood samples, blood pressure and electrocardiography, have been reported previously[18].

Overall obesity was defined on the basis of BMI $\geq 30 \, \mathrm{kg \, m}^{-2}$. Central obesity was defined on the basis of waist circumference (WC) and WHR: WC $> 102 \, \mathrm{cm}$ in men and $> 88 \, \mathrm{cm}$ in women (US Third Report of the National Cholesterol Education Program, Adult Treatment Panel); and WHR > 0.9 in men and > 0.85 in women (World Health Organization, WHO)[19,20]. We defined hypertension as participants with raised blood pressure (systolic blood pressure $> 140 \, \mathrm{mmHg}$ and/or diastolic blood pressure $> 90 \, \mathrm{mmHg}$) and/or self reported use of hypertensive

drugs. Details of the definition of pre-existing CVD have been reported[18]. Glucose intolerance was defined as those participants with type 2 diabetes or impaired fasting glucose, according to the current American Diabetes Association and WHO criteria[19,21]. Participants' risk of a first coronary heart disease (CHD) event was estimated using the Framingham risk equation, which incorporates and allows for the relative impact of age, sex, smoking status, systolic blood pressure, diabetes mellitus, ratio of total cholesterol to high-density lipoprotein (HDL)-cholesterol and left ventricular hypertrophy by electro-cardiographic criteria[18,22]. We defined 'high CHD risk' on the basis of an estimated absolute risk of first CHD event $> 20\%$ over 10 years.

Socio-economic status

Participants were classified by socio-economic categories, based on the standard occupational classification system of the Irish Central Statistics Office combined with educational attainment. When a participant defined herself as a housewife, the occupation of their partner was used for classification. We defined five socio-economic categories as follows: Category I (higher and lower professionals, employers/managers and own account workers with third level of education, $n = 161$); Category II (employers, managers or own account workers without third-level education, $n = 64$); Category III (farmers, $n = 138$); Category IV (non-manual workers, skilled and semi-skilled manual workers, $n = 371$); and Category V (agricultural workers and non-skilled manual workers, $n = 255$). Information for socio-economic status (SES) coding was not available for 29 participants.

Dietary data

Dietary data were collected by means of a food-frequency questionnaire (FFQ). The FFQ was an adapted version of that used in the UK arm of the European Prospective Investigation into Cancer study[23]. The latter was based on the original Willett FFQ. The questionnaire was modified by the National Nutritional Surveillance Unit, based in the Department of Health Promotion at University College Galway, to reflect the Irish diet[24,25]. The FFQ had 147 items. The FFQ was validated in an adult population using food diaries and urinary protein with *p*-aminobenzoic acid[26]. A Spearman's correlation of 0.40 was observed between protein estimates using the FFQ and the food diary, and 0.31 with the biomarker method. The relative validation of the FFQ and food diary estimates for total fat intake was 0.42 and 0.49 for saturated fats. Additional details of this FFQ have been given previously[25]. This questionnaire has been used in the Irish National Health and Lifestyle Survey[24].

Out of the 1018 participants, 937 completed the FFQ. We excluded from this study participants with implausible scores for total energy intake (< 500 or $> 3500 \, \mathrm{kcal \, day}^{-1}$ in women and < 800 or $> 4200 \, \mathrm{kcal \, day}^{-1}$ in men)[27], leaving

851 participants with dietary data for these analyses. The National Nutritional Surveillance Unit estimated the nutritional components from the food items in the FFQ using a computer program based on McCance and Widdowson's tables of composition of foods[28].

Items of the FFQ expressed in terms of the proportion of total mass of food consumed (g day^{-1} or ml day^{-1} in the case of alcoholic drinks or beverages) were aggregated into 22 mutually exclusive food groups similar to those used by Pryer *et al.*[17], which were based on the 51 food/drinks groups defined by Gregory *et al.*[29] (see Appendix).

Fasting blood samples

Blood samples were taken for fasting lipoprotein profile, glucose and homocysteine. Lipoprotein profile and blood glucose were analysed using the Roche Hitachi 747 Multichemistry analyser (Diamond Diagnostics, Holliston, MA, USA) and the Olympus 640 Discrete analyser (Olympus, Harnburg, Germany), respectively. Homocysteine was measured using a commercially available fluorescence polarisation immunoassay (Abbot Diagnostics Abbot Park, IL, USA). The inter-assay coefficients of variation were 7% at 7.6 mmol l^{-1}, 8% at 13.2 mmol l^{-1} and 10% at 26.3 mmol l^{-1}. Data on fasting homocysteine were available for 901 participants.

We excluded participants who did not fast for more than 8 h or did not provide information on their fasting status ($n = 101$). Following exclusions, the number of participants with valid data for glucose and homocysteine was 915 and 900, respectively. Data on triglycerides were available for 913 participants, and data on total cholesterol, HDL-cholesterol, low-density lipoprotein (LDL)-cholesterol and very-low-density lipoprotein (VLDL)-cholesterol were available for 900 participants. We defined high homocysteine levels as those at or above the 95th percentile of the distribution.

Statistical analysis

We used cluster analysis to identify dietary patterns and to segregate subjects based on the similarity of diet. Continuous food groups were standardised by converting to the standard normal deviate to ensure that clusters were not influenced by food groups with a high specific gravity, such as beverages. We chose food variables because we wanted to identify food patterns clusters. K-means cluster analysis was used to define clusters of subjects using the cluster analysis option in the MINITAB® software package, version 13 (Minitab Inc., State College, PA, USA).

This procedure attempts to identify relatively homogeneous groups of cases based on selected characteristics, using an algorithm that can handle large numbers of cases. In K-means cluster analysis, the homogeneity of cases within a cluster is measured by the total within-cluster sum of squares. Cluster memberships are determined by sequentially moving cases from one cluster to another so that the total within-cluster sum of squares is minimised.

The algorithm requires the number of clusters to be specified prior to analysis. It is possible to identify seeds using information derived from previous research. However, this approach is open to bias. Thus we adopted an *a posteriori* approach, using principal components analysis with the food groups to identify the seeds. From the initial exploratory analysis we judged that there were three clusters in these data. We initially chose the first, second and third components, but this did not produce three distinctive clusters. We therefore decided to base our seeds on the first, the second and one other component. We added consecutive components and found that the addition of component 14 yielded three distinct clusters with most variables showing significant between-cluster variation. The clusters are stable if a random subset of participants is used in the cluster. On average, 85.4% of the observations are correctly classified by subset analysis.

Three clusters were identified. Differences in food group consumption were investigated using non-parametric analysis of variance: the Kruskal–Wallis test. Differences in nutrient and daily energy intakes by cluster were assessed using analysis of variance. Nutrient variables that were not normally distributed were log-transformed. As ethanol intakes could not be transformed to a normal distribution, we used the Kruskal–Wallis test to compare intakes across clusters. As two of the three clusters accounted for 96% of study participants, all comparisons between clusters were repeated in analyses confined to the two major clusters.

We investigated differences in sociodemographic characteristics, lifestyle exposures and disease prevalence by cluster using the chi-square test. Age- and sex-adjusted means of BMI, WC, WHR, glucose, blood lipids, blood pressure measurements and homocysteine were calculated for each cluster and compared using analysis of covariance.

Results

We identified three distinct groups in this population on cluster analyses. A total of 480 participants (56.5%) were in cluster 1, 340 (38%) in cluster 2 and 31 (3.6%) in cluster 3. Median consumption (g day^{-1} or ml day^{-1}) of food groups in each cluster are shown in Table 1. Kruskal–Wallis tests revealed that only egg and egg products and miscellaneous foods were consumed at similar levels across the three clusters (data not shown).

Cluster 1 had the highest median intakes of beverages (non-alcoholic drinks), white bread and refined cereals, butter, whole milk and dairy products, desserts and sweets, and the lowest median intakes of fish and alcohol. We describe Cluster 1 as a 'Traditional Diet' pattern. Cluster 2 had the highest median intakes of pasta and rice, brown breads and unrefined cereals, poultry, fish, low-fat milk and dairy products, salad dressings, fruit and vegetables, and the lowest median intakes of chips, white bread and

Table 1 Median intakes of food groups (g day^{-1} or ml day^{-1}*) by cluster, $n = 851$

	Cluster 1 – Traditional Diet ($n = 480$)	Cluster 2 – Prudent Diet ($n = 340$)	Cluster 3 – Other [Alcohol & Convenience Foods] ($n = 31$)
Alcoholic drinks*	5	14	1344
Drinks*	13	9	7
Pasta/rice	0	12	0
White bread	42	28	35
Brown bread	77	98	75
Chips	24	13	36
Butter	8	0	4
Spreads	0	0	0
Salad dressings	3	14	1
High-fat dairy	284	0	142
Low-fat dairy	0	284	0
Meat	120	88	136
Poultry	19	56	19
Fish	17	33	27
Meat products	11	6	13
Egg	7	7	9
Fruits	147	279	125
Vegetables	333	379	317
Desserts	123	98	46
Sweets	129	6	15
Snacks	0	0	0
Miscellaneous	53	65	67

refined cereals, butter, high-fat dairy, meat, meat products and sweets. We use the term 'Prudent Diet' to describe this cluster. Cluster 3 had the highest median intakes of alcoholic drinks, meat, meat products and chips, and the lowest median intakes of desserts, fruits, vegetables, brown bread and unrefined cereals, and drinks. We use the term 'Alcohol & Convenience Foods' pattern to describe this dietary cluster.

Relative to the Prudent Diet cluster, participants in the Traditional Diet and the Alcohol & Convenience Foods clusters had lower intakes of more 'healthy' food groups (such as fruit, vegetables, low-fat dairy products, poultry, fish and whole-grain products) and higher intakes of foods richer in total fat and saturated fatty acids (SFA) (such as high-fat dairy products, butter, meat and meat products). The Prudent Diet cluster was characterised by relatively high intakes of food groups that are recommended in health promotion programmes and lower intakes of meat (red meat), meat products, sweets, high-fat dairy and white bread (white bread and refined cereals).

Table 2 shows the mean intakes of nutrients by cluster. Nutrient intakes varied significantly across the clusters with the exception of proteins, carbohydrates, starch and total sugars. The Prudent Diet cluster was characterised by significantly higher intakes of fibre, polyunsaturated fatty acids (PUFA), ratio of polyunsaturated to saturated fat (P:S) and antioxidant vitamins (vitamins C and E), and the lowest intakes of cholesterol, total fat, monounsaturated fatty acids (MUFA) and SFA. The Traditional Diet cluster had the highest intakes of total fat, MUFA, SFA and retinol, and the lowest P:S ratio and protein, vitamin B$_6$, folate and vitamin D intakes. The Alcohol & Convenience Foods group had the highest intakes of energy, cholesterol, ethanol, protein, vitamin B$_{12}$, vitamin B$_6$ and folate, and

the lowest intakes of PUFA, retinol and antioxidant vitamins (vitamins C and E).

In further analyses we compared the means of nutrient and energy intake in the two major clusters. Significant differences were observed for all of the major nutrients with the exception of carbohydrates and thiamin. Overall, the Prudent Diet was associated with lower intakes of nutrients linked with increased risk of chronic disease and higher intakes of nutrients regarded as offering protection against chronic disease.

Table 3 shows the sociodemographic characteristics of participants by dietary cluster. Almost two-thirds of subjects in the Prudent Diet cluster were women, whereas virtually all members of the smaller Alcohol & Convenience Foods cluster were men. In the latter cluster, a high proportion of participants were single and living alone. There was evidence of social class variation across the clusters, with a lower proportion of participants in the Traditional Diet cluster in SES groups I and II. No other significant differences in sociodemographic characteristics by cluster were observed.

Smoking status varied significantly by dietary pattern, with fewer current smokers in the Prudent Diet cluster than in either the Traditional Diet or the Alcohol & Convenience Foods cluster (Table 4). There was also significant variation in levels of physical activity by cluster, with lowest levels in the Traditional Diet and highest levels in the Alcohol & Convenience Foods cluster (Table 4).

No significant differences in the prevalence of glucose intolerance or in the CVD risk factor profile were observed by cluster, apart from evidence of greater obesity in the Alcohol & Convenience Foods cluster (Table 5). A higher proportion of participants in the Traditional Diet group were at high CHD risk (15.7%) as compared with the

Table 2 Mean daily nutrient intakes by cluster, $n = 851$

Nutrient	Cluster 1 – Traditional Diet ($n = 480$)	Cluster 2 – Prudent Diet ($n = 340$)	Cluster 3 – Other [Alcohol & Convenience Foods] ($n = 31$)	P-value*	P-value†
Energy (kcal day^{-1})	2196	2088	2479	<0.01	<0.01
Fat (g day^{-1})	77.6	62.6	69.8	<0.001	<0.001
MUFA (g day^{-1})	25	21	23	<0.001	0.01
PUFA (g day^{-1})‡	10	11	9	0.03	<0.001
SFA (g day^{-1})‡	31	19	26	<0.001	<0.001
P:S ratio‡	0.3	0.6	0.3	<0.001	<0.001
Cholesterol (mg day^{-1})	348	237	352	<0.001	<0.001
Ethanol (g day^{-1})§	0.7	1.4	54.4	<0.001¶	0.03¶
Protein (g day^{-1})	97	99	111	0.05	0.35
Carbohydrate (g day^{-1})	274	278	253	0.38	0.43
Total sugars (g day^{-1})‡	106	110	100	0.30	0.16
Starch (g day^{-1})	155	157	145	0.62	0.67
Fibre (Southgate) (g day^{-1})	21	25	20	<0.001	<0.001
Retinol (μg day^{-1})	608	396	530	<0.01	0.01
Vitamin C (mg day^{-1})‡	74	105	69	<0.001	0.001
Vitamin D (μg day^{-1})‡	2.9	3.7	3.7	<0.001	0.001
Vitamin E (mg day^{-1})‡	5.0	5.8	4.7	<0.001	0.001
Vitamin B$_6$ (mg day^{-1})	2.6	2.8	3.7	<0.001	<0.001
Vitamin B$_{12}$ (μg day^{-1})‡	5.4	4.4	5.9	<0.001	0.001
Folate (μg day^{-1})	270	318	339	<0.001	<0.001

MUFA – monounsaturated fatty acids; PUFA – polyunsaturated fatty acids; SFA – saturated fatty acids; P:S – ratio of polyunsaturated to saturated fat.
* Analysis of variance was used to compare means by cluster.
† Analysis of variance was used to compare means between clusters 1 and 2.
‡ Geometric mean.
§ Median is given instead of mean.
¶ Kruskal–Wallis test.

Table 3 Sociodemographic characteristics by cluster, $n = 851$*

	Cluster 1 – Traditional Diet ($n = 480$)	Cluster 2 – Prudent Diet ($n = 340$)	Cluster 3 – Other [Alcohol & Convenience Foods] ($n = 31$)	P-value†
Mean age (years)	59	60	60	0.31
Men (%)	54.2	37.6	96.8	<0.001
SES categories I + II (%)	19.7	26.2	33.3	0.03
Third-level education (%)	27.3	27.4	32.3	0.83
Single (%)	11.5	10.7	19.4	0.34
Living alone (%)	13.4	12.8	25.8	0.02
Car ownership (%)	84.3	86.7	77.4	0.30
House ownership (%)	89.1	89.7	90.3	0.52

SES – socio-economic status.
* n varies because of missing data.
† A chi-square test was used to compare proportions and the Kruskal signed rank test was used for age.

Table 4 Dietary cluster by lifestyle characteristics, $n = 851$*

	Cluster 1 – Traditional Diet ($n = 480$)	Cluster 2 – Prudent Diet ($n = 340$)	Cluster 3 – Other [Alcohol & Convenience Foods] ($n = 31$)	P-value†
Smoking status (%)				
Never	47.6	48.3	29.0	0.04
Ex-smoker	32.3	37.0	38.7	
Current	20.1	14.7	32.3	
Physical activity (%)				
Low	47.6	37.9	26.9	<0.02
Medium	40.3	42.0	42.3	
High	12.1	20.2	30.8	

* n varies because of missing data.
† Chi-square test.

Table 5 Glucose intolerance and CVD risk by cluster, $n = 851$*

	Cluster 1 – Traditional Diet ($n = 480$)	Cluster 2 – Prudent Diet ($n = 340$)	Cluster 3 – Other [Alcohol & Convenience Foods] ($n = 31$)	*P*-value†
Glucose intolerance (%)	6.9	6.5	3.4	0.75
Hypertension (%)	40.5	38.9	48.4	0.57
High homocysteine (%)	1.3	1.8	–	0.10
Overall obesity, BMI $> 30\,\mathrm{kg\,m}^{-2}$ (%)	23.1	27.2	25.8	0.33
Morbid obesity, BMI $> 35\,\mathrm{kg\,m}^{-2}$ (%)	4.0	6.2	3.2	0.30
High WC (%)	37.1	45.0	48.4	0.06
High WHR (%)	74.3	68.2	90.3	0.01
Pre-existing CVD (%)	12.7	15.6	6.5	0.20§
High CHD risk‡	15.7	11.2	19.4	0.12

CVD – cardiovascular disease; BMI – body mass index; WC – waist circumference; WHR – waist/hip ratio; CHD – coronary heart disease.
* *n* varies because of missing data.
† A chi-square test was used to compare proportions.
‡ Framingham risk $> 20\%$ over 10 years.
§ Cluster 1 vs. Cluster 2.

Prudent Diet group (11.2%); a finding of borderline significance ($P = 0.06$).

Table 6 shows differences in means, adjusted for age and sex, of continuous CVD risk factors by dietary cluster. Participants in the Traditional Diet cluster had the lowest BMI and WC, but the highest concentrations of serum homocysteine. BMI, WC, VLDL-cholesterol and triglyceride levels were highest in the Alcohol & Convenience Foods cluster. The findings from analyses comparing the two major clusters were broadly similar, although the differences in VLDL-cholesterol levels were no longer significant.

Discussion

In this cross-sectional study of middle-aged men and women we identified three dietary patterns by cluster analysis based on food groups: traditional and prudent diet patterns and a smaller group with high intakes of alcohol and convenience foods. These dietary patterns were associated with distinct nutrient intake profiles of potential biological significance. In particular, participants in the Prudent Diet group had a favourable nutrient profile relative to the Traditional Diet group, with higher intakes of polyunsaturated fat, antioxidant vitamins and fibre and, VLDL-cholesterolower intakes of saturated fat. There were also significant differences in the gender, socio-economic status and behaviour profiles of participants in the three dietary groups. Participants in the Prudent Diet group were predominantly female, drawn from higher socio-economic groups with relatively high levels of physical activity and a low prevalence of smoking. Virtually all participants in the Alcohol & Convenience Foods group were male, of whom a high proportion were living alone.

Table 6 Age- and sex-adjusted means of CVD risk factors by cluster, $n = 851$*

	Cluster 1 – Traditional Diet ($n = 480$)	Cluster 2 – Prudent Diet ($n = 340$)	Cluster 3 – Other [Alcohol & Convenience Foods] ($n = 31$)	*P*-value†	*P*-value‡
BMI ($\mathrm{kg\,m}^{-2}$)	27.27	28.05	28.11	0.03	0.01
WC (cm)	92.61	94.52	97.62	0.01	0.02
WHR	0.91	0.92	0.93	0.50	0.37
Total cholesterol ($\mathrm{mmol\,l}^{-1}$)	5.87	5.82	6.08	0.38	0.55
HDL-cholesterol ($\mathrm{mmol\,l}^{-1}$)	1.52	1.51	1.65	0.22	0.89
LDL-cholesterol ($\mathrm{mmol\,l}^{-1}$)	3.70	3.61	3.60	0.43	0.28
VLDL-cholesterol ($\mathrm{mmol\,l}^{-1}$)§	0.58	0.61	0.73	0.02	0.09
Triglycerides ($\mathrm{mmol\,l}^{-1}$)§	1.27	1.38	1.57	0.02	0.05
Glucose ($\mathrm{mmol\,l}^{-1}$)§	4.90	4.85	4.76	0.59	0.53
HbA1c (%)	5.05	5.01	4.92	0.51	0.52
Homocysteine ($\mathrm{mmol\,l}^{-1}$)§	11.1	10.3	10.2	< 0.01	< 0.01
SBP (mmHg)	136.84	134.58	136.54	0.31	0.12
DBP (mmHg)	81.12	80.61	83.88	0.27	0.49

CVD – cardiovascular disease; BMI – body mass index; WC – waist circumference; WHR – waist/hip ratio; HDL – high-density lipoprotein; LDL – low-density lipoprotein; VLDL – very-low-density lipoprotein; HbA1c – glycosylated haemoglobin; SBP – systolic blood pressure; DBP – diastolic blood pressure.
* *n* varies because of missing data.
† Analysis of covariance was used to compare means by cluster.
‡ Analysis of covariance was used to compare means between clusters 1 and 2 only.
§ Geometric mean.

Other studies have reported similar findings[4,7,8,30]. Using K-means cluster analysis, Tucker *et al.*[4] defined four clusters in a study of elderly participants in Boston: (1) alcohol, (2) milk, cereals and fruit, (3) bread and poultry and (4) meat and potatoes. In contrast to the current study, they defined intake as the proportion of total energy contributed by each of the food groups. However, in common with this study, they identified a dietary pattern characterised by high alcohol consumption. In a UK population study of 1087 men and 1110 women aged 16–64 years, four different dietary patterns were identified using hierarchical cluster analysis. As in the current study, differences in nutrient, social and behavioural profiles among clusters were observed[17]. There is now consistent evidence that dietary patterns are related to other behaviours such as smoking and physical activity[4,7,8,30]. Thus, our data adds to the evidence that risk factors for chronic disease cluster within individuals.

Williams *et al.* found that a healthy dietary pattern (identified by principal components analysis) was associated with a more favourable CVD risk factor profile in the Isle of Ely study[7]. By contrast, we did not find that the prudent diet pattern was associated with a more favourable cardiovascular risk factor profile in the current study. Surprisingly, participants in the Prudent Diet group had a higher BMI, WC and higher triglyceride levels than those in the Traditional Diet group. Clearly the findings from cross-sectional associations between dietary patterns and CVD risk factors must be interpreted cautiously given the possibility of reverse causation. It should also be noted that the current study is based on a sample of middle-aged men and women aged 50–69 years (with a relatively homogeneous diet), as opposed to adults aged 40–65 years in the Isle of Ely study[7].

Both the Prudent Diet and the Alcohol & Convenience Foods groups had lower homocysteine levels than the Traditional Diet group. This is biologically plausible given the relatively high intakes of folate, vitamin B_6 and vitamin B_{12} (inversely related to homocysteine levels) in these groups. Similar findings have emerged from the Male Professionals Health Study[8].

In conclusion, we have identified two major dietary patterns: prudent and traditional, and a smaller group with high intakes of alcohol and convenience foods in a general population sample of middle-aged men and women. The patterns identified segregate individuals within the population into groups with significant differences in intake of nutrients that are related to major chronic diseases and differences in behavioural risk factor profiles. Consideration of the multidimensional aspects of diet and other related behaviours such as smoking and physical activity will facilitate work on the aetiology of chronic disease and the development of multidisciplinary behaviour change strategies.

References

1 Hu FB. Dietary pattern analysis: a new direction in nutritional epidemiology. *Current Opinion in Lipidology* 2002; **13**(1): 3–9.

2 Schwerin HS, Stanton JL, Smith JL, Riley AM Jr, Brett BE. Food, eating habits, and health: a further examination of the relationship between food eating patterns and nutritional health. *American Journal of Clinical Nutrition* 1982; **35**(Suppl. 5): 1319–25.

3 Randall E, Marshall JR, Graham S, Brasure J. Patterns in food use and their associations with nutrient intakes. *American Journal of Clinical Nutrition* 1990; **52**(4): 739–45.

4 Tucker KL, Dallal GE, Rush D. Dietary patterns of elderly Boston-area residents defined by cluster analysis. *Journal of the American Dietetic Association* 1992; **92**(12): 1487–91.

5 Barker ME, McClean SI, Strain JJ, Thompson KA. Dietary behaviour and health in Northern Ireland: an exploration of biochemical and haematological associations. *Journal of Epidemiology and Community Health* 1992; **46**(2): 151–6.

6 McCann SE, Marshall JR, Brasure JR, Graham S, Freudenheim JL. Analysis of patterns of food intake in nutritional epidemiology: food classification in principal components analysis and the subsequent impact on estimates for endometrial cancer. *Public Health Nutrition* 2001; **4**(5): 989–97.

7 Williams DE, Prevost AT, Whichelow MJ, Cox BD, Day NE, Wareham NJ. A cross-sectional study of dietary patterns with glucose intolerance and other features of the metabolic syndrome. *British Journal of Nutrition* 2000; **83**(3): 257–66.

8 Fung TT, Rimm EB, Spiegelman D, Rifai N, Tofler GH, Willett WC, *et al*. Association between dietary patterns and plasma biomarkers of obesity and cardiovascular disease risk. *American Journal of Clinical Nutrition* 2001; **73**(1): 61–7.

9 Hu FB, Rimm EB, Stampfer MJ, Ascherio A, Spiegelman D, Willett WC. Prospective study of major dietary patterns and risk of coronary heart disease in men. *American Journal of Clinical Nutrition* 2000; **72**(4): 912–21.

10 Schwerin HS, Stanton JL, Riley AM Jr, Schaefer AE, Leveille GA, Elliott JG, *et al*. Food eating patterns and health: a reexamination of the Ten-State and HANES I surveys. *American Journal of Clinical Nutrition* 1981; **34**(4): 568–80.

11 Akin JS, Guilkey DK, Popkin BM, Fanelli MT. Cluster analysis of food consumption patterns of older Americans. *Journal of the American Dietetic Association* 1986; **86**(5): 616–24.

12 Hulshof KF, Wedel M, Lowik MR, Kok FJ, Kistmaker C, Hermus RJ, *et al*. Clustering of dietary variables and other lifestyle factors (Dutch Nutritional Surveillance System). *Journal of Epidemiology and Community Health* 1992; **46**: 417–24.

13 Huijbregts PP, Feskens EJ, Kromhout D. Dietary patterns and cardiovascular risk factors in elderly men: the Zutphen Elderly Study. *International Journal of Epidemiology* 1995; **24**(2): 313–20.

14 Farchi G, Mariotti S, Menotti A, Seccareccia F, Torsello S, Fidanza F. Diet and 20-y mortality in two rural population groups of middle-aged men in Italy. *American Journal of Clinical Nutrition* 1989; **50**(5): 1095–103.

15 Millen BE, Quatromoni PA, Gagnon DR, Cupples LA, Franz MM, D'Agonstino RB. Dietary patterns of men and women suggest targets for health promotion: the Framingham Nutrition Studies. *American Journal of Health Promotion* 1996; **11**(1): 42–52, discussion 52–3.

16 Wirfalt AK, Jeffery RW. Using cluster analysis to examine dietary patterns: nutrient intakes, gender, and weight status differ across food pattern clusters. *Journal of the American Dietetic Association* 1997; **97**(3): 272–9.

17 Pryer JA, Nichols R, Elliott P, Thakrar B, Brunner E, Marmot M. Dietary patterns among a national random sample of

British adults. *Journal of Epidemiology and Community Health* 2001; **55**(1): 29–37.

18 Creagh D, Neilson S, Collins A, Colwell N, Hinchion R, Drew C, *et al*. Established cardiovascular disease and CVD risk factors in a primary care population of middle-aged Irish men and women. *Irish Medical Journal* 2002; **95**(10): 298–301.

19 World Health Organization (WHO). *Definition, Diagnosis and Classification of Diabetes Mellitus and Its Complications*. Geneva: WHO, 1999.

20 Ford ES, Giles WH, Dietz WH. Prevalence of the metabolic syndrome among US adults: findings from the third National Health and Nutrition Examination Survey. *Journal of the American Medical Association* 2002; **287**(3): 356–9.

21 American Dietetic Association. Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. *Diabetes Care* 1997; **20**(7): 1183–97.

22 Anderson KM, Wilson PW, Odell PM, Kannel WB. An updated coronary risk profile. A statement for health professionals. *Circulation* 1991; **83**(1): 356–62.

23 Bingham SA, Gill C, Welch A, Cassidy A, Runswick SA, Oakes S, *et al*. Validation of dietary assessment methods in the UK arm of EPIC using weighed records, and 24-hour urinary nitrogen and potassium and serum vitamin C and carotenoids as biomarkers. *International Journal of Epidemiology* 1997; **26**(Suppl. 1): S137–51.

24 Friel S, Gadhainn N. *The National Health and Lifestyle Surveys – Regional Results from the Survey of Lifestyle, Attitudes and Nutrition*. Dublin/Galway: Health Promotion Unit, Department of Health and Children/Centre for Health Promotion Studies, National University of Ireland, 1999.

25 Friel S, Kelleher CC, Nolan G, Harrington J. Social diversity of Irish adults nutritional intake. *European Journal of Clinical Nutrition* 2003; **57**(7): 865–75.

26 Harrington J. *Validation of a Food Frequency Questionnaire as a Tool for Assessing Nutrient Intake*. Galway: Department of Health Promotion, University College Galway, 1997.

27 Joshipura KJ, Hu FB, Manson JE, Stampfer MJ, Rimm EB, Speizer FE, *et al*. The effect of fruit and vegetable intake on risk for coronary heart disease. *Annals of Internal Medicine* 2001; **134**(12): 1106–14.

28 Holland B, Welch AA, Unwin ID, Buss DH, Paul AA, Southgate DAT. *McCance & Widdowson's The Composition of Foods*, 5th ed. Royal Society of Chemistry and Ministry of Agriculture, Fisheries and Food. London: Her Majesty's Stationery Office, 1997.

29 Gregory J, Foster K, Tyler H, Wiseman M. *The Dietary and Nutritional Survey of British Adults*. London: HMSO, 1990.

30 Margetts BM, Jackson AA. Interactions between people's diet and their smoking habits: the dietary and nutritional survey of British adults. *British Medical Journal* 1993; **307**(6916): 1381–4.

Appendix – Food groups used in cluster analysis

Alcoholic drinks

Drinks: all non alcoholic beverages with the exception of fruit juice

Pasta/rice

White bread: white bread and refined cereals

Brown bread: brown bread, brown soda bread, crispbread and unrefined cereals

Chips: chips and roast potatoes

Butter

Spreads: margarine and fat spreads

Salad dressings: salad cream and other dressings

High-fat dairy: whole milk and dairy products made with whole milk (e.g. cheddar cheese)

Low-fat dairy: skimmed milk, low-fat milk, low-fat yoghurt and low-fat cheese

Meat: beef, lamb, pork, bacon and ham

Poultry: chicken

Fish: fish/shellfish

Meat products: processed meat products

Eggs: eggs and quiche

Fruits: all fruits and fruit juice

Vegetables: all vegetables, salads, boiled and mashed potatoes

Desserts: biscuits, cakes, pastries and puddings

Sweets: chocolate, chocolate bars, sweets and sugar

Snacks: crisps and nuts

Miscellaneous: sauces, chutney, jams, etc.