

## Comparison of cluster and principal component analysis techniques to derive dietary patterns in Irish adults

Áine P. Hearty\* and Michael J. Gibney

*Institute of Food and Health, Agriculture and Food Science Centre, University College Dublin, Belfield, Dublin 4, Ireland*

*(Received 13 February 2008 – Revised 2 April 2008 – Accepted 20 May 2008 – First published online 25 June 2008)*

The aims of the present study were to examine and compare dietary patterns in adults using cluster and factor analyses and to examine the format of the dietary variables on the pattern solutions (i.e. expressed as grams/day (g/d) of each food group or as the percentage contribution to total energy intake). Food intake data were derived from the North/South Ireland Food Consumption Survey 1997–9, which was a randomised cross-sectional study of 7 d recorded food and nutrient intakes of a representative sample of 1379 Irish adults aged 18–64 years. Cluster analysis was performed using the *k*-means algorithm and principal component analysis (PCA) was used to extract dietary factors. Food data were reduced to thirty-three food groups. For cluster analysis, the most suitable format of the food-group variable was found to be the percentage contribution to energy intake, which produced six clusters: ‘Traditional Irish’; ‘Continental’; ‘Unhealthy foods’; ‘Light-meal foods & low-fat milk’; ‘Healthy foods’; ‘Wholemeal bread & desserts’. For PCA, food groups in the format of g/d were found to be the most suitable format, and this revealed four dietary patterns: ‘Unhealthy foods & high alcohol’; ‘Traditional Irish’; ‘Healthy foods’; ‘Sweet convenience foods & low alcohol’. In summary, cluster and PCA identified similar dietary patterns when presented with the same dataset. However, the two dietary pattern methods required a different format of the food-group variable, and the most appropriate format of the input variable should be considered in future studies.

### Dietary patterns: Cluster analysis: Principal component analysis

Traditionally, nutrition research has been focused on the detailed examination of nutrients and dietary components. However, in more recent years, public health nutrition has seen a marked move from research at the nutrient level to the food level, with the driving force of this movement attributed to the accepted concept that people eat foods not nutrients. This approach at the food level aims to facilitate a more holistic assessment of dietary behaviour, which therefore is more poised to provide tangible food-based dietary advice to the general public<sup>(1)</sup>. However, foods are consumed in many complex combinations and studies of individual foods can be difficult to interpret because of strong intercorrelations<sup>(2–4)</sup>. Another major issue with this type of research is that dietary patterns cannot be measured directly. However, two fundamentally different analytical approaches have been developed for measuring dietary patterns in epidemiological studies: *a priori* and *a posteriori* methods. *A priori* methods explore the data using predefined combinations of foods in a dietary index, such as the healthy eating index<sup>(5)</sup>. *A posteriori* methods explore the available data *post hoc* by either factor or cluster analysis and produce dietary patterns in which nutritional variables (i.e. foods) are reduced to a smaller number of variables<sup>(4)</sup>.

Factor analysis, specifically principal component analysis (PCA), is a frequently used exploratory approach to identify dietary patterns in a population<sup>(6,7)</sup>. New dietary and food

pattern variables are derived on the basis of the correlation matrix of the original food variables and individuals receive a factor score (principal components (PC) score) for each of the derived factors. In cluster analysis, dietary data are reduced to patterns based on individual differences in mean dietary intakes. Cluster analysis creates patterns that are mutually exclusive, as each subject in the present study can belong to only one cluster. Multiple clustering algorithms exist. Of these, *k*-means is probably the most popular non-hierarchical clustering method used in the literature.

Factor and cluster procedures have been compared in relation to their ability to predict disease risk in a few studies<sup>(8,9)</sup>, and they showed that the results from the dietary score and factor analysis techniques were comparable. Bamia *et al.*<sup>(10)</sup> and Crozier *et al.*<sup>(11)</sup> also compared cluster and PCA methods in relation to elderly Europeans and young British women, respectively. They both found close similarities between patterns derived using either method. However, Crozier *et al.*<sup>(11)</sup> found only two meaningful dietary patterns (labelled as ‘Prudent’ and ‘High-energy’), and they felt that the dichotomous variable produced by cluster analysis was less informative than the continuous variable produced by PCA. As there are very few studies that have compared cluster and PCA, and as these few studies have been in specific population groups, there still exists the need to directly examine these two methods in a more general representative sample.

**Abbreviations:** PC, principal component; PCA, principal component analysis; Q, quartile; %TE, percentage contribution to total energy intake.

\* **Corresponding author:** Á. P. Hearty, fax +353 17161147, email aine.hearty@ucd.ie

In the studies of dietary patterns, food variables should be selected so that the emerging patterns make sense from a dietary perspective. Various formats of the food-group variable can be utilised, and it has been advised that research is needed on how the treatment of dietary variables affects the dietary pattern solution<sup>(12)</sup>.

Therefore, the aims of the present study were to examine and compare the dietary patterns in a representative sample of Irish adults using both the cluster and factor analyses and to examine the impact of the format of the dietary variables on the pattern solutions (i.e. expressed as either g/d of each food group or as the percentage contribution to total energy intake (%TE) from each food group).

## Methods

### Study sample details

The North/South Ireland Food Consumption Survey was a randomised cross-sectional study of food and nutrient intakes of a representative sample of adults aged 18–64 years from the Republic of Ireland and Northern Ireland between 1997 and 1999<sup>(13,14)</sup>. The dietary survey was completed by 1379 adults (662 males and 717 females; response rate was 63%), and food intake was collected using a semi-quantitative 7 d food diary. Food and nutrient analysis was conducted using Weighed Intake Software Program (Tinuviel Software, Anglesey, UK). Weighed Intake Software Program uses data from McCance & Widdowson's 'The Composition of Foods'<sup>(15)</sup> plus supplemental volumes to generate nutrient intake data.

### Food groups

Food data were reduced to thirty-three food groups to ease the interpretation of cluster and factor components. For the amount of each food group consumed, only edible fraction weights of all foods were considered in the present analysis. Foods expressed as the dry-weight version were corrected to represent the amount as consumed. Two beverage food groups were created to differentiate between the types of beverages consumed in the studies: low energy (e.g. water, tea, coffee, sugar-free drinks, sugar-free squashes) and high energy (e.g. soft drinks, squashes). Whole milk, low-fat milk and fruit juices remained in separate food groups. For both cluster and PCA, food groups were expressed as either g/d intake or the %TE from each food group (Appendix table).

### Identification of dietary patterns

**Cluster analysis.** Cluster analysis was performed using the *k*-means algorithm. This is based on geometric similarity, which gives a measure of the Euclidean distance from each record to the cluster centre, and from each cluster to the others. A series of steps were taken to select the most suitable number of clusters for the analysis. First, several runs were conducted with a varying number of clusters. For each run, cluster proximities for each cluster centre were examined and the number of iterations per each cluster was increased to ensure minimum error in cluster membership and that the model had converged to a solution. Clusters were also run without outliers to help find the best cluster solutions.

Finally, the resulting clusters were examined for sensible patterns to establish the robustness of the clusters.

**Principal component analysis.** PCA was used to extract dietary factors on the basis of the correlation of the thirty-three food groups consumed. PC with eigenvalues of  $\geq 1.5$  were retained and the retained factors were orthogonally rotated by the varimax method, so that the factors were uncorrelated, making them easier to interpret. Each rotated PC was interpreted based on the food groups with loadings of  $\geq 0.25$  or  $\leq -0.25$ , which were considered as significantly contributing to a pattern<sup>(16)</sup>. Factor scores were also saved for each PC for each respondent. The scores represent standardised variables with mean 0 and standard deviation 1.

### Statistical analysis

SPSS<sup>®</sup> version 12 for Windows (SPSS<sup>®</sup> Inc., Chicago, IL, USA) was used for data manipulation and basic statistical analysis of the datasets. SPSS Clementine<sup>®</sup> version 9.0 (SPSS<sup>®</sup> Inc.) was used to conduct the data reduction analysis, and this software standardises all variables before analysis. Differences in the mean percentage contribution of each food group across clusters and the differences in the mean nutrient intake across clusters were evaluated using one-way ANOVA. One-way ANOVA was also used to test for significant differences in mean nutrients across quartiles of factor components. Where statistically different effects were encountered ( $P < 0.05$ ), comparisons of mean nutrient intakes were made using Scheffe's *post hoc* multiple comparison test. For values that did not comply with Levene's test for homogeneity of variance, the Tamhane *post hoc* multiple comparison test was used<sup>(17)</sup>. Mean PC scores were also calculated and computed for each of the six cluster solutions.

In order to determine whether the clusters and PC of the dietary patterns found in the present study were comparable, binary logistic regression analysis was used. A model was constructed to predict the odds of being in quartile 4 (Q4) of each PC (dependent variable) based on the membership of one of the six clusters (independent variable). A binary variable 'yes/no' for being in Q4 of each PC was generated. Confounding factors (i.e. sex, age, social class and smoking) were also adjusted for and the model was run for different scenarios based on these confounders. OR, comparing the outcome of the dependent variable with the reference value, and the corresponding 95% CI were calculated.

## Results

### Cluster analysis

Dietary patterns derived by cluster analysis using either format of the food-group variable were compared (Table 1). When the patterns were examined with the food-group variables expressed as g/d, five clusters were found to best represent dietary patterns. However, one of these clusters, representing 39% of the sample ( $n = 541$ ), did not have any dominating food groups, making it difficult to interpret this major dietary pattern. However, when the food groups were expressed as %TE, a six-cluster solution was found to best represent dietary patterns and these clusters were fully interpretable based on their contributing food groups, so therefore these clusters

**Table 1.** Comparison of the dietary patterns derived by cluster and principal component analysis (PCA) methods using two forms of the food-group variable, g/d and percentage contribution to daily energy intake

Format of the dietary variable	Cluster analysis*			PCA	
	n†	Foods consumed in high quantities	Label	Foods with high-factor loadings‡	Label
1. Mean food-group intake (g/d)	541	No foods consumed in high quantities	Moderate all food groups	High-energy beverages, chips, meat products, savoury snacks, sugars and preserves, sauces, alcoholic beverages, savouries	Unhealthy and high alcohol
	239	Breakfast cereals, low-fat milk, yogurts, vegetables, fruit juices, fruit, fish, low-energy beverages, soups	Healthy foods	Potatoes, red meat, confectionery, butter and spreads, white bread, wholemeal bread, whole milk, alcoholic beverages	Traditional Irish
	229	Rice and pasta, low-fat spreads, red-meat dishes, poultry dishes	Rice and pasta dishes	Vegetables, fruit, fish, fruit juices, sauces, wholemeal bread, yogurts, low-energy beverages	Healthy foods
	202	White bread, wholemeal bread, biscuits and cakes, ice cream and desserts, eggs, butter and spreads, potatoes, red meat, confectionery	Traditional Irish and sweet foods	Biscuits and cakes, ice cream and desserts, confectionery, breakfast cereals, yogurts	Sweet foods and breakfast cereal
	163	Savouries, cheese chips, poultry, meat products, alcoholic beverages, sugars and preserves, savoury snacks, sauces, high-energy beverages, spreads, potatoes, red meat, confectionery	Unhealthy foods		
2. Energy intake from food groups (%)	303	White bread, whole milk, eggs, butter and spreads, potatoes, red meat, confectionery	Traditional Irish	Chips, fruit juices, meat products, high-energy beverages	Unhealthy foods and fruit juice
	285	Rice and pasta, savouries, cheese, red-meat dishes, poultry dishes, alcoholic beverages	Continental	Fruit, yogurts, breakfast cereals, vegetables, eggs	Healthy foods
	245	Chips, fruit juices, meat products, sugars and preserves, savoury snacks, high-energy beverages	Unhealthy foods	Low-fat milk	Low-fat milk
	221	Wholemeal bread, biscuits and cakes, ice cream and desserts	Wholemeal bread and desserts	Savoury snacks, sugars and preserves, high-energy beverages	Snacks and high-energy beverages
	208	White bread, low-fat milk, cheese, sugars and preserves, soups	Light meals and low-fat milk		
	117	Wholemeal bread, breakfast cereals, yogurts, low-fat spreads, vegetables, fruit, fish, poultry, low-energy beverages	Healthy foods		

\* *k*-Means algorithm used for the cluster analysis.

† *n* is in relation to the frequency of subjects in clusters only; for solutions derived using the g/d format, there were five outliers, so the total sample size represented by these clusters was 1374.

‡ All principal components described here had eigen values > 1.5, and factor loadings for food groups were > 0.25.

were chosen for the rest of the analysis presented in the present paper. These clusters were labelled as 'Traditional Irish', 'Continental', 'Unhealthy foods', 'Light-meal foods & low-fat milk', 'Healthy foods' and 'Wholemeal bread & desserts'. The differences in %TE per food group per cluster are depicted in Table 2. The mean daily nutrient intakes were also compared across clusters (Table 3). A summary of the cluster profiles is as follows.

**Cluster 1: 'Traditional Irish'.** This cluster was the most prevalent cluster (22% of the sample). It is characterised by providing a relatively high %TE from white bread, whole milk, eggs, butter and spreads, potatoes, red meat and confectionery, which are foods consumed frequently as part of the Irish diet. This cluster was associated with having high-energy (relative to clusters 4 and 5) and added sugar intakes (relative to clusters 2, 4–6) and a high %TE from saturated fat (relative to clusters 2, 3–5).

**Cluster 2: 'Continental'.** This cluster is characterised by providing a relatively high %TE from rice and pasta, savouries, cheese, red-meat dishes, poultry dishes, alcoholic

beverages, savoury snacks and sauces and low intakes of potatoes and red meat. This cluster was associated with having a low-fibre intake (relative to clusters 5 and 6) and low %TE from carbohydrate (relative to clusters 1, 4–6).

**Cluster 3: 'Unhealthy foods'.** This cluster is characterised by providing a relatively high %TE from chips, fruit juices, meat products, sugars and preserves, savoury snacks and high-energy beverages, which, apart from fruit juices, are usually considered as 'Unhealthy' foods. This cluster was associated with having a high %TE from fat (relative to clusters 4 and 5) and a low %TE from protein (relative to all other clusters). It was also found to have low calcium (relative to cluster 4), low folate (relative to clusters 5 and 6) and vitamin C (relative to cluster 5).

**Cluster 4: 'Light-meal foods & low-fat milk'.** This cluster is characterised by providing a relatively high %TE from foods often associated with light meals, i.e. white bread, low-fat milk, cheese, sugars and preserves and soups. This cluster was associated with having a high intake of calcium (relative to clusters 1–3).

**Table 2.** The dietary profile of the six clusters as described by the percentage contribution of each food-group variable to total energy intake\*† (Mean values and standard deviations)

Food group	Cluster 1		Cluster 2		Cluster 3		Cluster 4		Cluster 5		Cluster 6	
	n 303		n 285		n 245		n 208		n 117		n 221	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Rice and pasta	0.57 <sup>a</sup>	1.16	2.64 <sup>b†</sup>	2.56	0.74 <sup>ac</sup>	1.42	1.10 <sup>cd</sup>	1.60	1.86 <sup>bd</sup>	2.55	0.52 <sup>a</sup>	1.04
Savouries	1.27 <sup>a</sup>	2.57	3.26 <sup>b†</sup>	4.12	3.06 <sup>b</sup>	4.19	2.66 <sup>b</sup>	3.43	1.50 <sup>a</sup>	2.54	1.45 <sup>a</sup>	2.93
White bread	15.42 <sup>a†</sup>	5.59	10.41 <sup>b</sup>	4.97	10.07 <sup>b</sup>	5.39	13.34 <sup>c†</sup>	5.97	6.23 <sup>d</sup>	5.45	6.56 <sup>d</sup>	4.31
Wholemeal bread	2.03 <sup>a</sup>	2.58	3.32 <sup>b</sup>	3.81	2.20 <sup>a</sup>	3.20	2.79 <sup>ab</sup>	3.05	10.33 <sup>c†</sup>	6.73	10.98 <sup>c†</sup>	6.18
Breakfast cereals	2.74 <sup>ac</sup>	3.04	3.12 <sup>a</sup>	3.19	2.11 <sup>c</sup>	2.75	4.85 <sup>b</sup>	3.63	8.38 <sup>d†</sup>	5.78	4.11 <sup>b</sup>	3.70
Whole milk	7.23 <sup>a†</sup>	5.48	4.60 <sup>bc</sup>	4.32	3.86 <sup>b</sup>	4.79	0.80 <sup>d</sup>	1.97	1.60 <sup>d</sup>	2.73	5.79 <sup>c</sup>	5.11
Low-fat milk	0.26 <sup>a</sup>	0.82	0.79 <sup>b</sup>	1.49	1.35 <sup>c</sup>	2.22	6.41 <sup>d†</sup>	3.00	4.89 <sup>e</sup>	3.53	0.89 <sup>bc</sup>	1.59
Cheese	1.69 <sup>a</sup>	2.02	2.34 <sup>b†</sup>	2.35	1.91 <sup>ab</sup>	2.42	2.28 <sup>ab†</sup>	2.31	1.81 <sup>ab</sup>	2.04	2.00 <sup>ab</sup>	2.19
Yogurts	0.39 <sup>ab</sup>	1.04	0.49 <sup>ac</sup>	0.97	0.22 <sup>b</sup>	0.61	0.73 <sup>c</sup>	1.21	2.62 <sup>d†</sup>	2.72	0.61 <sup>ac</sup>	1.07
Eggs	1.76 <sup>a†</sup>	1.92	1.25 <sup>b</sup>	1.46	1.34 <sup>ab</sup>	1.47	1.03 <sup>b</sup>	1.34	1.38 <sup>ab</sup>	2.35	1.31 <sup>b</sup>	1.43
Butter and spreads	7.20 <sup>a†</sup>	4.92	3.89 <sup>b</sup>	3.07	4.11 <sup>b</sup>	3.22	4.72 <sup>b</sup>	3.74	2.27 <sup>c</sup>	2.73	5.97 <sup>a</sup>	4.61
Low-fat spreads	0.67 <sup>a</sup>	1.97	0.67 <sup>a</sup>	1.45	0.57 <sup>a</sup>	1.39	0.79 <sup>a</sup>	1.73	1.53 <sup>b†</sup>	2.31	1.03 <sup>ab</sup>	2.06
Potatoes	7.94 <sup>a†</sup>	5.78	3.60 <sup>b</sup>	3.12	3.80 <sup>bc</sup>	2.83	4.57 <sup>cd</sup>	3.77	6.09 <sup>d</sup>	5.60	7.25 <sup>ad</sup>	5.84
Chips	4.60 <sup>a</sup>	3.10	4.82 <sup>a</sup>	3.06	14.48 <sup>b†</sup>	4.39	4.54 <sup>a</sup>	3.03	2.13 <sup>c</sup>	2.62	3.30 <sup>d</sup>	2.98
Vegetables	3.10 <sup>a</sup>	2.27	3.71 <sup>abc</sup>	3.40	3.25 <sup>ab</sup>	2.77	3.76 <sup>bc</sup>	2.55	4.81 <sup>c†</sup>	4.10	3.38 <sup>ab</sup>	2.68
Fruit juices	4.60 <sup>a</sup>	3.10	4.82 <sup>a</sup>	3.06	14.48 <sup>b†</sup>	4.39	4.54 <sup>a</sup>	3.03	2.13 <sup>c</sup>	2.62	3.30 <sup>d</sup>	2.98
Fruit	1.40 <sup>a</sup>	1.81	1.94 <sup>b</sup>	2.15	1.03 <sup>a</sup>	1.55	2.32 <sup>bd</sup>	2.24	6.75 <sup>c†</sup>	4.39	2.70 <sup>d</sup>	2.54
Fish	1.58 <sup>a</sup>	1.88	1.89 <sup>ab</sup>	2.26	1.92 <sup>ab</sup>	2.42	2.48 <sup>bd</sup>	2.48	4.22 <sup>†</sup>	4.08	2.66 <sup>d</sup>	2.86
Red meat	8.33 <sup>a†</sup>	5.50	4.26 <sup>b</sup>	3.40	4.88 <sup>b</sup>	3.85	4.97 <sup>b</sup>	3.72	5.13 <sup>bc</sup>	4.65	6.35 <sup>c</sup>	4.36
Poultry	1.40 <sup>a</sup>	1.73	1.50 <sup>ab</sup>	1.84	1.93 <sup>bc</sup>	2.12	1.84 <sup>abc</sup>	1.97	2.25 <sup>c†</sup>	2.07	1.78 <sup>abc</sup>	2.21
Red-meat dishes	1.39 <sup>a</sup>	1.98	5.67 <sup>b†</sup>	3.82	1.87 <sup>a</sup>	2.66	1.41 <sup>a</sup>	1.94	1.83 <sup>a</sup>	2.54	1.40 <sup>a</sup>	2.00
Poultry dishes	0.71 <sup>a</sup>	1.59	2.03 <sup>b†</sup>	2.71	1.46 <sup>b</sup>	2.74	1.40 <sup>bc</sup>	3.09	1.38 <sup>abc</sup>	3.03	0.74 <sup>ac</sup>	1.64
Meat products	4.90 <sup>a</sup>	4.15	4.69 <sup>a</sup>	4.31	7.71 <sup>b†</sup>	5.42	4.30 <sup>a</sup>	3.65	2.08 <sup>c</sup>	2.58	3.19 <sup>d</sup>	4.04
Alcoholic beverages	4.46 <sup>ad</sup>	6.08	8.09 <sup>b†</sup>	9.10	6.38 <sup>bc</sup>	7.59	4.54 <sup>cd</sup>	6.40	3.40 <sup>de</sup>	4.57	2.76 <sup>e</sup>	4.85
Confectionery	4.66 <sup>a†</sup>	4.36	2.09 <sup>b</sup>	2.42	1.49 <sup>c</sup>	2.01	1.64 <sup>bc</sup>	2.13	1.56 <sup>bc</sup>	2.13	3.06 <sup>d</sup>	2.99
Desserts	2.25 <sup>a</sup>	2.53	3.06 <sup>b</sup>	3.48	1.76 <sup>a</sup>	2.78	3.52 <sup>b</sup>	3.53	1.93 <sup>a</sup>	2.30	5.49 <sup>†</sup>	4.75
Biscuits and cakes	5.40 <sup>a</sup>	5.15	5.30 <sup>a</sup>	4.92	3.94 <sup>b</sup>	3.98	7.03 <sup>c</sup>	5.93	5.07 <sup>ab</sup>	4.52	8.05 <sup>†</sup>	5.85
Sugars and preserves	2.14 <sup>a</sup>	2.78	2.98 <sup>bc</sup>	3.94	3.74 <sup>b†</sup>	4.46	3.82 <sup>b†</sup>	4.53	1.60 <sup>a</sup>	2.49	2.21 <sup>ac</sup>	3.76
Savoury snacks	1.02 <sup>a</sup>	1.79	2.43 <sup>b†</sup>	3.60	2.45 <sup>b†</sup>	2.91	1.88 <sup>b</sup>	2.40	0.70 <sup>a</sup>	1.40	0.98 <sup>a</sup>	1.90
Sauces	1.11 <sup>a</sup>	1.36	1.83 <sup>b†</sup>	2.27	1.62 <sup>bc</sup>	1.86	1.14 <sup>a</sup>	1.38	1.36 <sup>abc</sup>	1.76	1.20 <sup>ac</sup>	1.52
Low-energy beverages	0.10 <sup>NS</sup>	0.55	0.09	0.36	0.05	0.22	0.13	0.47	0.22 <sup>†</sup>	0.78	0.18	1.08
High-energy beverages	1.23 <sup>a</sup>	2.10	1.61 <sup>a</sup>	2.13	3.57 <sup>b†</sup>	3.97	1.33 <sup>a</sup>	1.97	0.59 <sup>†</sup>	1.40	0.78 <sup>c</sup>	1.29
Soups	0.47 <sup>ac</sup>	0.88	0.78 <sup>bc</sup>	1.49	0.45 <sup>ac</sup>	0.88	0.97 <sup>b†</sup>	1.68	0.68 <sup>abc</sup>	1.06	0.50 <sup>ac</sup>	0.90

a,b,c,d,e Unlike superscript letters denote significant differences between clusters at  $P < 0.05$ .

\* Percentage intake of energy from each of the food groups.

† Cluster 1, 'Traditional Irish'; cluster 2, 'Continental'; cluster 3, 'Unhealthy foods'; cluster 4, 'Light-meal foods and low-fat milk'; cluster 5, 'Healthy foods'; cluster 6, 'Wholemeal bread and desserts'.

‡ Important food groups in the cluster.

**Table 3.** Comparison of daily nutrient intakes between the six clusters\* (Mean values and standard deviations)

	Total		Cluster 1		Cluster 2		Cluster 3		Cluster 4		Cluster 5		Cluster 6	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Energy (MJ)	9.26	3.07	9.87 <sup>a</sup>	3.47	9.27 <sup>ab</sup>	2.56	9.69 <sup>a</sup>	3.31	8.59 <sup>b</sup>	2.63	7.24 <sup>c</sup>	2.37	9.65 <sup>a</sup>	3.05
TE protein (%)	15.59	2.81	15.40 <sup>a</sup>	2.67	15.35 <sup>a</sup>	2.67	14.56 <sup>b</sup>	2.69	15.97 <sup>a</sup>	2.52	18.28 <sup>c</sup>	3.32	15.53 <sup>a</sup>	2.34
TE carbohydrate (%)	44.33	6.29	44.50 <sup>a</sup>	6.11	41.46 <sup>b</sup>	5.78	42.27 <sup>b</sup>	5.50	45.77 <sup>a</sup>	5.89	49.75 <sup>c</sup>	6.21	45.83 <sup>a</sup>	5.58
TE saturated fat (%)	13.70	3.21	14.62 <sup>a</sup>	3.22	13.83 <sup>bcd</sup>	3.01	13.58 <sup>bd</sup>	2.75	13.46 <sup>d</sup>	2.82	10.25 <sup>e</sup>	2.68	14.46 <sup>ab</sup>	3.30
TE total fat (%)	35.37	5.75	36.29 <sup>a</sup>	5.28	35.95 <sup>a</sup>	5.69	37.32 <sup>a</sup>	5.09	33.95 <sup>b</sup>	4.72	28.60 <sup>c</sup>	5.74	36.14 <sup>a</sup>	5.21
Added sugars (g)	55.06	37.17	65.02 <sup>a</sup>	42.97	53.59 <sup>b</sup>	31.85	60.66 <sup>ab</sup>	43.28	51.29 <sup>b</sup>	33.25	29.15 <sup>c</sup>	18.54	54.35 <sup>b</sup>	31.10
Souffigate fibre (g)	21.60	8.21	20.78 <sup>a</sup>	7.83	19.67 <sup>a</sup>	6.57	20.07 <sup>a</sup>	7.12	20.55 <sup>a</sup>	7.52	25.45 <sup>b</sup>	8.81	25.87 <sup>b</sup>	9.80
Calcium (mg)	841.84	343.97	843.28 <sup>a</sup>	344.07	801.62 <sup>a</sup>	283.46	768.35 <sup>a</sup>	353.40	945.69 <sup>b</sup>	347.54	849.97 <sup>ab</sup>	356.25	871.18 <sup>ab</sup>	369.54
Iron (mg)	14.33	14.02	13.98 <sup>NS</sup>	15.92	13.79	9.92	13.57	14.50	13.99	11.86	16.83	23.47	15.35	9.76
Zinc (mg)	9.97	4.96	10.18 <sup>NS</sup>	4.54	10.30	6.43	8.93	4.28	9.65	4.57	9.70	4.67	10.86	4.32
Folate (µg)	295.50	141.54	288.88 <sup>ab</sup>	130.33	293.77 <sup>ab</sup>	132.60	267.29 <sup>a</sup>	122.82	300.41 <sup>ab</sup>	128.32	336.52 <sup>b</sup>	225.12	311.76 <sup>b</sup>	135.61
Vitamin C (mg)	112.31	203.20	98.47 <sup>ab</sup>	165.55	117.88 <sup>ab</sup>	284.31	87.86 <sup>a</sup>	133.30	114.79 <sup>ab</sup>	163.32	186.72 <sup>b</sup>	306.03	109.46 <sup>ab</sup>	136.18

TE, total energy contribution.

a,b,c,d,e Unlike superscript letters denote significant differences between clusters at  $P < 0.05$ .

\*Cluster 1, 'Traditional Irish'; cluster 2, 'Continental'; cluster 3, 'Unhealthy foods'; cluster 4, 'Light-meal foods and low-fat milk'; cluster 5, 'Healthy foods'; cluster 6, 'Wholemeal bread and desserts'.

*Cluster 5: 'Healthy foods'.* This was the smallest cluster (8.5% of the population). It is characterised by providing a relatively high %TE from wholemeal bread, breakfast cereals, yogurts, low-fat spreads, vegetables, fruit, fish, poultry and low-energy beverages. These foods are generally associated with providing a healthy diet. This cluster had the lowest-energy intake, the lowest %TE from total fat and saturated fat, and the lowest intake of added sugars, while it also had the highest %TE from protein and carbohydrate than the other clusters. It also had a high intake of folate (relative to clusters 3 and 6) and a high-vitamin C intake (relative to cluster 3).

*Cluster 6: 'Wholemeal bread & desserts'.* This cluster is characterised by providing a relatively high %TE from wholemeal bread, biscuits and cakes, and ice cream and desserts. This cluster had a high-fibre intake (relative to clusters 1–4).

*Principal component analysis*

When the format of the food-group variable was expressed as %TE, the dietary patterns derived using PCA were difficult to interpret as only a few foods were found to have high-factor loadings (i.e. >0.25) per PC (Table 1). However, when the food-group variable was expressed as g/d, many food groups had high-factor loadings, thus making the patterns easier to interpret and label. Therefore, the food groups based on the g/d format are used for the rest of the analysis presented in the present paper. For the g/d format, the extraction of eigenvalues >1.5 produced four dietary patterns, which explained 28.5% of the total variance. These were labelled as 'Unhealthy foods & high alcohol', 'Traditional Irish', 'Healthy foods' and 'Sweet convenience foods & low alcohol' (Table 4). For each factor, quartiles of the total component weights were calculated and compared across nutrient intakes, and the highest quartile (Q4) for each is presented in Table 5. The profiles of each PC are as follows.

*PC 1: 'Unhealthy foods & high alcohol'.* This pattern was characterised by high loadings for high-energy beverages, chips, meat products, savoury snacks, sugars and preserves, sauces, alcoholic beverages and savouries, and negative loadings for low-energy beverages and breakfast cereals. In relation to Q4, this PC was found to be high in energy (relative to PC 3 and 4), have a high %TE from total fat (relative to all other PC) and saturated fat (relative to PC 3), have the highest intake of added sugars and have the lowest %TE from protein and carbohydrate. It was also found to be low in zinc (relative to PC 2) and have the lowest vitamin C of all PC.

*PC 2: 'Traditional Irish'.* This pattern was characterised by high loadings for potatoes, red meat, confectionery, butter and spreads, white bread, wholemeal bread, whole milk, alcoholic beverages, and by negative loadings for savoury snacks, sugars and preserves, fruit juices, and rice and pasta. In relation to Q4, this PC was also found to be high in energy (relative to PC 3 and 4), calcium (relative to PC 4) and folate (relative to PC 1 and 4), and it had the highest zinc intake relative to the other PC.

*PC 3: 'Healthy foods'.* This pattern was characterised by high loadings for vegetables, fruit, fish, fruit juices, sauces, yogurts, wholemeal bread and low-energy beverages. In relation to Q4, this PC was found to have a high %TE from carbohydrate (relative to PC 1 and 4) and a lower %TE

**Table 4.** Loading weights from each food group per extracted principal component (PC)\*

Food group	PC††			
	PC 1	PC 2	PC 3	PC 4
High-energy beverages	0.73	–	–	–
Chips and processed potatoes	0.68	–	–	–
Meat products	0.63	–	–	–
Savoury snacks including nuts	0.52	–0.31	–	–
Sugars and preserves	0.47	–	–	0.31
Alcoholic beverages	0.33	0.26	–	–0.45
Savories	0.33	–	–	–
Sauces and salad dressings	0.30	–	0.38	–
Low-energy beverages	–0.27	–	0.28	–
Potatoes	–	0.76	–	–
Red meat	–	0.73	–	–
Confectionery	–	0.37	–	0.34
Butter and spreads	–	0.37	–	–
White bread and rolls	–	0.32	–	–
Wholemeal bread and rolls	–	0.25	0.29	–
Fruit juices	–	–0.26	0.44	–
Rice and pasta	–	–0.26	–	–
Vegetables	–	–	0.61	–
Fruit	–	–	0.60	–
Fish, fish dishes and products	–	–	0.55	–
Yogurts	–	–	0.33	0.26
Biscuits, cakes and pastries	–	–	–	0.72
Cream, ice cream and desserts	–	–	–	0.57
Breakfast cereals	–	–	–	0.34
Whole milk	–	–	–	–
Red-meat dishes	–	–	–	–
Low-fat milk	–	–	–	–
Poultry	–	–	–	–
Variance explained (%)	9.1	8.9	6.0	4.5

\* Extraction method: principal component analysis using varimax rotation.

† Total sample components: PC 1, 'Unhealthy foods and high alcohol'; PC 2, 'Traditional Irish'; PC 3, 'Healthy foods'; PC 4, 'Sweet foods and breakfast cereal'.

‡ Factor loadings are only displayed for values  $\leq -0.25$  or  $\geq 0.25$ , some food groups are excluded as they did not load onto any factor retained.

from saturated fat than other PC. It was also low in added sugars (relative to PC 1 and 2) but high in Southgate fibre (relative to factors 1 and 4), iron (relative to PC 4) and vitamin C (relative to PC 1 and 2).

**Table 5.** Daily nutrient intakes compared across highest quartile (Q4) of each principal component (PC)\* (Mean values and standard deviations)

	Q4 (n 345)							
	PC 1		PC 2		PC 3		PC 4	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Energy MJ	11.625 <sup>a</sup>	3.12	11.671 <sup>a</sup>	3.10	10.0132 <sup>b</sup>	3.15	9.56034 <sup>b</sup>	2.97
TE protein (%)	14.06 <sup>a</sup>	2.33	15.78 <sup>b</sup>	2.50	16.06 <sup>b</sup>	3.10	16.17 <sup>b</sup>	2.70
TE carbohydrate (%)	42.58 <sup>a</sup>	5.84	44.04 <sup>bc</sup>	7.19	44.62 <sup>b</sup>	6.61	43.37 <sup>ac</sup>	5.66
TE total fat (%)	36.70 <sup>a</sup>	4.87	34.54 <sup>b</sup>	5.94	34.20 <sup>b</sup>	6.11	35.36 <sup>b</sup>	5.55
TE saturated fat (%)	14.02 <sup>a</sup>	2.71	13.64 <sup>a</sup>	3.32	12.82 <sup>b</sup>	3.32	13.57 <sup>a</sup>	3.07
Added sugars (g)	83.59 <sup>a</sup>	44.33	69.42 <sup>b</sup>	39.73	50.27 <sup>c</sup>	35.43	51.71 <sup>c</sup>	34.85
Southgate fibre (g)	23.48 <sup>a</sup>	8.78	26.25 <sup>b</sup>	9.14	26.78 <sup>b</sup>	8.84	22.02 <sup>a</sup>	8.26
Calcium (mg)	953.05 <sup>a</sup>	372.30	992.06 <sup>a</sup>	348.14	947.52 <sup>a</sup>	351.25	857.21 <sup>b</sup>	335.15
Iron (mg)	14.83 <sup>ab</sup>	10.67	15.47 <sup>a</sup>	5.72	16.37 <sup>a</sup>	14.97	13.74 <sup>b</sup>	6.23
Zinc (mg)	10.66 <sup>a</sup>	4.62	12.84 <sup>b</sup>	4.30	11.11 <sup>a</sup>	4.74	10.93 <sup>a</sup>	6.65
Folate (µg)	316.33 <sup>a</sup>	132.40	357.57 <sup>b</sup>	121.36	355.62 <sup>b</sup>	169.95	68.00 <sup>a</sup>	138.68
Vitamin C (mg)	94.59 <sup>a</sup>	97.85	100.72 <sup>ac</sup>	123.17	173.57 <sup>b</sup>	242.28	128.61 <sup>bc</sup>	210.09

TE, total energy contribution.

<sup>a,b,c</sup> Unlike superscript letters denote significant differences across quartiles at  $P < 0.05$ .

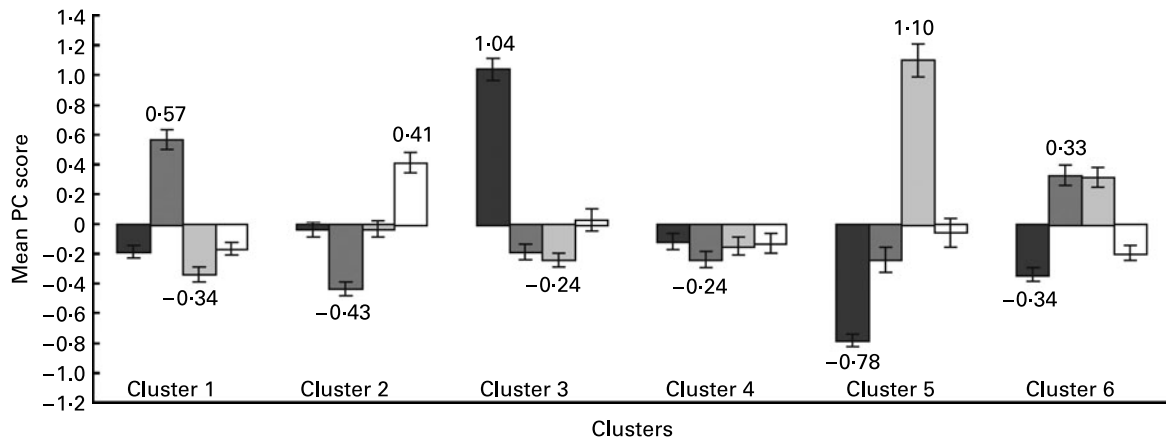
\* PC 1, 'Unhealthy foods and high alcohol'; PC 2, 'Traditional Irish'; PC 3, 'Healthy foods'; PC 4, 'Sweet foods and breakfast cereal'.

PC 4: 'Sweet foods & breakfast cereal'. This pattern was characterised by high loadings for biscuits and cakes, ice cream and desserts, confectionery, breakfast cereals, yogurts, and by a negative loading for alcoholic beverages. This PC had a low-energy intake (relative to PC 1 and 2) and fibre intake (relative to PC 2 and 3), low iron (relative to PC 2 and 3), zinc (relative to PC 2) and folate intake (relative to PC 2 and 3), and it had also had the lowest calcium intake.

#### Comparison of clusters and principal components

Mean PC scores were calculated and computed across the six cluster solutions, and this is illustrated in Fig. 1. The most striking features of this are in relation to clusters 1, 3 and 5. Cluster 1 scored the highest for PC 2 and the lowest for PC 3, indicating that close similarities exist between the dietary pattern 'Traditional Irish' derived by both methods, and that it is most different from the 'Healthy' PC. Cluster 3 scored the highest for PC 1, indicating also that for the 'Unhealthy' pattern, both cluster and PCA derived very similar patterns. Cluster 5 scored the highest for PC 3 and the lowest for PC 1, again illustrating the close similarities for the 'Healthy' pattern (and also how dissimilar this pattern is for the 'Unhealthy' PC). Cluster 6 also shares many similarities with the 'Healthy' and 'Traditional Irish' PC, due to the intake of wholemeal bread in both PC. In relation to the other two clusters, mean PC scores were less striking, but for cluster 4, it appears that none of the factors had a positive score for this, indicating that this cluster explains a dietary pattern not revealed by factor analysis.

In order to compare the cluster solutions and PCA, and to be able to quantify the relationship between them, a logistic regression model was run to predict Q4 of each principal component from each cluster with cluster 1 ('Traditional Irish') as the reference category (Table 6). The model was also adjusted for sex, age group, social class and smoking status, which were established as confounding factors. In relation to PC 1 ('Unhealthy foods & high alcohol'), membership of cluster 3 ('Unhealthy foods') had the highest OR for predicting this



**Fig. 1.** Principal component (PC) score compared across the six clusters of dietary patterns. PC 1 (■), 'Unhealthy foods and high alcohol'; PC 2 (▒), 'Traditional Irish'; PC 3 (□), 'Healthy foods'; PC 4 (□), 'Sweet foods and breakfast cereal'. Cluster 1, 'Traditional Irish'; cluster 2, 'Continental'; cluster 3, 'Unhealthy foods'; cluster 4, 'Light-meal foods and low-fat milk'; cluster 5, 'Healthy foods'; cluster 6, 'Wholemeal bread and desserts'.

(i.e. OR = 11.1). In relation to PC 2 ('Traditional Irish'), all other clusters had lower odds of predicting it relative to the reference cluster. In relation to PC 3 ('Healthy foods'), membership of cluster 5 ('Healthy foods') had the highest OR for predicting it (i.e. OR = 13.1). In relation to PC 4 ('Sweet foods & breakfast cereal'), membership of cluster 2 ('Continental') had the highest OR for predicting it (i.e. OR = 3.9).

## Discussion

An insight into the patterns of food intake may contribute to the successful implementation of dietary changes. Although there is a wealth of literature in the area of dietary patterns, there exists few published studies that have directly compared these methods using the same study sample<sup>(10,11,18)</sup>. This is despite the fact that, when conducting dietary pattern analysis, probably the most important issue is choosing the most appropriate pattern analysis technique. There is also the need to clarify the issue of the appropriate format of the food-group variable, as different formats may impact on the dietary patterns derived, and thus make it difficult to compare results across different studies.

The most popular unsupervised or *a posteriori* methods of data reduction are cluster and PCA. Both of these methods are statistically quite different from each other. Whereas cluster analysis separates persons into mutually exclusive groups based on the differences in food intakes, factor analysis separates foods into groups based on correlations between foods and persons receiving a score for each of the derived factors. From the present study, it was found that both methods derived quite similar and directly comparable dietary patterns for three of the dietary patterns (i.e. 'Traditional Irish', 'Healthy' and 'Unhealthy'). However, for cluster 4, there appeared to be no relationship with any of the patterns derived by PCA, based on the mean PCA score. This implies that PCA may not reveal all of the dietary patterns actually present in the dataset. This could be due to the subjective decisions applied in the present study (i.e. the eigenvalue cut-off), which resulted in choosing only the four PC presented here. Also, it has to be remembered that the four PC retained for further examination in the present study account for only 28.5% of the total variance, therefore the possibility remains that other dietary

patterns exist in the data. Schulze *et al.*<sup>(19)</sup> investigated the effect of PCA and variation in foods and nutrients associated with the dietary patterns in a sample of the European Prospective Investigation into Cancer and Nutrition-Potsdam cohort. They concluded that PC may explain food and nutrient intake quite differently, and therefore in some cases PCA may not uncover all dietary patterns in the dataset.

When the dietary patterns identified in the present study were compared using logistic regression analysis in order to quantify the relationships between them, it was also found that the 'Unhealthy foods', 'Healthy foods' and the 'Traditional Irish' clusters predicted the membership of similar patterns identified through PCA. Clusters 4 and 6 had OR of 2.1 and 2.9, respectively, for predicting the membership of PC 3 ('Healthy' pattern), which indicates that these clusters have properties similar to the 'Healthy' pattern, and that in the case for cluster 4, these were not picked up when the mean score alone was used. Cluster 2 had an OR of 3.9 for predicting the membership for PC 4, which supports the finding when the mean score was used. These results are similar to those found in a study by Costacou *et al.*<sup>(18)</sup>, who compared cluster analysis, PCA and a Mediterranean diet score in Greek adults, and they found that the Mediterranean dietary score closely predicted a Mediterranean dietary pattern derived by PCA.

Dietary patterns are inherently complex. Without more detailed analyses, they do not enable the specific identification of the particular dietary component within the pattern that may be responsible for the observed differences between population subgroups. Therefore, along with identifying the main patterns in the population sample, it is necessary to understand what these patterns mean, i.e. what foods are they high/low in, what nutrients are they high/low in and also to profile the people who consume them. The decision guiding the choice of dietary pattern technique should be construed according to reasons such as expertise of the research group and, perhaps most importantly, according to the format of the output required. Most published studies that have used either of these two methods have not provided a clear rationale as to their choice of one of these methods over another. When choosing between using either cluster or factor analysis, it should be appreciated that they each approach the data from

**Table 6.** Prediction of highest quartile (Q4) of each factor from each of the six cluster solutions using binary logistic regression (Odds ratios and 95% confidence intervals)

	Q4 PC 1*			Q4 PC 2			Q4 PC 3			Q4 PC 4		
	P	OR	95% CI	P	OR	95% CI	P	OR	95% CI	P	OR	95% CI
<b>Unadjusted model</b>												
Cluster 1 (reference)†	0.000	1.00		0.000	1.00		0.000	1.00		0.000	1.00	
Cluster 2	0.026	1.59	1.06, 2.40	0.000	0.16	0.11, 0.25	0.001	2.00	1.31, 3.07	0.000	4.05	2.76, 5.95
Cluster 3	0.000	10.12	6.75, 15.16	0.000	0.22	0.15, 0.33	0.297	1.28	0.80, 2.05	0.005	1.83	1.21, 2.78
Cluster 4	0.449	1.20	0.75, 1.90	0.000	0.22	0.14, 0.34	0.008	1.87	1.17, 2.96	0.303	1.27	0.80, 2.01
Cluster 5	0.002	0.04	0.01, 0.33	0.000	0.23	0.13, 0.40	0.000	11.85	7.17, 19.58	0.101	1.56	0.92, 2.64
Cluster 6	0.206	0.72	0.44, 1.20	0.085	0.73	0.51, 1.04	0.000	3.09	2.00, 4.76	0.656	1.11	0.70, 1.76
<b>Adjusted model‡</b>												
Cluster 1 (reference)†	0.000	1.00		0.000	1.00		0.000	1.00		0.000	1.00	
Cluster 2	0.179	1.40	0.86, 2.29	0.000	0.18	0.11, 0.29	0.002	2.09	1.32, 3.32	0.000	3.88	2.58, 5.83
Cluster 3	0.000	11.08	6.75, 18.20	0.000	0.22	0.13, 0.35	0.188	1.41	0.85, 2.35	0.011	1.78	1.14, 2.77
Cluster 4	0.137	1.52	0.88, 2.64	0.000	0.30	0.18, 0.49	0.003	2.09	1.28, 3.41	0.449	1.20	0.74, 1.95
Cluster 5	0.017	0.09	0.01, 0.65	0.005	0.41	0.22, 0.77	0.000	13.10	7.51, 22.84	0.145	1.53	0.86, 2.69
Cluster 6	0.585	1.18	0.66, 2.12	0.769	0.94	0.61, 1.44	0.000	2.90	1.82, 4.62	0.720	1.09	0.67, 1.77

PC, principal component.

\*PC 1, 'Unhealthy foods and high alcohol'; PC 2, 'Traditional Irish'; PC 3, 'Healthy foods'; PC 4, 'Sweet foods and breakfast cereal'.

†Cluster 1, 'Traditional Irish'; cluster 2, 'Continental'; cluster 3, 'Unhealthy foods'; cluster 4, 'Light-meal foods and low-fat milk'; cluster 5, 'Healthy foods'; cluster 6, 'Wholemeal bread and desserts'.

‡Model is adjusted for sex, age group, social class and smoking.

a different prospective, and thus answer different questions. Cluster analysis examines whether or not there are groups in the population that are distinctly different from one another, and if so, what typifies their diets? Factor analysis explores whether there are underlying patterns that explain variation in how people eat<sup>(20)</sup>. Factors identified also do not refer to identifiable groups within the population, and hence do not give an indication of the prevalence of a particular type of diet<sup>(21)</sup>. However, the actual procedure for PCA is more straightforward and logical as cluster analysis places considerable burden on the user in terms of selecting and deciding on the appropriate number of clusters.

Aside from uncovering the dietary patterns in the present study sample, the other aim was to explore the methodological analysis of dietary patterns using different formats of the food variable as unfortunately there is no gold standard technique for this. The comparative evaluations of patterns derived from the two variable formats described in the present paper produced slightly different dietary patterns, and in this case it was found that the food-group variables expressed as %TE produced the most interpretable cluster solution, while food groups expressed in g/d produced the most interpretable principal components. In a recent study by Bailey *et al.*<sup>(12)</sup>, they reported that the most appropriate variable for providing interpretable and relevant cluster solutions in their study was the number of servings from food subgroups. Therefore, it should be advised that for all dietary patterning methods, a clear rationale for the format of the dietary variable should be established before analysis in order to identify the most relevant pattern solutions in a particular dataset.

The most beneficial information that can be gathered from the dietary pattern analyses is on what foods in combination are culturally acceptable to the population. Hypothetical 'ideal' diets, despite their advantageous health implications to a population, are relatively useless unless they can be incorporated into the culture of the society. It appears from the present study that there are three main dietary patterns that dominate irrespective of the analytical methods used. These are an 'Unhealthy', a 'Traditional Irish' and a 'Healthy' pattern. This coincides with a recent paper which found, from a comprehensive literature review of articles using factor and cluster analyses, that 'Healthy', 'Traditional' and 'Sweets' patterns are fairly reproducible across populations<sup>(20)</sup>. The dietary patterns observed in the present study are also similar to those found in a previous study on a sample (non-representative) of Irish men and women<sup>(22)</sup>. They found three clusters in their sample of Irish adults: a 'Traditional diet'; a 'Prudent diet'; an 'Alcohol & convenience food' cluster. However, using cluster analysis, the present study also found evidence for two other dietary patterns, i.e. a 'Continental' and a 'Light-meal & low-fat milk' pattern, which may be a particular feature of this Irish population alone.

In general, the majority of studies in the literature, which have examined dietary patterns in different population groups, have revealed two predominant and opposing patterns, often referred to as a 'Prudent' and a 'Western' dietary pattern<sup>(23–27)</sup>. The cluster and factor patterns labelled as 'Healthy foods' in the present study correspond quite closely to the 'Prudent' pattern, and elements of both the 'Unhealthy foods' and the 'Traditional Irish' clusters and from the 'Unhealthy foods & high alcohol' and the 'Traditional Irish'



factors were similar to the 'Western' pattern. As would be suggested from the label, the 'Healthy' patterns represent the ideal diet, however only 8.5% of the population were found to exist in this cluster. This indicates that there is much room for growth into this dietary pattern, but also, most importantly, these results provide evidence that a healthy eating pattern is a part of Irish culture (if somewhat relatively small), and that recommendations based on this eating pattern should be acceptable to the population. The 'Traditional Irish' pattern, which was the most dominant cluster, had the highest energy, fat and sugar intakes. Changing this dietary pattern to one of a more 'Healthy' pattern provides a realistic but difficult challenge due to the cultural position of this pattern in the Irish society. Information concerning non-use of foods may prove to be useful, particularly for the design of successful nutrition intervention programmes. From both the cluster and factor analyses, it was found that with regard to the 'Healthy' and 'Unhealthy' patterns, a high consumption of 'Healthy' foods was associated with a low consumption of 'Unhealthy' foods and vice versa.

Although cluster and factor methods are data driven and may be considered objective because they are conducted *a posteriori*, the analytical process is filled with subjective decisions. In the *k*-means cluster algorithm, the user needs to predefine the number of cluster solutions, which can force the data into unrealistic clusters. To prevent this potential pitfall, many scenarios with varying number of clusters are required and cluster solutions need to be compared and contrasted. In the PCA process, the user is also faced with many subjective decisions, such as the criteria used to determine how many factors to extract (e.g. the eigenvalue cut-off), and the method with which to rotate the selected factors. The majority of studies using PCA to derive dietary patterns in the literature have used criteria such as selecting PC based on eigenvalues of >1.0–1.5, orthogonal rotation and factor loadings of the food groups of >0.2 for pattern interpretation<sup>(28)</sup>. Therefore, the present paper followed similar criteria. Schulze *et al.*<sup>(29)</sup> described an approach to calculate factor scores in the form of a sum of the standardised value of the six most loading foods on each factor. These simplified measures are proposed as an approach to overcome the limitation of factor analyses in relation to their non-comparable risk estimates over different populations and studies. Another method of overcoming the reliance on subjective criteria is to use confirmatory PCA. In confirmatory PCA, only food groups with factor loadings above a defined cut-off are retained to measure a food pattern in order to isolate the 'core' of the pattern<sup>(30)</sup>. However, only a few studies have used confirmatory PCA in nutritional epidemiology<sup>(16,31–34)</sup>. Other subjective decisions revolve about selecting the food groups, and even the manner in which the clusters and factors are ultimately labelled is also based on subjective criteria and is liable to different interpretations<sup>(7,33,35–39)</sup>.

In summary, the present study has shown that cluster and PCA, although statistically different methods, identify similar dietary patterns when presented with the same dataset, and that these patterns are directly comparable. However, caution needs to be applied to the subjective decisions involved when using PCA, as this can have a direct impact on the number and type of dietary patterns revealed in the data. Also, both methods in the present study required a different format of

the food-group variable, and therefore consideration of the input variable should always be given separately for each method. In terms of further analysis of the dietary pattern variables generated, cluster analysis produces mutually exclusive groups that are amenable to profiling, thus offering this method considerable advantage over the other. However, as stated earlier, each research study will have its own specific hypotheses and aims, and it is these that should ultimately drive the choice of the dietary pattern analytical technique.

### Acknowledgements

We would like to acknowledge the Irish Department of Agriculture, Fisheries and Food for funding the present analysis. A. P. H. carried out the analysis for the present paper and drafted the manuscript. M. J. G. provided guidance for the analysis and reviewed the final drafts of the manuscript. The authors report no conflicts of interest.

### References

1. Tucker KL, Chen H, Hannan MT, Cupples LA, Wilson PW, Felson D & Kiel DP (2002) Bone mineral density and dietary patterns in older adults: the Framingham osteoporosis study. *Am J Clin Nutr* **76**, 245–252.
2. Randall E, Marshall JR, Graham S & Brasure J (1990) Patterns in food use and their associations with nutrient intakes. *Am J Clin Nutr* **52**, 739–745.
3. 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.
4. Newby PK & Tucker KL (2004) Empirically derived eating patterns using factor or cluster analysis: a review. *Nutr Rev* **62**, 177–203.
5. Kennedy ET, Ohls J, Carlson S & Fleming K (1995) The healthy eating index: design and applications. *J Am Diet Assoc* **95**, 1103–1108.
6. van Dam RM (2005) New approaches to the study of dietary patterns. *Br J Nutr* **93**, 573–574.
7. Martinez ME, Marshall JR & Sechrest L (1998) Invited Commentary: factor analysis and the search for objectivity. *Am J Epidemiol* **148**, 17–19.
8. Newby PK, Muller D & Tucker KL (2004) Associations of empirically derived eating patterns with plasma lipid biomarkers: a comparison of factor and cluster analysis methods. *Am J Clin Nutr* **80**, 759–767.
9. Kant AK, Graubard BI & Schatzkin A (2004) Dietary patterns predict mortality in a national cohort: the National Health Interview Surveys, 1987 and 1992. *J Nutr* **134**, 1793–1799.
10. Bamia C, Orfanos P, Ferrari P, *et al.* (2005) Dietary patterns among older Europeans: the EPIC-elderly study. *Br J Nutr* **94**, 100–113.
11. Crozier SR, Robinson SM, Borland SE, Inskip HM & SWS Study Group (2006) Dietary patterns in the Southampton Women's Survey. *Eur J Clin Nutr* **60**, 1391–1399.
12. Bailey RL, Gutschall MD, Mitchell DC, Miller CK, Lawrence FR & Smiciklas-Wright H (2006) Comparative strategies for using cluster analysis to assess dietary patterns. *J Am Diet Assoc* **106**, 1194–1200.
13. Kiely M, Flynn A, Harrington KE, Robson PJ & Cran G (2001) Sampling description and procedures used to conduct the North/South Ireland Food Consumption Survey. *Public Health Nutr* **4**, 1029–1035.

14. Harrington KE, Robson PJ, Kiely M, Livingstone MB, Lambe J & Gibney MJ (2001) The North/South Ireland Food Consumption Survey: survey design and methodology. *Pub Health Nutr* **4**, 1037–1042.
15. Holland B, Welch AA, Unwin ID, Buss DH, Paul AA & Southgate DAT (1995) *McCance & Widdowson's the Composition of Foods*. London: HMSO.
16. Togo P, Heitmann BL, Sorensen TI & Osler M (2003) Consistency of food intake factors by different dietary assessment methods and population groups. *Br J Nutr* **90**, 667–678.
17. Coakes SJ & Steed LG (1999) *SPSS without Anguish Versions 7.0, 7.5, 8.0 for Windows*. Brisbane: John Wiley and Sons.
18. 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.
19. 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.
20. Moeller SM, Reedy J, Millen AE, Dixon LB, Newby PK, Tucker KL, Krebs-Smith SM & Guenther PM (2007) Dietary patterns: challenges and opportunities in dietary patterns research: an experimental biology workshop, April 1, 2006. *J Am Diet Assoc* **107**, 1233–1239.
21. Kant AK (2004) Dietary patterns and health outcomes. *J Am Diet Assoc* **104**, 615–635.
22. Villegas R, Salim A, Collins MM, Flynn A & Perry IJ (2004) Dietary patterns in middle-aged Irish men and women defined by cluster analysis. *Pub Health Nutr* **7**, 1017–1024.
23. Lopez-Garcia E, Schulze MB, Fung TT, Meigs JB, Rifai N, Manson JE & Hu FB (2004) Major dietary patterns are related to plasma concentrations of markers of inflammation and endothelial dysfunction. *Am J Clin Nutr* **80**, 1029–1035.
24. Fung TT, Hu FB, Holmes MD, Rosner BA, Hunter DJ, Colditz GA & Willett WC (2005) Dietary patterns and the risk of postmenopausal breast cancer. *Int J Cancer* **116**, 116–121.
25. Michaud DS, Skinner HG, Wu K, Hu F, Giovannucci E, Willett WC, Colditz GA & Fuchs CS (2005) Dietary patterns and pancreatic cancer risk in men and women. *J Natl Cancer Inst* **97**, 518–524.
26. Wu K, Hu FB, Willett WC & Giovannucci E (2006) Dietary patterns and risk of prostate cancer in US men. *Cancer Epidemiol Biomarkers Prev* **15**, 167–171.
27. Gao X, Chen H, Fung TT, Logroscino G, Schwarzschild MA, Hu FB & Ascherio A (2007) Prospective study of dietary pattern and risk of Parkinson disease. *Am J Clin Nutr* **86**, 1486–1494.
28. Michels KB & Schulze MB (2005) Can dietary patterns help us detect diet-disease associations? *Nutr Res Rev* **18**, 241–248.
29. Schulze MB, Hoffmann K, Kroke A & Boeing H (2003) Risk of hypertension among women in the EPIC-Potsdam study: comparison of relative risk estimates for exploratory and hypothesis-oriented dietary patterns. *Am J Epidemiol* **158**, 365–373.
30. Newby PK, Weismayer C, Akesson A, Tucker KL & Wolk A (2006) Longitudinal changes in food patterns predict changes in weight and body mass index and the effects are greatest in obese women. *J Nutr* **136**, 2580–2587.
31. Maskarinec G, Novotny R & Tasaki K (2000) Dietary patterns are associated with body mass index in multiethnic women. *J Nutr* **130**, 3068–3072.
32. Togo P, Osler M, Sorensen TI & Heitmann BL (2004) A longitudinal study of food intake patterns and obesity in adult Danish men and women. *Int J Obes Rel Metabol Disord* **28**, 583–593.
33. Tseng M, Breslow RA, DeVellis RF & Ziegler RG (2004) Dietary patterns and prostate cancer risk in the National Health and Nutrition Examination Survey epidemiological follow-up study cohort. *Cancer Epidemiol Biomarkers Prev* **13**, 71–77.
34. Weismayer C, Anderson JG & Wolk A (2006) Changes in the stability of dietary patterns in a study of middle-aged Swedish women. *J Nutr* **136**, 1582–1587.
35. McCann SE, Weiner J, Graham S & Freudenheim JL (2001) Is principal components analysis necessary to characterise dietary behaviour in studies of diet and disease? *Pub Health Nutr* **4**, 903–908.
36. 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.
37. Osler M, Helms Andreasen A, Heitmann B, Hoidrup S, Gerdes U, Mørch Jørgensen L & Schroll M (2002) Food intake patterns and risk of coronary heart disease: a prospective cohort study examining the use of traditional scoring techniques. *Eur J Clin Nutr* **56**, 568–574.
38. Dixon LB, Balder HF, Virtanen MJ, *et al.* (2004) Dietary patterns associated with colon and rectal cancer: results from the dietary patterns and cancer (DIETSCAN) project. *Am J Clin Nutr* **80**, 1003–1011.
39. Newby PK, Muller D, Hallfrisch J, Andres R & Tucker KL (2004) Food patterns measured by factor analysis and anthropometric changes in adults. *Am J Clin Nutr* **80**, 504–513.

**Appendix Table.** Intake of each food group for the total population expressed as g/d or as percentage contribution to total energy intake (% TE) (Mean values and standard deviations)

Food group*	Mean daily food-group intake			
	g/d		%TE	
	Mean	SD	Mean	SD
Alcoholic beverages	332.86	587.01	5.20	7.10
Biscuits and cakes	31.53	32.62	5.76	5.28
Breakfast cereals	35.42	53.97	3.72	3.91
Butter and spreads	17.36	16.46	4.98	4.18
Cheese	12.37	15.28	2.01	2.25
Chips	65.51	62.30	5.97	5.20
Confectionery	18.06	23.00	2.59	3.17
Ice cream and desserts	30.48	39.45	3.01	3.57
Eggs	17.22	21.33	1.36	1.66
Fish	25.92	29.91	2.24	2.65
Fruit	80.03	94.70	2.25	2.77
Fruit juices	33.27	59.83	5.97	5.20
High-energy beverages	106.45	169.91	1.61	2.58
Low-energy beverages	1007.28	556.78	0.12	0.61
Low-fat milk	30.03	48.14	1.98	3.03
Low-fat spreads	4.48	10.72	0.80	1.81
Meat products	40.85	40.43	4.75	4.53
Potatoes	157.94	165.16	5.53	4.92
Poultry	20.65	22.99	1.71	1.98
Poultry dishes	20.86	37.25	1.28	2.50
Red meat	55.00	46.32	5.78	4.56
Red-meat dishes	38.32	51.91	2.40	3.11
Rice and pasta	20.25	31.72	1.21	1.94
Sauces	20.10	21.27	1.39	1.75
Savouries	23.77	40.31	2.26	3.53
Savoury snacks	7.22	12.31	1.66	2.64
Soups	28.52	49.33	0.64	1.20
Sugars and preserves	14.03	22.37	2.82	3.85
Vegetables	120.94	77.33	3.54	2.93
White bread	93.80	62.20	10.92	6.19
Wholemeal bread	44.56	57.96	4.58	5.50
Whole milk	155.39	189.49	4.41	4.96
Yogurts	15.80	32.84	0.66	1.38

\* Intakes of all foods considered as edible fraction only.