

## Irish Section Postgraduate Symposium

# The use of cluster analysis to derive dietary patterns: methodological considerations, reproducibility, validity and the effect of energy mis-reporting

Una M. Devlin\*, Breige A. McNulty, Anne P. Nugent and Micheal J. Gibney  
UCD Institute of Food and Health, University College Dublin, Belfield, Dublin 4, Republic of Ireland

Over the last three decades, dietary pattern analysis has come to the forefront of nutritional epidemiology, where the combined effects of total diet on health can be examined. Two analytical approaches are commonly used: *a priori* and *a posteriori*. Cluster analysis is a commonly used *a posteriori* approach, where dietary patterns are derived based on differences in mean dietary intake separating individuals into mutually exclusive, non-overlapping groups. This review examines the literature on dietary patterns derived by cluster analysis in adult population groups, focusing, in particular, on methodological considerations, reproducibility, validity and the effect of energy mis-reporting. There is a wealth of research suggesting that the human diet can be described in terms of a limited number of eating patterns in healthy population groups using cluster analysis, where studies have accounted for differences in sex, age, socio-economic status, geographical area and weight status. Furthermore, patterns have been used to explore relationships with health and chronic diseases and more recently with nutritional biomarkers, suggesting that these patterns are biologically meaningful. Overall, it is apparent that consistent trends emerge when using cluster analysis to derive dietary patterns; however, future studies should focus on the inconsistencies in methodology and the effect of energy mis-reporting.

### Dietary patterns: Cluster analysis: Adults: Energy mis-reporting

With the global prevalence of chronic diseases increasing, it is now widely accepted that diet has an important role to play, as many of these diseases may have a nutritional base or may be promoted by inappropriate dietary habits<sup>(1,2)</sup>. Traditionally, nutritional epidemiology focused on a detailed examination of single nutrient intake; however, over the last three decades research has moved towards examining the combined effect of total food intake. This significant shift reflects a need to explore the complexity of individual total dietary intake and it is hoped that this alternative approach will help to increase our understanding of the role of diet in chronic diseases and improve the effectiveness of public health recommendations<sup>(3)</sup>. Furthermore, it has been recognised that individuals consume diverse diets consisting of many foods containing

complex combinations of nutrients and it is likely that these nutrients will interact with each other, an effect that may be confounded within the single nutrient approach<sup>(4)</sup>.

One way to examine the combined effect of total food intake on health is to derive dietary patterns. Dietary patterns are typically characterised on the basis of habitual food intake and can be described as a measure of usual intake of food combination in individuals and groups where nutritional variables are grouped according to some criterion of nutritional status<sup>(5)</sup>. Two analytical approaches are commonly used: *a priori* and *a posteriori*. The *a priori* approach is a theoretically driven method that focuses on constructing dietary scores using a predefined combination of diet quality based on published dietary guidelines<sup>(6)</sup>. The *a posteriori* approach is an exploratory method that

**Abbreviations:** hcy, homocysteine; %TE food, percentage total energy contribution from food.

**\*Corresponding author:** Una M. Devlin, email una.devlin@ucd.ie

uses multivariate statistical techniques to derive dietary patterns where large datasets representing total food intake are aggregated and reduced to smaller datasets to summarise total dietary exposure<sup>(7)</sup>. Factor analysis and cluster analysis are two *a posteriori* methods commonly used to derive dietary patterns in nutritional epidemiology. In factor analysis, linear combinations (factors) are created based on correlations between dietary intakes where each individual receives a score for the derived factors; however, these scores are difficult to interpret as an individual can belong to more than one factor<sup>(8)</sup>. Cluster analysis, on the other hand, offers the advantage of deriving dietary patterns which represent homogenous groups that can be related to other variables<sup>(4)</sup>.

In studies where factor and cluster analysis were used simultaneously to derive dietary patterns, results have shown good evidence of comparability. Two studies have indicated that there is a high resemblance between some of the clusters and factors identified due to similarities in food types<sup>(9,10)</sup>. In addition, one study reported that three patterns dominated irrespective of which method was used<sup>(8)</sup>. Dietary patterns derived using both methods have also been compared with plasma lipid markers. Newby *et al.* reported that a cluster and a factor dominated by healthy foods were both inversely associated with plasma TAG, whereas a cluster and a factor dominated by alcohol were both directly associated with HDL and cholesterol<sup>(11)</sup>. Although both methods are directly comparable, it has been suggested that the choice of the dietary pattern analysis technique should depend on the type of outcome that is needed from the dataset as each method approaches the data from different angles and thus answers different questions<sup>(8)</sup>. Other authors have suggested that the ultimate way to approach dietary pattern analysis is to use a combination of factor and cluster analysis as complementary approaches<sup>(12)</sup> in order to give a better perspective and understanding of dietary habits<sup>(13)</sup>.

Clustering methods separate individuals into mutually exclusive, non-overlapping clusters, where an individual can belong to one cluster only, therefore representing a unique cluster or dietary pattern<sup>(8)</sup>. Differences between clusters are based on mean dietary intake of each individual, where the dietary patterns derived are specific to individuals within each cluster and each cluster has a specific food and nutrient composition<sup>(14)</sup>. Clusters are then labelled based on shared characteristics of dietary intake, where individuals with similar dietary intake will cluster together, away from others in dissimilar clusters. Dietary input variables can include nutrients, foods or food groups or a combination of all three<sup>(15)</sup>. However, within the literature, food groups are most commonly used<sup>(8,16–19)</sup>. One reason for using food groups as the preferred dietary input variable is that these groups can represent total dietary intake, accounting for any interaction between nutrients within the groups. Furthermore, various algorithms can also be used in the clustering procedure. The principle of all clustering algorithms is to calculate the Euclidean distance, which measures the distance between each dietary variable consumed together by similar individuals. Individuals are then grouped into clusters where the distance is maximised between the defined centre of each cluster from

others, while the distance is minimised between any single individual and the centre of their closest cluster<sup>(5)</sup>. Of these algorithms, the *k*-means approach is most frequently used<sup>(8,19–21)</sup>, although this algorithm has limitations which will be discussed later. This review examines the literature on dietary patterns derived by cluster analysis in adult population groups only, focusing in particular on methodological considerations, reproducibility, validity and the effect of energy mis-reporting.

### Methodological considerations

Many dietary assessment tools are available to researchers to estimate dietary intake of an individual or a population group. These methods can be split into two categories: one is the prospective method, i.e. those that record data at the time of eating (dietary records) and the other is the retrospective method, i.e. those that collect data about the diet eaten in the past (diet histories, FFQ and dietary recalls)<sup>(22)</sup>. Within dietary pattern analysis, consideration should be given to the most appropriate method, as some may provide more 'favourable' results than others as several may not accurately identify the usual food pattern<sup>(23)</sup>. The impact of the dietary assessment methods used in cluster analysis will be discussed later in the review.

In recent years, scrutiny of the statistical methodology concerning cluster analysis has been undertaken by many researchers, due to its highly exploratory nature. One issue of concern is researcher bias, which can ultimately influence the grouping of the dietary variables and the number of clusters in the final solution<sup>(8)</sup>. The frequently used *k*-means approach has a subjective element as the number of clusters needs to be predefined prior to analysis. To overcome this problem varying cluster solutions are usually run and then the clusters are examined for the best fit using cross-validation methods. Two approaches that can be used to examine the final cluster solution are to calculate the within cluster variance ratio<sup>(20,24,25)</sup> or to generate scree plots<sup>(26,27)</sup>, where higher ratios indicate a better separation of clusters. It has been suggested, however, that there is no gold standard for determining the number of clusters<sup>(15)</sup>. In many cases, the appropriate number of clusters is determined by the author, taking into consideration those which are clearly distinct and nutritionally meaningful, while also maintaining a reasonable sample size<sup>(25)</sup>. In a similar way, there is no gold standard concerning the format of the dietary variable for the clustering procedure. Preferably, the dietary variables should be grouped to suitably represent the dataset to increase the likelihood of identifying sensible dietary patterns. When using food groups as the dietary variable, it has been suggested that food items consumed need to be aggregated into a limited number of groups avoiding the exclusion of subjects due to missing data<sup>(28)</sup>. Previous studies have joined food groups together based on similarities in food group types<sup>(8,16,18)</sup> or on nutrient content and culinary preference<sup>(19,29,30)</sup>. In most cases authors have also differentiated between food groups, e.g. low- or high-energy and low- or high-fat<sup>(8,16,19,29,30)</sup>. Food groups are usually presented using three different methods (1) the frequency of

the food consumed (servings)<sup>(17,19)</sup>, (2) the portion size of the food consumed (grams)<sup>(8,21)</sup> or (3) the percentage total energy contribution from food (%TE food)<sup>(8,30,31)</sup>. Few studies have examined the impact of the methodological differences between these different methods. One author has proposed that when using the %TE food method, differences in energy needs due to sex, age, body weight and level of physical activity can be accounted for<sup>(25)</sup>. One study that compared two methods (servings and %TE food) reported similar clusters for food groups high in energy. However, clusters arising from %TE food were less likely to differentiate between low-energy foods such as fruit and vegetables. The authors therefore concluded that the servings approach best represented the patterns<sup>(32)</sup>. In contrast, a second study that clustered using the grams and %TE food methods showed that the %TE food method best characterised the patterns, which were fully interpretable based on their contributing food group<sup>(8)</sup>. To the best of our knowledge no studies have examined the results obtained comparing all three methods in one dataset, therefore, it is difficult to make firm conclusions on the best method to use. One way to overcome the issue of high- or low-energy food groups affecting the patterns is to standardise the variables prior to analysis ensuring that variables with large variances which may have greater effects on resulting patterns than those with small variances can be accounted for<sup>(24)</sup>. Ideally, by standardising the input variables, all food groups will have equal influence on the clustering procedure. Research carried out by Wirfalt *et al.* examining the effect of standardising variables found that the distribution of individuals was more evenly spread and differences in nutrient intake across patterns were improved when using the un-standardised approach<sup>(33)</sup>. Furthermore, in a follow-up study, Wirfalt reported that the transformation of variables by standardisation may have an effect on the dietary patterns identified as low-energy foods may be given equal weights to high-energy foods, which may represent poor dietary patterns<sup>(34)</sup>. Overall, there is insufficient evidence regarding the standardisation procedure and more research is needed.

### Dietary patterns in healthy population groups

Throughout the last three decades many studies have identified meaningful dietary patterns in healthy population groups using cluster analysis as the patterning method. Initial studies focused on identifying patterns where nutrient intakes were inadequate *v.* published dietary recommendations, thus acknowledging that cluster analysis is a useful tool for identifying groups of people who may be at nutritional risk<sup>(35,36)</sup>. Later studies have accounted for the influence of sex, age, socio-economic status, geographical area and weight status. A range of dietary assessment methods were used including FFQ, dietary recalls and diet records. Only one study used nutrients as the clustering variable<sup>(35)</sup>, whereas another used meal type<sup>(37)</sup>; therefore, food groups were predominantly used and were presented using servings<sup>(9,19,36,38–44)</sup>, grams<sup>(13,16,21,45–47)</sup> and %TE food<sup>(8,18,31)</sup>. It is noteworthy that no matter which dietary

assessment method or clustering variable was used, similar dietary patterns have been found across a collection of studies in healthy population groups.

In all studies, labels or names are normally assigned to characterise each pattern, based on the dietary intake that contributes relatively greater proportions<sup>(11,31,48)</sup>. Two commonly used terms are 'healthy' patterns characterised by the consumption of fruits and vegetables and 'unhealthy' patterns characterised by the consumption of foods high in fat and salt<sup>(9,31,38,39)</sup>. 'Healthy' patterns can also be referred to as 'prudent', while 'unhealthy' patterns can also be referred to as 'western' or 'traditional'<sup>(8,21,45)</sup>. A strength of these studies is large sample size ( $n > 1379$ )<sup>(8,21,35,38,39)</sup> (only one study of sample size  $n < 213$ )<sup>(45)</sup> though many were carried out in female<sup>(9,36)</sup> or older adults<sup>(31)</sup> only. In one study of London adults aged 39–63 years, differences were reported in the type of 'healthy' patterns identified by using terms such as 'very healthy' or 'moderately healthy', similarly for 'unhealthy' patterns<sup>(39)</sup>. Other descriptive labels used to characterise dietary patterns relate to 'high- or low-nutrient density'<sup>(40,43)</sup> or 'glycaemic level'<sup>(42)</sup>; however, these findings are limited to three US studies in either females or older adults. Furthermore, many studies have examined differences in socio-economic status according to dietary patterns, reporting that typically 'healthy' patterns are associated with increased socio-economic status in males and females<sup>(13,21,36,39,46)</sup>.

Significant differences among dietary patterns by sex have also been reported, highlighting the need to examine males and females separately in healthy population groups<sup>(26,49)</sup>. In a study carried out in a representative sample of UK adults aged 16–64 years, it was reported that dietary patterns differ by sex<sup>(16)</sup>, but these differences were lost in an older cohort aged 65+ years of the same study<sup>(46)</sup>. Confirmation that dietary patterns differ by sex was reported in a cohort of older Italian adults aged 65+ years<sup>(41)</sup>, Swedish adults aged 30–60 years<sup>(19)</sup>, African-American adults aged 18+ years<sup>(44)</sup> and American adults aged 20–70 years<sup>(17)</sup>. These studies suggest that dietary patterns differ by sex and this should therefore be accounted for in public health recommendations. Few studies have reported differences among age across dietary patterns<sup>(16,35,40,45)</sup> and to the best of our knowledge no studies have examined the effect of age groups on dietary patterns in a large representative sample.

Dietary pattern analysis is also influenced by geography. Within large cohorts of older European adults, specific dietary patterns have been found to represent those living in Northern and Southern regions where one of these patterns is usually considered as more healthy<sup>(18,41,47,50)</sup>. Differences have also been found at a national level; in a large study of Norwegian females aged 41–56 years, one dietary pattern was dominated by those living in a certain region of Norway<sup>(13)</sup>. These results could therefore indicate that dietary patterns are influenced by geography and are associated with cultural perceptions, beliefs and attitudes about foods which can ultimately affect food choice. Although these studies are of large sample sizes, a limitation is that they are limited to groups of older adults and female populations only.

Three studies have also examined differences in weight status according to dietary patterns in healthy population groups. These studies have reported that BMI of individuals is significantly different across all patterns after controlling for age, sex, exercise and total energy intake in US adults (mean age 37 years)<sup>(26)</sup> and UK adults aged 16–64 years<sup>(16)</sup>. In the US study, the dietary pattern with the highest mean BMI was found to be predominantly male and had high intake of soft drinks. In contrast, in a large sample of Swedish adults aged 47–68 years, Holmback reported that the ‘fruit’ pattern had the greatest proportion of overweight individuals<sup>(37)</sup>. These differences may perhaps be explained by the different types of clustering variables used (servings, %TE and meal type); however, further research is required.

The earlier studies in general show consistent findings across dietary patterns in healthy population groups. One issue of concern is that few have accounted for energy misreporters, with only two studies excluding such reporters from their analysis. This issue will be discussed later in the review. It is evident that from these studies, literature is accumulating in relation to using cluster analysis to derive dietary patterns taking into account sex, age, socio-economic status, geographical area and weight status; however, the lack of consensus of some studies warrants further research in this area.

### Dietary patterns and associations with chronic diseases

The effect of diet on chronic diseases is a key consideration in nutritional epidemiology. By considering the effect of total diet using dietary pattern analysis, it is believed that various patterns may influence the development and possibly increase the risk of many diet related chronic diseases over time. An overview of the literature examining the association of dietary patterns and chronic diseases is outlined in Table 1 and reviewed briefly later.

As previously discussed, evidence has suggested that weight status can differ according to dietary patterns in cross-sectional cohorts<sup>(16,26,37)</sup>. In studies, specifically examining the risk of obesity, it has been reported that in comparison with ‘healthy’ patterns and after adjustments for confounders, patterns that are considered ‘less healthy’ have a significantly larger BMI and waist circumference<sup>(29,51)</sup>, higher total percentage body fat (males only)<sup>(25)</sup> and are associated with an increased risk of overweight (14–17%)<sup>(52,53)</sup> and obesity (20%)<sup>(53)</sup>. Interestingly, Carrera *et al.* found that no one pattern was associated with increased risks of obesity as it was reported that BMI and waist circumference were high among all patterns identified<sup>(54)</sup>. Overall, arising from these large studies involving a wide variety of age groups, the consensus appears that subjects in ‘healthy’ patterns following current dietary recommendations are at lesser risk of becoming overweight or obese. Furthermore, it has been suggested that due to the complexity of total diet, future studies should consider the influence of total food volume on energy balance<sup>(29)</sup>.

Dietary patterns have also been associated with CVD risk mainly in prospective studies. As before, ‘healthy’

patterns have been shown to be protective, lowering the risk of subclinical heart disease<sup>(55)</sup> and carotid atherosclerosis<sup>(56)</sup> by 4% and are favourably associated with anthropometric, blood pressure and blood lipid values<sup>(28)</sup> and with markers of inflammation<sup>(57)</sup> in comparison with the other patterns identified. However, one study relied on the analysis of non-fasting blood samples<sup>(28)</sup>. In one case–control study, food groups associated with increased risk of acute myocardial infarction after adjustments for confounders were a ‘red meat and alcohol’ pattern in males and females and a ‘low fruit and vegetables’ pattern in females only, where the ‘red meat and alcohol’ pattern had significantly higher risks of CVD risk markers than those in a ‘healthy’ pattern<sup>(58)</sup>. Interestingly, in one study no one pattern was associated with increased CVD risk although a ‘sweets’ pattern, showed a protective effect against CVD risk factors as significant associations were reported among HDL and elevated systolic blood pressure<sup>(59)</sup>. These results provide support for the protective effects of ‘healthy’ dietary patterns against CVD.

Dietary patterns have also been linked to risk factors for diabetes. In one study, where 67 and 33% of subjects had normal and impaired glucose tolerance, respectively, it was reported that the ‘white bread’ pattern was associated with poorest insulin sensitivity and adiposity levels, whereas a ‘wine’ and ‘dark bread’ pattern was associated with improving these markers<sup>(60)</sup>. In non-diabetic cohorts, it has been reported that a pattern that is high in dairy products and low in staple foods is associated with a lower prevalence of type-2 diabetes<sup>(61)</sup>, and a ‘healthy’ pattern improves insulin concentration and anthropometric profiles<sup>(62)</sup>. One study also reported that a pattern with high intake of animal and soyabean products had a higher prevalence of glucose tolerance abnormalities, after adjustment for confounders<sup>(63)</sup>. The cross-sectional study design of most of these studies is a limitation as information on diet (mainly collected using FFQ) and indicators of diabetes were collected at one specific point in time. This highlights the need for more prospective studies to be carried out in order to determine how the dietary patterns affect diabetes over a certain time frame.

Specific dietary patterns have also been associated with cancer risk, mainly in case–control studies. As before, ‘healthy’ dietary patterns were shown to have protective effects, and to reduce the risk of oesophageal cancer<sup>(64)</sup>, gastric cancer<sup>(65)</sup>, ovarian cancer<sup>(66)</sup> and lung cancer in subjects who smoke<sup>(67)</sup>. ‘Unhealthy’ patterns increased the risk of oesophageal and colorectal cancer<sup>(64,68)</sup> and one pattern with high intake of bread and pasta was unfavourable for breast and ovarian cancer risk<sup>(66)</sup>. Although these results have shown patterns that may increase cancer risk and others that are protective, a difficulty in epidemiological studies of diet and cancer is lack of specific biomarkers for the disease. Further research needs to be carried out to establish environmental factors that may increase cancer risk.

The effect of dietary patterns on a combination of chronic diseases has also been evaluated. In one study, it was reported that after 16 years of follow-up, levels of overweight and obesity increased from 67 to 76% and 81 to 91%, respectively, whereas the rates of diabetes

**Table 1.** Associations between dietary patterns and chronic diseases

Reference	Study type	n*	Disease	Cohort	Patterns associated with chronic disease		
					Labels	Food groups (highest contribution <sup>†</sup> )	Risk
(29)	L	459	Obesity	Adults (30–80 years)	Meat and potatoes White bread	Red and processed meat, potatoes and fast food White bread	Increase in mean annual change: BMI 0.30 v. 0.05 in healthy pattern ( $P<0.01$ ) Increase in mean annual change: WC 1.32 v. 0.43 in healthy pattern ( $P<0.05$ )
(51)	CS	825	Obesity	Adults (60–92 years)	Rice	Rice, added fats (mainly cooking oil), beans and poultry	BMI was greater than all other patterns ( $P<0.05$ ) and WC was greater than a fruit and cereal cluster ( $P<0.05$ )
(25)	P	3075	Obesity	Older adults (70–79 years)	Meat, snacks, fat and alcohol Breakfast cereal	Processed meat, meat, fried poultry, alcohol, high energy drinks, snacks, nuts, salad dressings and miscellaneous fats Breakfast cereals	Higher total percentage body fat ( $P<0.05$ ) in comparison with healthy pattern – males only Higher total percentage body fat ( $P<0.05$ ) and higher abdominal visceral fat ( $P<0.05$ ) in comparison with healthy pattern – males only
(52)	L	737	Obesity	Females (30–89 years)	Empty calorie pattern	Animal and vegetable fats, sweets and desserts, meat and sweetened beverages	17% absolute increased risk in comparison with heart healthy pattern
(53)	CS	15 890	Obesity	Adults (20–59 years)	Refined foods and sweets Diverse	Alcohol, soft drinks, white bread, fast food, sweets and snacks Whole fat dairy, rice & pasta, meat, poultry, eggs, saturated fat, fruits, vegetables	Both patterns were associated with a 14–17% increased risk of being overweight ( $P<0.01$ ) and 20% increased risk of being obese ( $P<0.001$ ) in comparison with a traditional pattern
(54)	CS	659	Obesity	Adults (18+ years)	N/A	N/A	No one pattern was associated with lowering of the risk of obesity, as the levels of BMI and WC were high across all patterns
(55)	P	1423	CVD	Females (18–76 years)	Less heart healthy	High-fat foods	Higher total, LDL cholesterol and total to HDL cholesterol ratio ( $P<0.05$ ) than heart healthy pattern. In all, 11% of the sample had subclinical heart disease at follow-up, in comparison with 7% of the heart healthy pattern
(56)	L	1423	CVD	Females (18–76 years)	Light eating Empty energy	Lowest energy content <sup>†</sup> Sweetened beverages, red meat and desserts	Follow up: 11% had carotid atherosclerosis in comparison with 7% of the heart healthy pattern ( $P<0.05$ ) Follow up: 18% had carotid atherosclerosis in comparison with 7% of the heart healthy pattern ( $P<0.05$ )
(28)	CS	3452	CVD	Adults (25–74 years)	Traditional Fast energy	Medium fat milk, offal, boiled coffee and potatoes Soft drinks, white bread, fast food, full fat milk, cheese, alcohol, sweets and snacks	Significantly higher BMI, WHR and lower serum HDL ( $P<0.05$ ) in comparison with the healthy pattern Significantly higher BMI, WHR, blood pressure, serum TAG and lower serum HDL ( $P<0.05$ ) in comparison with the healthy pattern

Dietary patterns derived by cluster analysis

Table 1 (Continued)

Reference	Study type	n*	Disease	Cohort	Patterns associated with chronic disease		
					Labels	Food groups (highest contribution <sup>†</sup> )	Risk
(57)	P	4999	CVD	Adults (46–73 years)	Milk fat Sweets and cakes	Cheese, whole milk, white bread and sweets Sugar, sweets, jam, cakes, biscuits and soft drinks	Associated with high-leucocyte count in females, increased Lp-PLA2 mass in males Associated with high-leucocyte count and increased Lp-PLA2 mass in females
(58)	CC	820 v. 2196	CVD	Adults (18+ years)	Red meat and alcohol	Red meat, fast food and alcohol	Association between CVD risk markers in comparison with the healthy pattern – Females: higher adjusted total to HDL; Males: higher blood pressure, hs-CRP, uric acid levels
(59)	CS	1313	CVD	Females (50+ years)	N/A	N/A	No one pattern associated with CVD risk
(60)	CS	980	Diabetes	Adults (40–69 years)	White bread	White bread, tomatoes, cheese, dried beans, eggs, meat, fats and oils and beer	Significant differences found in insulin sensitivity, fasting insulin, BMI and waist ( $P < 0.05$ ) in comparison with all other patterns
(61)	P	64 191	Diabetes	Females (40–70 years)	Cluster 1	Staple foods	Higher incidence of type 2 diabetes after 6.9 years follow-up in comparison with a pattern high in dairy milk
(62)	CS	2875	Diabetes	Adults (age NR)	Soda Refined grains and sweets	Meat, chocolate and miscellaneous sweets Refined grains, sweets, beer and soda	Higher mean fasting insulin concentration ( $P < 0.001$ ) Higher WC ( $P < 0.008$ ) and BMI ( $P < 0.02$ )
(63)	CS	20, 210	Diabetes	Adults (45–69 years)	New affluence	Animal and soyabean products	Highest prevalence of glucose abnormality (8%) in comparison with pattern with the lowest prevalence (3.9%)
(64)	CC	124, 124 v. 449	Oesophageal and stomach cancer	Adults (21+ years)	High meat	Red meat, processed meat and beans	Associated with a 3.6-fold higher risk of esophageal cancer and a 2.6-fold higher risk of stomach cancer in comparison with a healthy pattern
(65)	CC	591 v. 1463	Gastric cancer	Adults (18–93 years)	Pattern II	Low intake of fruits, salads, vegetables, meat, fish and dairy products <sup>†</sup>	Associated with a significant 1.7-fold increase risk of gastric cancer in comparison with pattern I (healthy pattern)
(66)	CC	2569, 1031 v. 3413	Breast and ovarian cancer	Females (17–79 years)	G5	Bread, pasta	Unfavourable for both cancers – breast cancer, OR 1.23, ovarian cancer, OR 1.21
(67)	CC	254 v. 184	Lung cancer	Adults (age NR)	Unhealthy	Total fat, saturated fat, animal fat, cholesterol and alcohol	Higher risk of lung cancer than a healthy pattern
(68)	CC	465 v. 426, 171 v. 309	Colorectal cancer	Adults (30–79 years)	Cluster 2	White bread, pork, processed meat, potatoes, rice and pasta	Significant risk of cancer as compared with cluster 1 (high intake of healthy foods). No other pattern associated with risk
(69)	L	1666	Chronic diseases	Males (18–77 years)	N/A	N/A	Chronic disease risk factor levels were high in all dietary patterns at baseline and follow up
(70)	P	7731	Chronic diseases	Adults (age NR)	Unhealthy	White bread, processed meat, fries and full cream milk	Compared with the healthy pattern, this pattern increased risk of coronary death or myocardial infarction and diabetes, after adjustments

(71)	CS	1052	Metabolic syndrome	Adults (42–74 years)	Starch	Refined grains (bread, rice and pasta)	High TAG levels (41.9%) and abdominal obesity (30.6%)
(34)	P	4999	Metabolic syndrome	Adults (45–68 years)	Animal products White bread  Many foods and drinks	Meat, eggs and dairy products White bread  Cheese, fat meat, cakes, fruits, white bread, low fat meat, boiled potato, medium fat spread, low fat milk, regular milk, low fat spread and fibre bread	High impaired fasting glucose levels (38.1%) Females: increased risk of hyperinsulinaemia (33%); Males and females: increased risk of dyslipidaemia Males: increased risk of hyperglycaemia (27%) and central obesity
(72)	CS	4730	Metabolic syndrome	Adults (20+ years)	Meat and alcohol	Processed meat and alcohol	Increased risk: 33% elevated blood glucose, 21% elevated serum TAG, 21% elevated blood pressure
(73)	L	907	Osteoporosis	Adults (69–93 years)	Candy	Candy	Males: low bone mineral density compared with fruits and vegetables, cereal pattern ( $P<0.05$ ); Females: low bone mineral density compared with four other patterns but not the sweet baked products pattern ( $P<0.01$ )

L, longitudinal; WC, waist circumference; P, prospective; CC, case-control; CS, cross-sectional; hsCRP, high-sensitivity C-reactive protein; WHR, waist:hip ratio; Lp-Pla2, lipoprotein.

\*Disease v. control (CC studies only).

<sup>†</sup>Lowest contribution.

nearly doubled from 10 to 18% in the total population<sup>(69)</sup>. No significant difference in risk was found according to dietary patterns, as it was reported that chronic disease risk factors were high in all patterns; however, the sample consisted of only males living in one suburban community of the US. In another study, a pattern characterised by the consumption of wholemeal bread, fruits, vegetables, pasta and rice lowered cancer mortality rate and myocardial infarction rates and a pattern characterised by wholemeal bread, fruits, vegetables and polyunsaturated margarine lowered the incidence of obesity<sup>(70)</sup>. This provides extra support for the health promoting effects of healthy diets.

Dietary patterns have also been explored in relation to the metabolic syndrome. In one study of Italian non-diabetic adults, the highest prevalence of the metabolic syndrome was found in the ‘starch’ and ‘animal products’ patterns and the lowest prevalence found in a ‘vegetable oil and fat spread’ pattern and a ‘vegetable and fruit’ pattern<sup>(71)</sup>. Furthermore, in a Swedish study, it was reported that in males the ‘many foods and drinks’ and the ‘white bread’ pattern and in females the ‘white bread’ pattern only had increased risks of metabolic risk factors<sup>(34)</sup>. Song *et al.* also found increased risks of metabolic risk factors, although this time with a ‘meat and alcohol’ pattern, where it was also reported that a ‘traditional’ pattern that was characterised by high intake of white rice and vegetables had a 23% lower likelihood of having low HDL-cholesterol<sup>(72)</sup>. One limitation of these studies is that divergent definitions were used to define the metabolic syndrome prior to analysis.

Few studies have examined the association of dietary patterns with a risk of osteoporosis. In one study an association with bone mineral density was reported, as it was demonstrated that a diet consisting of high intake of fruits, vegetables and breakfast cereals and limited in less nutrient dense foods may contribute to better bone mineral density in both males and females, though this association was not as strong in females, as levels of bone mineral density were fairly equal among all patterns identified<sup>(73)</sup>.

Overall strength of these studies includes large sample sizes where a wide variety of clustering variables were also used; nevertheless as with healthy population groups the issue of energy mis-reporting is overlooked, as few authors have excluded these mis-reporters from their analysis. Findings mostly from cross-sectional studies have linked dietary patterns and numerous foods associated with these patterns to chronic diseases; however, further research including targeted nutrition interventions is warranted to fully assess the relationship taking into account all other environmental factors that may influence the disease. As it is well known that the progression of these chronic diseases gradually worsens over time, future studies should also consider the importance of prospective and case-control studies, to help advancements in the area.

### Dietary patterns and associations with nutritional biomarkers

More recently, cluster analysis has been used firstly to derive dietary patterns, and thereafter differences in

nutritional biomarkers explored in an attempt to examine the relationship between the two. It is hoped that this will enhance the knowledge base as to whether these dietary patterns are biologically meaningful.

In addition to the earlier studies on markers of lipid metabolism and inflammation, dietary patterns have been associated with markers of homocysteine (hcy) and vitamin B status. Hcy is an important and well-recognised biomarker in nutritional epidemiology as high levels have been linked to increasing the risk of CVD<sup>(74)</sup>. In a sample of 119 Chinese adults aged 35–49 years, it was found that relative to the ‘fruit and milk’ pattern, those subjects consuming a ‘refined cereals’ pattern were 4 and 5.2 times more likely to have high hcy and low vitamin B<sub>12</sub> concentration, respectively<sup>(75)</sup>. Another study investigated the levels of folate and hcy in a sample of 354 American males aged 21–88 years, following the folic acid fortification programme in the US. Within this study it was reported that plasma folate increased in all three dietary patterns identified, although plasma hcy decreased in the low fruit and vegetable pattern only<sup>(76)</sup>. Limitations of these studies include small sample sizes where one study was limited to males only.

A study has also linked dietary patterns to metabolic profiles in a small sample of Irish adults aged 18–63 years. Three dietary patterns were identified, and when compared with metabolic profiles (using metabolomics<sup>(77)</sup>), it was reported that food groups within patterns could be associated with concentration of metabolites<sup>(30)</sup>. A pattern that had high intake of fruits and vegetables and a pattern that had high intake of red meat were associated with phenylacetylglutamine and *O*-acetylcarnitine, respectively. Although one major limitation of this study is its small sample size, the findings of this study underline the ability of metabolomics to identify novel biomarkers of dietary intake. Future studies should consider advancing these results in larger studies, in order to strengthen findings.

### Reproducibility and validity

Although dietary pattern analysis has become of major interest in the field of nutritional epidemiology, the reproducibility and validity of the patterns derived are not clear, and few studies have fully evaluated this issue. As part of the Framingham Nutrition Studies, dietary patterns were identified for adult males and females aged 18–76 years separately. Five patterns were found to best represent each sex, with some patterns being associated with healthier nutrient profiles, while others were associated with disease risk<sup>(17)</sup>. The internal validity of the five dietary patterns identified for women was assessed and it was found that 80% of the sample was correctly classified when using a discriminant analysis technique to measure the stability of the patterns<sup>(48)</sup>. Furthermore, the authors used the results of this study to derive a statistical scoring system or algorithm that would classify a subject from a newer Framingham Nutrition Study into one of the previously identified patterns for males and females. Using the scoring system it was reported that 80% of new males and females under study were correctly classified into one of the previous

patterns already identified<sup>(78)</sup>. The results from this large population based study show that dietary patterns are reproducible across similar population groups, although it should be noted that reproducibility does not guarantee validity. As mentioned previously, cluster analysis can be carried out using different algorithms; however, to date just one study has investigated the differences between these. Lo Siou *et al.* reported that when the clustering variable was presented as the %TE food method, the *k*-means approach (in comparison with Ward’s and flexible beta methods) had the highest reproducibility of cluster solutions for Canadian adults aged 35–69 years<sup>(20)</sup>. When the sample was split by sex, a strong relationship was only seen for males; similar results were not found in females, therefore, highlighting the need for further research in the area. One study has also evaluated the influence of the dietary assessment method used (FFQ and 3-d diary), by comparing the classification rate of subjects into the same dietary patterns using either method, where it was found that four out of ten subjects were misclassified<sup>(79)</sup>. Furthermore, the question is raised as to what is the appropriate threshold for acceptable correct classification. As few studies have assessed both reproducibility and validity, it is clear that there is insufficient evidence to make firm conclusions; therefore highlighting the need for further research.

### Energy mis-reporting

Energy mis-reporting is a major issue in dietary surveys<sup>(22)</sup>. Research has indicated consistent errors in self-reported dietary intake, using the available dietary assessment methods<sup>(80)</sup>. Dietary intake is commonly over- or under-reported leading to implausible energy intake in population groups, where the latter may be considered the most detrimental to research studies. Under-reporting of dietary intake can happen in three ways, where subjects can (1) deny ever eating the food at all; (2) fail to report the correct portion size consumed or (3) fail to report how many times the food is actually consumed. Approaches to identify under-reporters are to calculate the ratio of energy intake to BMR where cut-off values are applied described by Goldberg *et al.*<sup>(81)</sup> or by using the gold standard doubly labelled water technique<sup>(82)</sup>. In studies of under-reporting, it has been found that females, overweight and obese subjects are more likely to under-report their dietary intake<sup>(83–86)</sup>. This is no exception in dietary pattern analysis studies as significant differences have been reported among males and females<sup>(37,46)</sup> and healthy dietary patterns have been found to contain the greatest proportion of females and overweight subjects<sup>(19,37)</sup>. In contrast, Pryer *et al.* found that there was no difference in the proportion of under-reporters across the patterns<sup>(16)</sup>, although Martikainen *et al.* demonstrated that differences in the numbers of under-reporters exist across all patterns; however, these differences are not systematically associated with good or bad diets<sup>(39)</sup>. Other studies have found that under-reporting of energy intake is not uniformly distributed among dietary patterns<sup>(87,88)</sup>. In one study the highest prevalence of under-reporting fell among those in the healthy

pattern, where although this study measured under-reporting using the doubly labelled water method, the sample consisted of females only aged 18–57 years<sup>(88)</sup>.

To the best of our knowledge, no studies have examined the effects of energy mis-reporting by identifying patterns for adequate and under-reporters separately. Two studies have although demonstrated that patterns generated following the removal of under-reporters are relatively similar in comparison with patterns of the total sample (including adequate and under-reporters)<sup>(19,39)</sup>. In one study 70% of participants fall into the same pattern regardless of their reporting status<sup>(39)</sup>. The limitations of both these studies are that the authors have only briefly acknowledged under-reporting and there is a lack of published statistical analysis. Similarly, in another study patterns were identified in the total population and adequate reporters, where it was shown that the correlation between energy intake and weight status was improved for females only after removal of under-reporters<sup>(89)</sup>. Although it is not clear the effect energy mis-reporting may have on dietary pattern analysis, only two studies have removed such reporters from their analysis in healthy population groups<sup>(13,21)</sup> and eight studies in chronic disease groups<sup>(25,53,54,61,62,70,72,73)</sup>.

### Conclusion and future work

From the numerous studies mentioned in this review, some consistent trends emerge when using cluster analysis to derive dietary patterns. It can be argued that there is homogeneity of dietary patterns across populations, where the consistency of patterns identified suggests that they are reproducible. Despite this, given the data driven nature of this statistical technique, the extent to which the identified patterns are reproducible and the extent to which they can be used to develop the understanding of nutritional epidemiology remains debatable. Several important issues have been highlighted, specifically regarding the methodological aspect of cluster analysis and these should be considered in future studies. However, in the earlier studies, different clustering techniques and procedures have been used, making it difficult to draw firm conclusions. Few studies have examined the effect of energy mis-reporting and it is clear that this effect is not fully understood. This review demonstrates the need for large representative cross-sectional and longitudinal studies to assess the effects of energy mis-reporting by carrying dietary pattern analysis on (1) the total population, (2) adequate reporters and (3) under-reporters.

### Acknowledgements

U.M.D. wrote the review, B.A.McN., A.P.N. and M.J.G. provided expert advice in the drafting of the paper and commented on drafts of the paper. The authors declare no conflict of interest. The work was supported by joint funding from the Irish Department of Agriculture, Fisheries and Food and the Health Research Board under the Food for Health Research Initiative (2007–2012).

### References

1. World Health Organization (2003) Diet, Nutrition and the Prevention of Chronic Diseases. Joint WHO/FAO Expert Consultation. *WHO Technical Report Series no. 916*. Geneva: WHO.
2. Kafatos A & Codrington CA (1999) Nutrition and diet for healthy lifestyles in Europe: the Eurodiet Project. *Public Health Nutr* **2**, 327–328.
3. Jacques P & Tucker K (2001) Are dietary patterns useful for understanding the role of diet in chronic disease? *Am J Clin Nutr* **73**, 1–2.
4. Hu F (2002) Dietary pattern analysis: a new direction in nutritional epidemiology. *Curr Opin Lipidol* **13**, 3–9.
5. Tucker KL (2010) Dietary patterns, approaches, and multicultural perspective. *Appl Physiol Nutr Metab* **35**, 211–218.
6. Kennedy ET, Ohls J, Carlson S *et al.* (1995) The Healthy Eating Index: design and applications. *J Am Diet Assoc* **95**, 1103–1108.
7. Kant AK (2004) Dietary patterns and health outcomes. *J Am Diet Assoc* **104**, 615–635.
8. Hearty AP & Gibney MJ (2009) Comparison of cluster and principal component analysis techniques to derive dietary patterns in Irish adults. *Br J Nutr* **101**, 598–608.
9. Crozier SR, Robinson SM, Borland SE *et al.* (2006) Dietary patterns in the Southampton Women's Survey. *Eur J Clin Nutr* **60**, 1391–1399.
10. Cunha DB, Almeida RM & Pereira RA (2010) A comparison of three statistical methods applied in the identification of eating patterns. *Cad Saude Publica* **26**, 2138–2148.
11. 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.
12. Hoffman K, Schulze MB, Boeing H *et al.* (2002) Dietary patterns: report of an international workshop. *Public Health Nutr* **5**, 89–90.
13. Engeset D, Alsaker E, Ciampi A *et al.* (2005) Dietary patterns and lifestyle factors in the Norwegian EPIC cohort: the Norwegian Women and Cancer (NOWAC) study. *Eur J Clin Nutr* **59**, 675–684.
14. Newby PK & Tucker KL (2004) Empirically derived eating patterns using factor or cluster analysis: a review. *Nutr Rev* **62**, 177–203.
15. Togo P, Osler M, Sorensen TI *et al.* (2001) Food intake patterns and body mass index in observational studies. *Int J Obes Relat Metab Disord* **25**, 1741–1751.
16. Pryer JA, Nichols R, Elliott P *et al.* (2001) Dietary patterns among a national random sample of British adults. *J Epidemiol Community Health* **55**, 29–37.
17. Millen BE, Quatromoni PA, Gagnon DR *et al.* (1996) Dietary patterns of men and women suggest targets for health promotion: the Framingham Nutrition Studies. *Am J Health Promot* **11**, 42–52.
18. Haveman-Nies A, Tucker KL, de Groot LC *et al.* (2001) Evaluation of dietary quality in relationship to nutritional and lifestyle factors in elderly people of the US Framingham Heart Study and the European SENECA study. *Eur J Clin Nutr* **55**, 870–880.
19. Winkvist A, Hornell A, Hallmans G *et al.* (2009) More distinct food intake patterns among women than men in northern Sweden: a population-based survey. *Nutr J* **8**, 12.
20. Lo Siou G, Yasui Y, Csizmadi I *et al.* (2011) Exploring statistical approaches to diminish subjectivity of cluster analysis to derive dietary patterns: the Tomorrow Project. *Am J Epidemiol* **173**, 956–967.

21. Villegas R, Salim A, Collins M *et al.* (2004) Dietary patterns in middle-aged Irish men and women defined by cluster analysis. *Public Health Nutr* **7**, 1017–1024.
22. European Food Safety Authority (2009) General principles for the collection of national food consumption data in the view of a pan-European dietary survey. *EFSA J* **7**, 1435.
23. Moeller SM, Reedy J, Millen AE *et al.* (2007) Dietary patterns: challenges and opportunities in dietary patterns research an Experimental Biology workshop, April 1, 2006. *J Am Diet Assoc* **107**, 1233–1239.
24. Michels KB & Schulze MB (2005) Can dietary patterns help us detect diet disease associations? *Nutr Res Rev* **18**, 241–248.
25. Anderson AL, Harris TB, Houston DK *et al.* Relationships of dietary patterns with body composition in older adults differ by gender and PPAR- $\gamma$  Pro12Ala genotype. *Eur J Nutr* **49**, 385–394.
26. Wirfalt AK & Jeffery RW (1997) Using cluster analysis to examine dietary patterns: nutrient intakes, gender, and weight status differ across food pattern clusters. *J Am Diet Assoc* **97**, 272–279.
27. Bailey RL, Mitchell DC, Miller CK *et al.* (2007) A dietary screening questionnaire identifies dietary patterns in older adults. *J Nutr* **137**, 421–426.
28. Berg CM, Lappas G, Strandhagen E *et al.* (2008) Food patterns and cardiovascular disease risk factors: the Swedish INTERGENE research program. *Am J Clin Nutr* **88**, 289–297.
29. Newby PK, Muller D, Hallfrisch J *et al.* (2003) Dietary patterns and changes in body mass index and waist circumference in adults. *Am J Clin Nutr* **77**, 1417–1425.
30. O'Sullivan A, Gibney MJ & Brennan L (2011) Dietary intake patterns are reflected in metabolomic profiles: potential role in dietary assessment studies. *Am J Clin Nutr* **93**, 314–321.
31. Anderson AL, Harris TB, Tyllavsky FA *et al.* (2011) Dietary patterns and survival of older adults. *J Am Diet Assoc* **111**, 84–91.
32. Bailey RL, Gutschall MD, Mitchell DC *et al.* (2006) Comparative strategies for using cluster analysis to assess dietary patterns. *J Am Diet Assoc* **106**, 1194–1200.
33. Wirfalt E, Mattisson I, Gullberg B *et al.* (2000) Food patterns defined by cluster analysis and their utility as dietary exposure variables: a report from the Malmö Diet and Cancer Study. *Public Health Nutr* **3**, 159–173.
34. Wirfalt E, Hedblad B, Gullberg B *et al.* (2001) Food patterns and components of the metabolic syndrome in men and women: a cross-sectional study within the Malmo Diet and Cancer cohort. *Am J Epidemiol* **154**, 1150–1159.
35. Hulshof KF, Wedel M, Lowik MR *et al.* (1992) Clustering of dietary variables and other lifestyle factors (Dutch Nutritional Surveillance System). *J Epidemiol Community Health* **46**, 417–424.
36. Greenwood DC, Cade JE, Draper A *et al.* (2000) Seven unique food consumption patterns identified among women in the UK Women's Cohort Study. *Eur J Clin Nutr* **54**, 314–320.
37. Holmback I, Ericson U, Gullberg B *et al.* (2009) Five meal patterns are differently associated with nutrient intakes, lifestyle factors and energy misreporting in a sub-sample of the Malmo Diet and Cancer cohort. *Food Nutr Res*: DOI:10.3402/fnr.v53i0.1970.
38. Margetts BM, Thompson RL, Speller V *et al.* (1998) Factors which influence 'healthy' eating patterns: results from the 1993 Health Education Authority health and lifestyle survey in England. *Public Health Nutr* **1**, 193–198.
39. Martikainen P, Brunner E & Marmot M (2003) Socio-economic differences in dietary patterns among middle-aged men and women. *Soc Sci Med* **56**, 1397–1410.
40. Millen BE, Quatromoni PA, Copenhafer DL *et al.* (2001) Validation of a dietary pattern approach for evaluating nutritional risk: the Framingham Nutrition Studies. *J Am Diet Assoc* **101**, 187–194.
41. Correa Leite ML, Nicolosi A, Cristina S *et al.* (2003) Dietary and nutritional patterns in an elderly rural population in Northern and Southern Italy: (I). A cluster analysis of food consumption. *Eur J Clin Nutr* **57**, 1514–1521.
42. Davis MS, Miller CK & Mitchell DC (2004) More favorable dietary patterns are associated with lower glycemic load in older adults. *J Am Diet Assoc* **104**, 1828–1835.
43. Ledikwe JH, Smiciklas-Wright H, Mitchell DC *et al.* (2004) Dietary patterns of rural older adults are associated with weight and nutritional status. *J Am Geriatr Soc* **52**, 589–595.
44. James DC (2009) Cluster analysis defines distinct dietary patterns for African-American men and women. *J Am Diet Assoc* **109**, 255–262.
45. Delisle HF, Vioque J & Gil A (2009) Dietary patterns and quality in West-African immigrants in Madrid. *Nutr J* **8**, 3.
46. Pryer JA, Cook A & Shetty P (2001) Identification of groups who report similar patterns of diet among a representative national sample of British adults aged 65 years of age or more. *Public Health Nutr* **4**, 787–795.
47. Bamia C, Orfanos P, Ferrari P *et al.* (2005) Dietary patterns among older Europeans: the EPIC-Elderly study. *Br J Nutr* **94**, 100–113.
48. Quatromoni PA, Copenhafer DL, Demissie S *et al.* (2002) The internal validity of a dietary pattern analysis. The Framingham Nutrition Studies. *J Epidemiol Community Health* **56**, 381–388.
49. Tucker KL, Dallal GE & Rush D (1992) Dietary patterns of elderly Boston-area residents defined by cluster analysis. *J Am Diet Assoc* **92**, 1487–1491.
50. Schroll K, Carbajal A, Decarli B *et al.* (1996) Food patterns of elderly Europeans. SENECA Investigators. *Eur J Clin Nutr* **50**, 86–100.
51. Lin H, Bermudez OI & Tucker KL (2003) Dietary patterns of Hispanic elders are associated with acculturation and obesity. *J Nutr* **133**, 3651–3657.
52. Quatromoni PA, Copenhafer DL, D'Agostino RB *et al.* (2002) Dietary patterns predict the development of overweight in women: the Framingham Nutrition Studies. *J Am Diet Assoc* **102**, 1239–1246.
53. Flores M, Macias N, Rivera M *et al.* (2010) Dietary patterns in Mexican adults are associated with risk of being overweight or obese. *J Nutr* **140**, 1869–1873.
54. Carrera PM, Gao X & Tucker KL (2007) A study of dietary patterns in the Mexican-American population and their association with obesity. *J Am Diet Assoc* **107**, 1735–1742.
55. Millen BE, Quatromoni PA, Nam BH *et al.* (2004) Dietary patterns, smoking, and subclinical heart disease in women: opportunities for primary prevention from the Framingham Nutrition Studies. *J Am Diet Assoc* **104**, 208–214.
56. Millen BE, Quatromoni PA, Nam BH *et al.* (2002) Dietary patterns and the odds of carotid atherosclerosis in women: the Framingham Nutrition Studies. *Prev Med* **35**, 540–547.
57. Hlebowicz J, Persson M, Gullberg B *et al.* (2011) Food patterns, inflammation markers and incidence of cardiovascular disease: the Malmo Diet and Cancer study. *J Intern Med* **270**, 365–376.
58. Oliveira A, Rodriguez-Artalejo F, Gaio R *et al.* (2011) Major habitual dietary patterns are associated with acute myocardial

- infarction and cardiovascular risk markers in a southern European population. *J Am Diet Assoc* **111**, 241–250.
59. Lopez EP, Rice C, Weddle DO *et al.* (2008) The relationship among cardiovascular risk factors, diet patterns, alcohol consumption, and ethnicity among women aged 50 years and older. *J Am Diet Assoc* **108**, 248–256.
  60. Liese AD, Schulz M, Moore CG *et al.* (2004) Dietary patterns, insulin sensitivity and adiposity in the multi-ethnic Insulin Resistance Atherosclerosis Study population. *Br J Nutr* **92**, 973–984.
  61. Villegas R, Yang G, Gao YT *et al.* (2010) Dietary patterns are associated with lower incidence of type 2 diabetes in middle-aged women: the Shanghai Women's Health Study. *Int J Epidemiol* **39**, 889–899.
  62. Liu E, McKeown NM, Newby PK *et al.* (2009) Cross-sectional association of dietary patterns with insulin-resistant phenotypes among adults without diabetes in the Framingham Offspring Study. *Br J Nutr* **102**, 576–583.
  63. He Y, Ma G, Zhai F, Li Y *et al.* (2009) Dietary patterns and glucose tolerance abnormalities in Chinese adults. *Diabetes Care* **32**, 1972–1976.
  64. Chen H, Ward MH, Graubard BI *et al.* (2002) Dietary patterns and adenocarcinoma of the esophagus and distal stomach. *Am J Clin Nutr* **75**, 137–144.
  65. Bastos J, Lunet N, Peleteiro B *et al.* (2010) Dietary patterns and gastric cancer in a Portuguese urban population. *Int J Cancer* **127**, 433–441.
  66. Edefonti V, Randi G, Decarli A *et al.* (2009) Clustering dietary habits and the risk of breast and ovarian cancers. *Ann Oncol* **20**, 581–590.
  67. Tsai YY, McGlynn KA, Hu Y *et al.* (2003) Genetic susceptibility and dietary patterns in lung cancer. *Lung Cancer* **41**, 269–281.
  68. Rouillier P, Senesse P, Cottet V *et al.* (2005) Dietary patterns and the adenomacarcinoma sequence of colorectal cancer. *Eur J Nutr* **44**, 311–318.
  69. Millen BE, Quatromoni PA, Pencina M *et al.* (2005) Unique dietary patterns and chronic disease risk profiles of adult men: the Framingham nutrition studies. *J Am Diet Assoc* **105**, 1723–1734.
  70. Brunner EJ, Mosdol A, Witte DR *et al.* (2008) Dietary patterns and 15-y risks of major coronary events, diabetes, and mortality. *Am J Clin Nutr* **87**, 1414–1421.
  71. Leite ML & Nicolosi A (2009) Dietary patterns and metabolic syndrome factors in a non-diabetic Italian population. *Public Health Nutr* **12**, 1494–1503.
  72. Song Y & Joung H (2012) A traditional Korean dietary pattern and metabolic syndrome abnormalities. *Nutr Metab Cardiovasc Dis* **22**, 456–462.
  73. Tucker KL, Chen H, Hannan MT *et al.* (2002) Bone mineral density and dietary patterns in older adults: the Framingham Osteoporosis Study. *Am J Clin Nutr* **76**, 245–252.
  74. McNulty H & Scott JM (2008) Intake and status of folate and related B-vitamins: considerations and challenges in achieving optimal status. *Br J Nutr* **99**, 48–54.
  75. Gao X, Yao M, McCrory MA *et al.* (2003) Dietary pattern is associated with homocysteine and B vitamin status in an urban Chinese population. *J Nutr* **133**, 3636–3642.
  76. Knuops KT, Spiro A, de Groot LC *et al.* (2009) Do dietary patterns in older men influence change in homocysteine through folate fortification? The Normative Aging Study. *Public Health Nutr* **12**, 1760–1766.
  77. Brennan L (2008) Session 2: personalised nutrition Metabonomic applications in nutritional research. *Proc Nutr Soc* **67**, 404–408.
  78. Pencina MJ, Millen BE, Hayes LJ *et al.* (2008) Performance of a method for identifying the unique dietary patterns of adult women and men: the Framingham nutrition studies. *J Am Diet Assoc* **108**, 1453–1460.
  79. Bountziouka V, Tzavelas G, Polychronopoulos E *et al.* (2011) Validity of dietary patterns derived in nutrition surveys using *a priori* and *a posteriori* multivariate statistical methods. *Int J Food Sci Nutr* **62**, 617–627.
  80. Black AE, Prentice AM, Goldberg GR *et al.* (1993) Measurements of total energy expenditure provide insights into the validity of dietary measurements of energy intake. *J Am Diet Assoc* **93**, 572–579.
  81. Goldberg GR, Black AE, Jebb SA *et al.* (1991) Critical evaluation of energy intake data using fundamental principles of energy physiology: 1. Derivation of cut-off limits to identify under-reporting. *Eur J Clin Nutr* **45**, 569–581.
  82. Livingstone MB & Black AE (2003) Markers of the validity of reported energy intake. *J Nutr* **133**, 895S–920S.
  83. Pryer JA, Vrijheid M, Nichols R *et al.* (1997) Who are the 'low energy reporters' in the dietary and nutritional survey of British adults? *Int J Epidemiol* **26**, 146–154.
  84. Johansson L, Solvoll K, Bjorneboe GE *et al.* (1998) Under- and overreporting of energy intake related to weight status and lifestyle in a nationwide sample. *Am J Clin Nutr* **68**, 266–274.
  85. Mendez MA, Wynter S, Wilks R *et al.* (2004) Under- and overreporting of energy is related to obesity, lifestyle factors and food group intakes in Jamaican adults. *Public Health Nutr* **7**, 9–19.
  86. Mattisson I, Wirfalt E, Aronsson CA *et al.* (2005) Misreporting of energy: prevalence, characteristics of misreporters and influence on observed risk estimates in the Malmo Diet and Cancer cohort. *Br J Nutr* **94**, 832–842.
  87. Hornell A, Winkvist A, Hallmans G *et al.* (2010) Misreporting, previous health status and health status of family may seriously bias the association between food patterns and disease. *Nutr J* **9**, 48.
  88. Scagliusi FB, Ferrioli E, Pfrimer K *et al.* (2008) Under-reporting of energy intake is more prevalent in a healthy dietary pattern cluster. *Br J Nutr* **100**, 1060–1068.
  89. Bailey RL, Mitchell DC, Miller C *et al.* (2007) Assessing the effect of underreporting energy intake on dietary patterns and weight status. *J Am Diet Assoc* **107**, 64–71.