RESEARCH REPORT

The internal validity of a dietary pattern analysis. The Framingham Nutrition Studies

P A Quatromoni, D L Copenhafer, S Demissie, R B D'Agostino, C E O'Horo, B-H Nam, B E Millen

....

Study objectives: To examine the internal validity of a dietary pattern analysis and its ability to discriminate clusters of people with similar dietary patterns using independently assessed nutrient intakes and heart disease risk factors.

Design and participants: Population based study characterising dietary patterns using cluster analysis applied to data from the semiquantitative Framingham food frequency questionnaire collected from 1942 women ages 18–76 years, between 1984–88.

Setting: Framingham, Massachusetts.

Main results: Of 1942 women included in the cluster analysis, 1828 (94%) were assigned to one of the five dietary pattern clusters: Heart Healthy, Light Eating, Wine and Moderate Eating, High Fat, and Empty Calorie. Dietary patterns differed substantially in terms of individual nutrient intakes, overall dietary risk, heart disease risk factors, and predicted heart disease risk. Women in the Heart Healthy cluster had the most nutrient dense eating pattern, the lowest level of dietary risk, more favourable risk factor levels, and the lowest probability of developing heart disease. Those in the Empty Calorie cluster had a less nutritious dietary pattern, the greatest level of dietary risk, a heavier burden of heart disease risk factors, and a relatively higher probability of developing heart disease. Cluster reproducibility using discriminant analysis showed that 80% of the sample was correctly classified. The cluster technique was highly sensitive and specific (75% to 100%).

Conclusions: These findings support the internal validity of a dietary pattern analysis for characterising dietary exposures in epidemiological research. The authors encourage other researchers to explore this technique when investigating relations between nutrition, health, and disease.

See end of article for authors' affiliations

. . . .

Correspondence to:
Dr B E Millen, Department of Social and Behavioral Sciences, Boston University School of Public Health, 715 Albany Street, Boston, MA 02118, USA; bmillen@bu.edu.

Accepted for publication 3 August 2001

. . . .

With advancements in dietary assessment methodologies, many of the observed epidemiological associations between diet and chronic disease risk have been supported by clinical research and now inform our public policy recommendations for health promotion. None the less, nutrition researchers continue to critically evaluate investigative methods and strive for further enhancements that minimise the well established limitations of dietary assessment and analytical techniques.[1–6]

Food frequency questionnaires (FFQs) have facilitated the widespread investigation of diet and disease relations in large cohorts. With their generally established reliability and validity,[7–9] FFQs provide a cost effective method for assessing usual dietary intake over a specified period of time within a population. FFQ data have typically been used to rank individuals according to varying levels of nutrient intake exposure, enabling calculation of relative risk estimates. As exposures to nutrients are not independent,[3 5 10 11] interest in overall dietary patterns has emerged in the recent literature.[6 12] In fact, FFQs are now commonly used to characterise intake of food groups and to explore total dietary patterns in relation to cardiovascular and cancer risk.[13–19]

An analytical approach that characterises overall dietary patterns will provide an innovative strategy for assessing dietary exposures in chronic disease epidemiology.[3–5 10] Cluster analysis is one multivariate technique that has been successfully applied to nutrition data for characterising food intake patterns.[6 12] In this report, we discuss the use of cluster analysis applied to a FFQ that was administered to women in the Framingham Nutrition Studies. We examine the internal validity of the dietary patterns using two sets of independent criterion measures: (1) nutrient intake assessed by dietary records and (2) heart disease risk factor profiles.

METHODS

The Framingham Study was started in 1948 as a longitudinal population based study of cardiovascular disease. In 1971, a second generation cohort was enrolled, forming the Framingham Offspring/Spouse (FOS) study.[20] This cohort consisted of 5124 Framingham Study offspring, between the ages of 5 and 70, and their spouses if married.

Subjects and ethics

The data reported were collected among women in the Framingham Offspring/Spouse cohort at Exam 3 (1984–88). The data collection protocols and procedures were approved for human subjects by the Office of the Institutional Review Board at the Boston University School of Medicine, Boston Medical Center. At Exam 3, some 2005 women aged 18 to 76 years participated (83% of eligible). Dietary patterns were characterised using cluster analysis applied to food consumption data from the semiquantitative Framingham FFQ. Participants were grouped into one of five clusters based on food intake patterns identified from individual responses to a FFQ. The clusters were then compared for independent assessments of nutrient intake (based on separate three day dietary records) and cardiovascular risk factors (based on standardised Framingham exam protocols), as described in detail below.

Instruments

The self administered Framingham FFQ is a modified version of the original Willett questionnaire.[21] This instrument was validated in the Framingham Offspring cohort.[22] It contains 145 food items with seven non-overlapping response categories, ranging from "rarely or never use" to use "four or more times each day." Respondents reported how often, on average,

they consumed a standard portion of each food item during the past year. Reported frequencies were used to estimate the number of usual daily servings of each food item. Among women who came to FOS Exam 3, 97% completed the FFQ (n=1944). Of those, only two women (0.1%) were excluded from these analyses because of incomplete FFQ data.

Cluster analysis

Analytical details of the cluster technique were previously published and the identification of the dietary clusters has been described.[6] In brief, the 145 food item listings on the FFQ were first classified into 42 nutrient based food categories. The food categories were consistent with subgroups of foods defined by the American Dietetic Association's Exchange List for Meal Planning.[23 24] Food item listings in a particular category contained similar levels of macronutrients and other key nutrients (for example, all vitamin C rich fruits (>30 mg/serving) were grouped together). An estimate of the usual number of daily servings for each of the 42 food categories was derived for each study participant by summing across the category's component items, weighted by the individual frequency responses.

We used two procedures to identify dietary patterns. Firstly, VARCLUS [25] identified and grouped the 42 food categories that women consumed with a similar level of frequency. This SAS procedure is similar to a factor analysis, with the added requirement that food categories are separated into non-overlapping groups. The VARCLUS technique does not require that foods be eaten at the same time of day, at the same meal, or in similar quantities. Thus, foods that appear in the same cluster were consumed with a similar daily frequency (that is, relatively frequent consumption versus relatively infrequent consumption). For example, women who reported relatively higher daily intakes of fish also reported relatively higher daily intakes of other lower fat foods including skinless poultry, whole grains, and low fat dairy. Using this method, 13 food groupings were identified, each containing multiple food categories (see appendix).

In the second step, Ward's clustering method [26] was used to separate women into non-overlapping groups based upon similarities in their frequency of consumption of the food group clusters. This method considered how women differed in their consumption of the 13 food groupings and used the pseudo t^2 statistic, a criterion for identifying the optimal number of clusters of women with distinctive food consumption patterns. The pseudo t^2 statistic is plotted against the number of potential clusters, and when it changes little, it is concluded that adding clusters does not provide a better fit to the data.

Discriminant analysis

We evaluated the reproducibility of the dietary pattern clusters by examining their stability and classification ability using discriminant analysis. This analysis was performed on the five clusters using the 13 food groupings as discriminant variables and the 1828 subjects originally clustered as the sample. From the discriminant analysis, classification functions were developed and used to assign each subject into one of the five clusters. Subjects were then cross classified according to their original cluster classification and the one obtained from the discriminant analysis.

Nutrient intake analyses

We estimated mean nutrient intake levels among women in the clusters using three day dietary records as an independent assessment of nutrient intake. As previously published,[22] independent estimates of nutrient intake in the Framingham cohort are derived from either 24 hour dietary recalls or three day dietary records. In our sample, 1802 women provided 24 hour dietary recalls (90% of the cohort) and 1265 women

completed three day dietary records (63% of the cohort). Women who completed the dietary records were similar to those who did not complete them with respect to most heart disease risk factors. As estimates of nutrient intake derived from multiple days are considered more representative of usual intake than estimates from a single day,[22] we present findings based on the three day dietary records. We reviewed the dietary data based on three day dietary records in comparison with data derived from the 24 hour recalls. The analysis using recall data did not change our interpretation of the differences in nutrient profiles between the clusters as presented in this report.

Dietary records (two weekdays and one weekend day) were completed by participants using published research protocols.[22 27] Nutrient calculations were performed using the Minnesota Nutrition Data System (NDS) software (NDS version 2.6; Food Database 6A; Nutrient Database 23).[28] Three day mean nutrient intakes were determined for each individual. Macronutrient intakes were expressed as percentage of total calories whereas all other nutrients were expressed in absolute amounts (g, mg, µg).

Cardiovascular disease risk factor analyses

Risk factors for cardiovascular disease are routinely measured at all Framingham exams. Framingham protocols for risk factor measurement are standardised and have been summarised by Cupples *et al*.[29] The two year probability of developing coronary heart disease (coronary heart disease risk score) was estimated using a published Framingham algorithm.[30] This multivariate model is gender specific and includes the following risk factors for determining primary coronary heart disease risk: age, fasting lipid levels (total cholesterol, high density lipoprotein-cholesterol, and triglycerides), systolic blood pressure, smoking status, presence of left ventricular hypertrophy, diagnosis of diabetes, alcohol consumption (ounces per week), use of antihypertensive medication, postmenopausal status, and body mass index. The coronary heart disease risk score was calculated only for women who were free of heart disease at Exam 3 (n=1722).

Statistics

Analysis of variance was used to test for differences in food group consumption across clusters (table 1). Next, for each nutrient (table 2), we assigned ranks ranging from 1 to 1265 to the three day mean intake estimate for each of the 1265 women who completed the dietary records. Ranks were assigned so that an individual with a desirable nutrient intake level (that is, lower fat intake) received a lower rank, whereas an individual with an undesirable nutrient intake level (that is, lower fibre intake) received a higher rank. This facilitated a consistent interpretation linking lower risk with lower ranks and higher risk with higher ranks. For "risk nutrients" (total and saturated fat, cholesterol, alcohol, and sodium) and for energy, protein and monounsaturated fat, intakes were ranked from lowest (1) to highest (1265). For polyunsaturated fat and the "cardioprotective nutrients," intakes were ranked in the reverse order, from highest (1) to lowest (1265). In these analyses, monounsaturated fat was not considered a protective nutrient as the majority of its consumption by Framingham Offspring/Spouse participants was derived from animal products, which also contribute saturated fat, rather than from plant sources.[31] Age adjusted mean ranks were determined for each nutrient in each cluster, because age distributions differed between clusters.

Ranks of overall dietary risk were computed using the mean of the ranks of the individual variables in the nutrient list. Firstly, we computed the mean rank across the 19 nutrients for each of the 1265 women with dietary records. For a given individual, this mean of the ranks over 19 nutrients represented their overall nutrient rank. The age adjusted least

Table 1 Daily consumption of food groupings derived from the Food Frequency Questionnaire among Framingham women, by dietary cluster (n=1828). Mean number of servings per day and 95% CI

Food groupings	Clusters mean (95% CI)				
	Heart healthy n=366	Light eating n=872	Wine and moderate eating n=64	High fat n=370	Empty calorie n=156
Vegetables	**3.91[b] (3.70 to 4.11)**	2.59[a] (2.51 to 2.67)	3.16[a] (2.80 to 3.52)	2.75[a] (2.61 to 2.88)	**2.55[a] (2.33 to 2.77)**
Fruits and low fat milk	**4.33[b] (4.10 to 4.56)**	2.45[a] (2.36 to 2.54)	2.44[a] (2.07 to 2.82)	**2.39[a] (2.23 to 2.54)**	2.56[a] (2.31 to 2.80)
Other low fat foods	**5.07[c] (4.79 to 5.34)**	2.7[b] (2.63 to 2.87)	2.79[a,b] (2.36 to 3.22)	**2.44[a] (2.27 to 2.62)**	3.06[b] (2.69 to 3.42)
Legumes, soups, and miscellaneous foods	**0.36[b] (0.33 to 0.40)**	**0.19[a] (0.18 to 0.20)**	0.22[a] (0.18 to 0.26)	**0.19[a] (0.17 to 0.20)**	0.21[a] (0.18 to 0.24)
Refined grains, soft margarine and oils	3.42[a] (3.23 to 3.61)	**2.6[b] (2.51 to 2.71)**	2.86[b,c] (2.41 to 3.32)	3.66[b] (3.42 to 3.91)	2.96[c] (2.67 to 3.24)
Diet beverages and firm vegetable fats	**2.44[a] (2.24 to 2.64)**	3.37[b] (3.23 to 3.51)	3.49[b] (3.09 to 3.89)	4.17[a] (3.97 to 4.37)	3.54[b] (3.23 to 3.86)
Sweets and animal fats	1.12[b] (0.99 to 1.24)	**1.05[b] (0.98 to 1.12)**	1.75[c] (1.28 to 2.23)	4.45[a] (4.22 to 4.68)	2.49[d] (2.16 to 2.82)
Desserts	0.85[b] (0.77 to 0.93)	1.15[a] (1.06 to 1.23)	**0.58[b] (0.41 to 0.75)**	1.25[a] (1.14 to 1.35)	**1.30[a] (1.13 to 1.46)**
Sweetened beverages	0.30[a,b] (0.25 to 0.35)	0.27[b] (0.24 to 0.30)	**0.23[a,b] (0.15 to 0.31)**	0.34[a] (0.29 to 0.39)	**2.74[c] (2.58 to 2.90)**
Wine and cholesterol rich foods	0.43[b] (0.39 to 0.46)	0.46[b] (0.43 to 0.48)	**2.77[c] (2.66 to 2.89)**	0.36[a] (0.33 to 0.39)	**0.34[a] (0.30 to 0.39)**
High fat dairy and snack foods	0.69[b] (0.64 to 0.75)	0.78[b] (0.73 to 0.82)	**0.92[b] (0.76 to 1.08)**	**0.67[c] (0.62 to 0.71)**	0.79[a] (0.70 to 0.88)
Meats and mixed dishes	0.72[a] (0.67 to 0.76)	0.70[a] (0.68 to 0.73)	0.80[a] (0.69 to 0.92)	0.88[a] (0.82 to 0.94)	0.92[a] (0.84 to 0.99)
Fattier poultry and beer	0.12[b] (0.10 to 0.14)	0.20[a] (0.17 to 0.22)	0.15[a] (0.09 to 0.20)	0.14[a] (0.12 to 0.16)	0.16[a] (0.12 to 0.19)

[a,b,c,d]Means with the same letter are not significantly different from each other (p≥.05); means that are in **bold face** are significantly **highest**; means that are underlined and **bold** are significantly **lowest**.
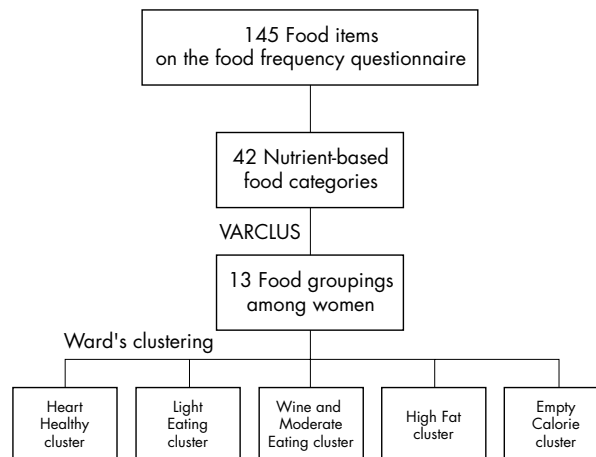

Figure 1 Cluster analysis.

square mean of these overall nutrient ranks was then computed for each cluster to provide a measure of overall dietary risk.

For each of the cardiovascular disease risk factors considered (table 3), each woman who participated in the cluster analysis (n=1828) was assigned a rank ranging from 1 to 1828. Here too, ranks were assigned so that a woman with a desirable risk factor level (higher high density lipoprotein-cholesterol) received a lower rank and a woman with an undesirable risk factor level (raised low density lipoprotein-cholesterol) received a higher rank. All but three of the cardiovascular disease risk factors were ranked from lowest (1) to highest (1828). The exceptions were the protective factors (physical activity, high density lipoprotein-cholesterol, and oestrogen use), which were reverse ranked, from highest (1) to lowest (1828) to coincide with a consistent interpretation of desirability. Age adjusted mean ranks were determined for each risk factor in each cluster.

The coronary heart disease risk score was ranked in a similar fashion to provide a measure of overall heart disease risk. This score was ranked in ascending order for all women who were free of heart disease at Exam 3 (n=1722). Finally, the age adjusted least square mean of the coronary heart disease risk score ranks was computed for each cluster.

RESULTS
Identifying dietary clusters
Five groups of women were identified by the cluster analysis, representing distinct segments of the population with unique dietary patterns (fig 1). Of the 1942 women included in the cluster analysis, 1828 (94%) were assigned cluster membership. The remaining 114 women (6%) were excluded because of extreme food behaviours that deviated markedly from the food consumption patterns of the five clusters that emerged. Dietary patterns were characterised by differences in levels of consumption of 11 of the 13 food groupings (table 1). Each cluster was assigned a label based on the key distinguishing features of its dietary pattern: Heart Healthy, Light Eating, Wine and Moderate Eating, High Fat, and Empty Calorie.

Compared with other clusters, women in the Heart Healthy cluster consumed notably more daily servings of vegetables, fruits and low fat milk, other low fat foods (non-fat and low fat dairy, skinless poultry, fish, whole grains, etc), and legumes, soups and miscellaneous foods (including other vegetarian products and shellfish). They also ate fewer servings of diet beverages and firm vegetable fats and had relatively lower intakes of sweets and animal fats, desserts, and high fat dairy and snack foods. Among the dietary patterns observed in this cohort, the Heart Healthy pattern was most consistent with current population based dietary guidelines for health promotion.[6]

Table 2   Age adjusted mean ranks of nutrient intake levels derived from three day dietary records, by cluster (n=1265), LS Mean

Nutrients	Clusters				
	Heart healthy n=267	Light eating n=618	Wine and moderate eating n=40	High fat n=247	Empty calorie n=93
Selected macronutrients					
Energy	666[a]	595[b]	688[a,b]	667[a]	675[a]
Protein	**590** [b]	612[b]	636[a,b]	701[a]	**718**[a]
Monounsaturated fat	**516**[b]	649[a]	661[a]	683[a]	**721**[a]
Polyunsaturated fat	631[a]	620[a]	642[a]	663[a]	645[a]
Risk nutrients					
Total fat	**520**[b]	648[a]	589[a,b]	698[a]	**704**[a]
Saturated fat	**496**[c]	642[b]	526[c]	**746**[a]	711[a,b]
Alcohol	613[a,b]	655[b]	1061[c]	577[a,d]	**515**[d]
Cholesterol	593[a]	633[a]	730[a]	660[a]	631[a]
Sodium	661[a]	605[a]	669[a]	649[a]	680[a]
Protective nutrients					
Carbohydrate	**535**[c]	671[b]	**917**[d]	608[a]	603[a,b,c]
Fibre	**471**[b]	676[a]	621[a]	671[a]	**718**[a]
Calcium	**550**[b]	663[a]	**701**[a]	621[a]	676[a]
Selenium	571[b]	656[a]	563[a,b]	652[a]	639[a,b]
Vitamin C	**496**[b]	658[a]	622[a,c]	674[a,c]	**753**[c]
Vitamin B6	**526**[b]	668[a,c]	556[a,b]	649[a,c]	**701**[c]
Vitamin B12	618[a]	634[a]	549[a]	649[a]	662[a]
Folate	**506**[b]	658[a]	596[a,b]	679[a]	**726**[a]
Vitamin E	**543**[b]	659[a]	594[a,b]	656[a]	**673**[a]
β carotene	**562**[b]	633[a]	637[a,b,c]	668[a,c]	**741**[c]
Overall dietary risk	**561**[b]	644[a]	661[a,c]	662[a,c]	**679**[c]

[a,b,c,d]Means with the same letter are not significantly different from each other (p≥.05); means that are in **bold face** are significantly **highest**; high ranks indicate undesirable nutrient intake levels; means that are **underlined** and **bold** are significantly **lowest**; low ranks indicate desirable nutrient intake levels.

Women in the Light Eating cluster chose fewer servings of legumes, soups and miscellaneous foods, refined grains, soft margarines and oils, and sweets and animal fats. Those in the Wine and Moderate Eating cluster consumed more servings of wine and cholesterol rich foods (eggs and organ meats) and high fat dairy and (salty and fatty) snack foods. Contributing to their otherwise moderate eating pattern, their intake levels of desserts and sweetened beverages were lower than other groups of women.

The High Fat dietary pattern was distinguished by substantially higher consumption of fats of all types (animal fats, firm and soft vegetable fats, and oils). Women in this cluster consumed more servings of sweets, refined grains, and diet beverages and relatively more servings of desserts. Their dietary pattern was notable for lower intakes of fruits and low fat milk, other low fat foods, legumes, soups and miscellaneous foods, high fat dairy products, and snack foods. The Empty Calorie pattern was characterised by markedly higher intakes of desserts and sweetened beverages with fewer servings of vegetables, wine, and cholesterol rich foods. The Empty Calorie label for this pattern reflects its relative lack of nutrient density despite its relatively higher caloric content, a result of the predominant food choices made by these women.

Ranking of nutrient intake

An analysis of age adjusted cluster mean ranks of independently assessed nutrient intake was one approach to validating the dietary pattern methodology (table 2). Dietary patterns differed significantly from one another on the basis of nutrient ranks for 15 of the 19 nutrients considered (79%). As well, the patterns differed on the composite measure of overall dietary risk. These individual and composite differences in nutrient intake ranks confirmed the placement of the clusters along a continuum of dietary risk, ranging from the most desirable Heart Healthy pattern to the least desirable Empty Calorie pattern.

It is notable that the Heart Healthy pattern was most desirable (lowest ranks) in terms of protein, total, saturated and monounsaturated fat, carbohydrate, fibre, calcium, vitamins C, B6, and E, folate, and β carotene intakes. Not surprisingly, this cluster displayed the lowest level of overall dietary risk. At the other end of the spectrum, the Empty Calorie dietary pattern was least desirable (highest ranks) for intakes of these same nutrients. The only exception was for saturated fat intake. However, the mean rank for saturated fat in the Empty Calorie cluster was not significantly different from the mean rank in the High Fat cluster, which was the absolute highest. On the positive side, women in the Empty Calorie cluster had the most desirable (lowest) rank for alcohol intake. The Empty Calorie dietary pattern achieved the highest level of overall dietary risk, anchoring its position at the high risk end of the continuum.

The ranking analysis also confirmed distinguishing nutrient profiles of other dietary patterns. The Light Eating cluster displayed the lowest rank for energy intake, consistent with its lower caloric density and overall pattern of lower food and nutrient intake relative to other clusters. The Wine and Moderate Eating cluster had the highest rank for alcohol and a relatively high rank for total calories. Women in this cluster also had the least desirable (lowest intake) levels of carbohydrate and calcium in their diets. Yet their overall micronutrient profiles and total and saturated fat intake levels were relatively closer to the desirable end of the spectrum than most other clusters. Finally, the High Fat cluster was distinguished by the least desirable (highest) intake of saturated fat. Ranks for other dietary fats, protein, and most micronutrients were relatively closer to the less desirable range.

Ranking of coronary heart disease risk factors

A second level of validation of the dietary pattern analysis was supported by the differences in cardiovascular disease risk factors observed between the five clusters (table 3). As with

Table 3 Age adjusted mean ranks of risk factor levels, by dietary cluster (n=1828), LS Mean

Risk factors	Heart healthy n=366	Light eating n=872	Wine and moderate eating n=64	High fat n=370	Empty calorie n=156
Mean age (SD)	**50.1[a] (10.0)**	48.2[a,b] (10.0)	47.7[b] (7.6)	47.4[b,c] (9.8)	**45.5[c] (10.0)**
Dichotomous outcomes					
Overweight*	929[a,b]	896[a]	**864[a]**	876[a]	**994[b]**
Obesity†	893[a,b]	904[a,b]	967[b,c]	**880[a]**	**976[c]**
Hypertension‡	951[b]	892[a]	973[a,b]	883[a]	972[b]
Diabetes§	915[a]	915[a]	935[a]	910[a]	913[a]
Current smoker	**815[c]**	889[b]	918[a,b,c]	1014[a,d]	**1045[d]**
Oestrogen use	904[a]	909[a]	900[a]	924[a]	922[a]
Continuous outcomes					
Body mass index (BMI)¶	918[b]	913[b]	899[a,b]	**842[a]**	**991[b]**
Total cholesterol	**810[b]**	881[a]	833[a,b]	895[a]	**916[a]**
HDL cholesterol	855[b]	828[b]	**633[c]**	964[a]	**989[a]**
T/H ratio**	816[b]	837[b]	**663[c]**	971[a]	**1000[a]**
LDL cholesterol	801[b]	854[b,c]	**754[b]**	930[a]	**916[a,c]**
Triglyceride	**837[a]**	852[a]	855[a]	893[a]	**1012[b]**
Systolic blood pressure	908[a,b]	919[b]	**1083[c]**	**858[a]**	970[b,c]
Diastolic blood pressure	937[b]	905[a,b]	**1091[c]**	**853[a]**	983[b,c]
Physical activity index††	**718[b]**	822[a]	754[a,b]	832[a]	**833[a]**
CHD risk score‡‡	**841[a]**	852[a,c]	**918[c,d]**	875[a,b,c]	**908[b,d]**

[a,b,c,d]Means with the same letter are not significantly different from each other (p≥.05); means that are in **bold face** are significantly **highest**; high ranks indicate undesirable risk factor levels; means that are **underlined and bold** are significantly **lowest**; low ranks indicate desirable risk factor levels. *Overweight defined as BMI ≥25,[32] therefore this variable includes those who are obese in its ranking scheme; †obesity defined as BMI ≥30 [32]; ‡hypertension defined as systolic blood pressure ≥140 mm Hg, diastolic blood pressure ≥90 mm Hg, or use of antihypertensive medication [29]; §diabetes defined as a history of use of insulin or oral hypoglycaemic agent, or a fasting blood glucose level ≥7.8 mmol/l (140 mg/dl) [29]; ¶BMI calculated as weight (kg) / height (m²) from height and weight measurements taken in the clinic; **T/H ratio is the ratio of total cholesterol to HDL-cholesterol; ††Physical Activity Index [33] measured at Exam 2 in the FOS cohort; ‡‡CHD Risk Score [30] is the average two year probability of developing coronary heart disease; (n=1722) includes only those women who were free of CHD at Exam 3.

the nutrient ranks, the analysis of age adjusted cluster mean ranks of risk factors and the estimated coronary heart disease risk score demonstrated a continuum of risk among the five groups of women. Not only did the clusters differ by age, but they also differed in terms of the predicted likelihood of developing coronary heart disease and in levels of 13 of the 15 risk factors considered (87%).

Heart Healthy women were again positioned at the most desirable end of the risk spectrum and Empty Calorie women maintained the highest risk status. Women who ate a Heart Healthy diet had the most desirable profiles for smoking, physical activity, and blood concentrations of total cholesterol and triglycerides. These women were the oldest, yet they had the lowest probability for developing coronary heart disease in the next two years. In contrast, women who exhibited the Empty Calorie dietary pattern had the least desirable profiles for BMI, overweight and obesity, smoking, physical activity, and dyslipidaemia in general. They were also relatively more likely to have undesirable ranks for blood pressure measures, hypertension prevalence, and low density lipoprotein-cholesterol levels. Though women in the Empty Calorie cluster were the youngest, they had a relatively higher likelihood of developing coronary heart disease in the next two years. Their risk was not substantially different from women in the cluster with the absolute highest predicted risk.

Relative to other groups, those in the Light Eating cluster had relatively lower to moderate ranks for most risk factors and for the coronary heart disease risk score. Women in the Wine and Moderate Eating cluster had the least desirable blood pressure levels and a relatively higher prevalence of hypertension. This clinical profile, along with their higher level of alcohol consumption, was a major contributor to their highest ranking on the coronary heart disease risk score. Levels of overweight and dyslipidaemia were more desirable among women in this cluster relative to other groups of women. While women in the High Fat cluster had less desirable levels of low density lipoprotein-cholesterol relative to other groups, they had more desirable profiles for BMI, obesity, hypertension, and blood pressure measures.

Reliability of the cluster analysis

Discriminant analysis correctly classified 80% of the sample. When each of the original clusters was considered separately, the sensitivity of the discriminant analysis ranged from 75% to 100% over the clusters as follows: 75% for the High Fat cluster, 78% for Heart Healthy, 80% for Light Eating, 90% for Empty Calorie, and 100% for Wine and Moderate Eating. Similarly, the specificity ranged from 75% to 100% over the five clusters as follows: 75% for the High Fat cluster, 89% for Light Eating, 90% for Heart Healthy, 98% for Empty Calorie, and 100% for Wine and Moderate Eating.

DISCUSSION

Using cluster analysis with food frequency data from a large population based cohort of women, we characterised unique patterns of dietary exposure within five distinct segments of the population. Groups of women whose dietary patterns were distinguished by cluster analysis differed from one another in terms of food group consumption, desirability of individual nutrient intakes, overall dietary risk, desirability of cardiovascular disease risk factors, and predicted risk for developing heart disease. In fact, the dietary patterns ranged along a continuum, from a low fat, nutrient dense "Heart Healthy" eating pattern to a higher fat, less nutritious "Empty Calorie" pattern. The continuum corresponded with both overall dietary risk and risk for the predicted development of coronary heart disease. These data support the internal validity of a dietary pattern analysis for characterising dietary exposures in epidemiological research. While our analyses were focused on heart disease risk factors and predicted coronary heart disease risk, it is believed that the utility of the dietary pattern approach would be comparable for exploring other chronic disease outcomes, including hypertension and cancer.[12 18 34] This is an area for further research.

Key points

- Multiple unique eating patterns were identified in a cohort of women.
- Dietary patterns display a range of nutritional and cardiovascular risk.
- Cluster analysis is a valid approach to dietary exposure measurement, enabling the exploration of relations between eating patterns, health, and disease.
- Dietary patterns are informative for the development and targeting of effective nutrition intervention messages.

Analytical approaches that define the food consumption patterns of people within a population have been used, albeit relatively rarely, in a variety of research settings since as early as 1981. Their popularity has increased in the past decade as investigators have realised the utility of grouping individual food items into interpretable eating patterns that lend themselves to epidemiological investigations.[12 35] Aggregating foods into meaningful food groupings or nutrient based categories (that is, fruits high in antioxidants) and then statistically clustering people on the basis of dietary behaviours is now considered to be an important feature of studies of diet and disease.[36] Recently, Pryer et al [12] used cluster analysis and recorded food intake data to identify four diet groups among men and four among women in Great Britian. These authors described important differences in nutrient profiles, sociodemographic characteristics, and behavioural features that are relevant for the development and targeting of health promotion strategies and public health nutrition policy.

While the validity of FFQs has been established for relative ranking of nutrient intake,[7 21 37] dietary pattern analyses using FFQ data are relatively recent. FFQs are attractive for this use because they effectively measure longer term usual diet and are cost efficient tools to use with large cohorts. Hu et al [38] used factor analysis and data from a FFQ to identify two predominant eating patterns, a prudent pattern and a Western pattern, within a small cohort of 127 men. The prudent pattern was similar to the Heart Healthy pattern observed among Framingham women, and the Western pattern was similar to the High Fat and Empty Calorie patterns described in our cohort. Recent prospective analyses involving 44 875 men followed up over eight years by Hu et al [19] demonstrated that the prudent and Western dietary patterns predict risk of coronary heart disease, independent of other lifestyle variables including smoking and body mass index.

These Framingham analyses lend support to the internal validity of the dietary pattern approach and provide further justification for its application in epidemiological research. The data of Hu et al[38] are important but are somewhat limited in scope because the validity of the dietary pattern analysis was established only at the level of food groupings using factor analysis. Our analyses identify the unique dietary patterns of individuals in the population and demonstrate the validity of this technique in comparison with two independent criteria.

Approaches that identify dietary patterns enable the examination of broader aspects of the diet, rather than single nutrient or food exposures, in relation to health outcomes.[3 39 40] In light of the complexities of the biological mechanisms known to cause disease, it seems important to understand relations within the context of the overall diet and to explore analytical techniques that discern patterns of intake that confer greater or lesser risk. This perspective is supported by recent findings in the literature where dietary pattern interventions, such as the DASH trial,[41] contributed to risk reduction in clinical settings.

Our five dietary clusters included eating patterns similar to the prudent and Western diets described by Hu et al [38] at the extremes of the risk continuum. Yet, differences between dietary patterns were not limited to a comparison of the two extremes within our sample. We were also able to show important differences in nutrient and risk factor profiles between patterns that fell in the middle of the continuum. One notable example is the Wine and Moderate Eating pattern, which was associated with higher blood pressure levels and higher predicted CHD risk.

We recognise that there is a level of non-specificity that accompanies dietary pattern analyses. Dietary patterns are inherently complex. Without more detailed analyses, they do not enable the specific identification of the particular nutrient or dietary components within the pattern that may be responsible for the observed differences in disease risk between population subgroups.[38] While the dietary pattern approach may be complex, the dietary patterns and clusters of people are in fact unique and non-overlapping. Each person belongs to only one cluster and the clusters are distinguished from each other by levels of food group consumption.

Thus, while we do not know which specific individual component(s) of the dietary patterns relate to better health outcomes or the relative importance of the components in conferring risk, we have begun to identify what those individual components are within the context of the overall eating pattern. We have also established that certain dietary patterns are associated with healthier profiles while others are associated with disease risk. Furthermore, we have identified personal characteristics (gender and age) and other behavioural characteristics (alcohol consumption, smoking habits, etc) that are closely linked to dietary behaviour. This information is likely to be beneficial for developing interventions based upon dietary pattern messages and may facilitate the targeting of intervention messages to relevant subgroups of the population.

With established validity, the dietary pattern approach offers an innovative analytical strategy for measuring dietary exposures in epidemiological research settings and in other populations. This approach may prove well suited for addressing nutrition research questions and offers an alternative to traditional methods that consider isolated dietary components as exposures of interest. Dietary pattern analyses may be especially useful when traditional analyses have failed to identify associations between individual nutrients and the outcome of interest, or to use the dietary pattern as a covariate in traditional single nutrient analyses to assess the effect of the exposure nutrient independent of the dietary pattern.[38] We demonstrated the internal validity and stability of a dietary pattern analysis and documented how the dietary behaviours of distinct groups of people relate directly to nutritional risk and heart disease risk. The availability of this analytical approach will probably increase the ability of investigators to explore relations between nutrition, health, and disease.

. . . .

Authors' affiliations
P A Quatromoni, C E O'Horo, B E Millen, Department of Social and Behavioral Sciences, Boston University School of Public Health, USA
D L Copenhafer, R B D'Agostino, B-H Nam, Department of Mathematics, Boston University, USA
S Demissie, The Data Coordinating Center, Boston University School of Public Health

Appendix Clustering of 42 food groups derived from the Food Frequency Questionnaire

Food group cluster	Food group
Vegetables	Vitamin A rich vegetables
	Vitamin C rich vegetables
	Higher fibre vegetables
	Other vegetables
Fruits and low fat milk	Vitamin C rich fruits
	Higher fibre fruits
	Other fruits
	Low fat dairy beverages
	Liquor and cordials
Other lower fat foods	Lower fat poultry
	Fish
	Whole grains
	Lower fat dairy products
	Lower fat health foods
	Non-fat dairy beverages
	Non-caloric decaffeinated beverages
Legumes, soups, and miscellaneous foods	Soups
	Legumes
	Shellfish
	High fat health foods
Refined grains, soft margarine, and oils	Refined grains
	Soft vegetable fats and oils
Diet beverages and firm vegetable fats	Firm vegetable fats
	Non-caloric caffeinated beverages
Sweets and animal fats	High fat dairy beverages
	Butter, creams, and gravies
	Lower fat sweets
Desserts	High fat desserts
	Lower fat desserts
	High fat sweets
Sweetened beverages	Sweet caffeinated beverages
	Sweet decaffeinated beverages
Wine and cholesterol rich foods	Organ meats
	Eggs
	Wine
High fat dairy and snack foods	High fat dairy and cheese
	Salty or fatty snacks
Meats and mixed dishes	High fat meats
	Lower fat meats
	Mixed protein dishes
Fattier poultry and beer	High fat poultry
	Beer

REFERENCES

1 **Kushi LH**. Gaps in epidemiologic research methods: design considerations for studies that use food-frequency questionnaires. *Am J Clin Nutr* 1994;**59** (suppl):180–4S.
2 **Kohlmeier L**. Gaps in dietary assessment methodology: meal- vs. list-based methods. *Am J Clin Nutr* 1994;**59** (suppl):175–9S.
3 **Baghurst KI**, Baghurst PA, Record SJ. Demographic and dietary profiles of high and low fat consumers in Australia. *J Epidemiol Community Health* 1994;**48**:26–32.
4 **Ursin G**, Zeigler RG, Subar AF, *et al*. Dietary patterns associated with a low-fat diet in the National Health and Nutrition Examination Follow-up Study: identification of potential confounders for epidemiologic analyses. *Am J Epidemiol* 1993;**137**:916–27.
5 **Patterson RE**, Haines PS, Popkin BM. Diet Quality Index: capturing a multidimensional behavior. *J Am Diet Assoc* 1994;**94**:57–64.
6 **Millen BE**, Quatromoni PA, Gagnon DR, *et al*. Dietary patterns of men and women suggest targets for health promotion: The Framingham Nutrition Studies. *Am J Health Promotion* 1996;**11**:42–52.
7 **Willett WC**, Reynolds RD, Cottrell-Hoehner S, *et al*. Validation of a semi-quantitative food frequency questionnaire: comparison with a 1-year diet record. *J Am Diet Assoc* 1987;**87**:43–7.
8 **Kristal AR**, Feng Z, Coates RJ, *et al*. Associations of race/ethnicity, education, and dietary intervention with the validity and reliability of a food frequency questionnaire: the Women's Health Trial Feasibility Study in Minority Populations. *Am J Epidemiol* 1997;**146**:856–69.
9 **Hankin JH**, Wilkens LR. Development and validation of dietary assessment methods for culturally diverse populations. *Am J Clin Nutr* 1994;**59** (suppl 1):198–200S.
10 **Randall E**, Marshall JR, Graham S, *et al*. Patterns in food use and their associations with nutrient intakes. *Am J Clin Nutr* 1990;**52**:739–45.
11 **Kant AK**, Schatzkin A, Block G, *et al*. Food group intake patterns and associated nutrient profiles of the US population. *J Am Diet Assoc* 1991;**91**:1532–7.
12 **Pryer JA**, Nichols R, Elliott P, *et al*. Dietary patterns among a national random sample of British adults. *J Epidemiol Community Health* 2001;**55**:29–37.
13 **Gillman MW**, Cupples AL, Gagnon D, *et al*. Protective effect of fruits and vegetables on development of stroke in men. *JAMA* 1995;**273**:1113–17.

14 **Gillum FR**, Mussolino ME, Madans JH. The relationship between fish consumption and stroke incidence. The NHANES I Epidemiologic Follow-up Study (National Health and Nutrition Examination Survey). *Arch Intern Med* 1996;**156**:537–42.
15 **Le Marchand L**, Hankin JH, Wilkens LR, *et al*. Dietary fiber and colorectal cancer risk. *Epidemiology* 1997;**8**:658–65.
16 **Franceschi S**, Favero A, Conti E, *et al*. Food groups, oils and butters, and cancer of the oral cavity and pharynx. *Br J Cancer* 1999;**80**:614–20.
17 **Ocke MC**, Bueno-de-Mesquita HB, Feskens EJ, *et al*. Repeated measurements of vegetables, fruits, beta-carotene, and vitamins C and E in relation to lung cancer. The Zutphen Study. *Am J Epidemiol* 1997;**145**:358–65.
18 **WCRF/AICR**. Food, nutrition and the prevention of cancer: a global perspective. Washington, DC: WCRF in association with American Institute for Cancer Research 1997. Available from the American Institute for Cancer Research, 1759 R Street, NW, Washington, DC 20009, USA.
19 **Hu FB**, Rimm EB, Stampfer MJ, *et al*. Prospective study of major dietary patterns and risk of coronary heart disease in men. *Am J Clin Nutr* 2000;**72**:912–21.
20 **Dawber TR**. *The Framingham Study. The epidemiology of atherosclerotic disease*. Cambridge MA: Harvard University Press, 1980.
21 **Willett WC**, Sampson L, Stampfer MJ, *et al*. Reproducibility and validity of a semiquantitative food frequency questionnaire. *Am J Epidemiol* 1985;**122**:51–65.
22 **Posner BM**, Martin-Munley SS, Smigelski C, *et al*. Comparison of techniques for estimating nutrient intake – the Framingham Study. *Epidemiology* 1992;**3**:171–7.
23 **The American Dietetic Association**. *Handbook of clinical dietetics*. New Haven, CT: Yale University Press, 1981:A3–5, F20–6.
24 **The American Diabetes Association, Inc, and the American Dietetic Association.** *Exchange lists for meal planning*. Chicago, IL: The American Dietetic Association, 1989.
25 **SAS Institute**. *SAS user's guide*, version 6, vol 1 and 2. 4th edn. Cary, NC: SAS Institute, 1989:1641–59.
26 **Ward JH**. Hierarchical grouping to optimize an objective function. *J Am Stat Assoc* 1963;**58**:236–44.

27 **Posner BM**, Smigelski C, Duggal A, *et al*. Validation of two-dimensional models for estimation of portion size in nutrition research. *J Am Diet Assoc* 1992;**92**:738–41.

28 **Schakel SF**, Sievert YA, Buzzard IM. Sources of data for developing and maintaining a nutrient database. *J Am Diet Assoc* 1988;**88**:1268–71.

29 **Cupples LA**, D'Agostino RB. Some risk factors related to the annual incidence of cardiovascular disease and death using pooled repeated biennial measurements: Framingham Heart Study, 30-year follow-up. In: Kannel WB, Wolf PA, Garrison RJ, eds. *The Framingham Study, an epidemiological investigation of cardiovascular disease*. Washington, DC: Department of Health and Human Services, 1987. NIH Publication 87–2703. (NTIS PB87–177499).

30 **D'Agostino RB**, Russell MW, Huse DM, *et al*. Primary and subsequent coronary risk appraisal: new results from the Framingham Study. *Am Heart J* 2000;**139**:272–81.

31 **Posner BM**, Cupples LA, Franz MM, *et al*. Diet and heart disease risk factors in adult American men and women: The Framingham Offspring-Spouse Nutrition Studies. *Int J Epidemiol* 1993;**22**:1014–25.

32 **The National Heart**, Lung, and Blood Institute Expert Panel on the Identification, Evaluation, and Treatment of Overweight and Obesity in Adults. Executive summary of the clinical guidelines on the identification, evaluation, and treatment of overweight and obesity in adults. *J Am Diet Assoc* 1998;**98**:1178–91.

33 **Kannel WB**, Sorlie P. Some health benefits of physical activity. The Framingham Study. *Arch Intern Med* 1979;**139**:857–61.

34 **Slattery ML**, Boucher KM, Caan BJ, *et al*. Eating patterns and risk of colon cancer. *Am J Epidemiol* 1998;**148**:4–16.

35 **Wirfalt AKE**, Jeffery RW. Using cluster analysis to examine dietary patterns: nutrient intakes, gender, and weight status differ across food pattern clusters. *J Am Diet Assoc* 1997;**97**:272–9.

36 **Huijbregts PPCW**, Feskens EJM, Kromhout D. Dietary patterns and cardiovascular risk factors in elderly men: The Zutphen Elderly Study. *Int J Epidemiol* 1995;**24**:313–20.

37 **Longnecker MP**, Lissner L, Holden JM, *et al*. The reproducibility and validity of a self-administered semiquantitative food frequency questionnaire in subjects from South Dakota and Wyoming. *Epidemiology* 1993;**4**:356–65.

38 **Hu FB**, Rimm E, Smith-Warner SA, *et al*. Reproducibility and validity of dietary patterns assessed with a food frequency questionnaire. *Am J Clin Nutr* 1999;**69**:243–9.

39 **Kant AK**. Indexes of overall diet quality: A review. *J Am Diet Assoc* 1996;**96**:785–91.

40 **Kant AK**, Schatzkin A, Harris TB, *et al*. Dietary diversity and subsequent mortality in the First National Health and Nutrition Examination Survey Epidemiologic Follow-up Study. *Am J Clin Nutr* 1993;**57**:434–40.

41 **Svetkey LP**, Simons-Morton D, Vollmer WM, *et al*. Effects of dietary patterns on blood pressure: subgroup analysis of the Dietary Approaches to Stop Hypertension (DASH) randomized clinical trial. *Arch Intern Med* 1999;**159**:285–93.