Published in final edited form as:

J Am Diet Assoc. 2008 September ; 108(9): 1453–1460. doi:10.1016/j.jada.2008.06.442.

Performance of a Method for identifying the Unique Dietary Patterns of Adult Women and Men: The Framingham Nutrition Studies

Michael J. Pencina¹, Barbara E. Millen², Laura J. Hayes¹, and Ralph B. D'Agostino¹

¹Statistics and Consulting Unit, Department of Mathematics and Statistics, Boston University, 111 Cummington St., Boston, MA 02215

²Department of Family Medicine and Graduate Medical Sciences Division, BU School of Medicine, One Boston Medical Center Plaza, Dowling 5S, Boston, MA 02118

Abstract

Background—Identification of population dietary patterns has been recommended by experts as a key to developing innovative and targeted nutrition interventions and achieving long-term dietary behavior change for health promotion and chronic disease risk reduction. Essential in this task is the evaluation of methods to accurately identify these unique dietary patterns.

Objective—To evaluate the validity and test the performance of a method for classifying adult men and women into one of five *a priori* dietary patterns.

Design—The first examination of Framingham Nutrition Studies took place between 1984 and 1987 and included 1,828 women and 1,666 men who completed the Framingham semi-quantitative food frequency questionnaire. Five unique dietary patterns for each gender were identified. Here we used Fisher's discriminant functions to create classification algorithms for identifying the dietary patterns of new individuals. Its validity and performance was evaluated using a variety of statistical tools.

Results—The new Framingham Dietary Pattern algorithm classified about 80% of women or men correctly and kappa statistics exceeded 0.70. The cluster-specific sensitivities ranged from 0.70 to 1.00 and specificities were all above 0.88. The pooled conditional c statistics were 0.95 and 0.96 for women and men, respectively. Overall, we can be confident that our methodology offers a valid identification of male and female dietary patterns when applied in practice.

Conclusions—The Framingham Dietary Pattern technique is a valid and reliable method for identifying the unique dietary behavior of adults. Our approach can also be used to guide the development and evaluation of other composite dietary quality indices.

Keywords

classification; cluster analysis; risk index; discriminant analysis; food frequency questionnaire

Address correspondence to: Michael J. Pencina, Statistics and Consulting Unit, Department of Mathematics and Statistics, Boston University, 111 Cummington St., Boston, MA 02215 phone: (617) 358-3386, fax: (617) 353-4767, e-mail: mpencina@bu.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

INTRODUCTION

The characterization of a population's unique dietary patterns has been advocated by experts as an important step in promotion of optimal food and nutrient intake and as a basic element to achieve more comprehensive approaches to disease prevention and health promotion (1–3). The dietary pattern approach is increasingly viewed as an innovative method in nutrition epidemiology because of its capacity to characterize food behavior and overall dietary quality. The features of the dietary pattern methodology have potentially important public health implications. Population dietary patterns have been increasingly widely linked to the emergence of specific diseases and morbidity and mortality in diverse populations. Current expert recommendations and recent multi-center clinical trials increasingly emphasize dietary patterns of importance in cancer, obesity, cardiovascular disease, and diabetes risk reduction. Furthermore, the food-related focus of the dietary patterns approach may enhance understanding of human behavior sufficient to enable more targeted, personalized nutrition interventions, educational programs, and public health campaigns.

In recent years, the Framingham Nutrition Studies identified and validated five unique dietary patterns of its adult male and female participants using statistical clustering techniques applied to a validated, semi-quantitative food frequency questionnaire. The methods have been described in detail by Millen et al. (3–5) and Quatromoni et al. (6) and noted by Hu (1) and elsewhere and summarized below.

The Framingham dietary patterns have been shown to be associated with the development of a variety of metabolic conditions and preclinical diseases (7–9). Sonnenberg et al. (7) reported that the prevalence of metabolic syndrome, abdominal obesity, low HDL cholesterol and high triglycerides is higher in women who follow the *empty calorie* pattern, while women in the *wine and moderate eating* group have the highest proportion of elevated blood pressure. Quatromoni et al. (8) showed that the *empty calorie* women have an increased risk of becoming overweight. Millen et al. (9) showed that the *heart healthy* eating pattern, in combination with the avoidance of smoking, leads to decreased odds of carotid atherosclerosis in women. Recently, Millen et al. (10) further demonstrated that nutritional risk, a composite index of dietary quality that is based upon intake levels of 19 nutrients and which varies across the dietary patterns, is predictive of the development of abdominal obesity and the overall metabolic syndrome in Framingham women.

In this article, we summarize the development of a method for classifying new adult male and female individuals into one of the five *a priori* Framingham Dietary patterns based on Fisher's discriminant functions (11,12). In addition, we utilize several performance measures to assess the validity of the Framingham Dietary Pattern classification algorithm.

SUBJECTS AND METHODS

The Framingham Heart Study was initiated in 1948 as a longitudinal population-based study of cardiovascular disease. More recently, other chronic diseases were added to the study design. The original Framingham cohort consisted of 5,209 men and women, between the ages of 28 and 62 at the beginning of the study, who represented a two-thirds random sample of the residents of Framingham, Massachusetts (13). In 1971, a second generation cohort was recruited and 5,124 Framingham Study offspring and their spouses were invited to participate in the Framingham Offspring/Spouse (FOS) study (14). Members of the FOS cohort participate in a standardized clinical assessment every four years, involving a complete physical exam, laboratory tests, non-invasive diagnostic testing, and updating of medical histories and other pertinent information. At certain exams, detailed dietary data are collected as part of the ongoing Framingham Nutrition Studies. The Boston University Medical Center's Human

Subjects Institutional Review Board approved the study protocol and all participants provided written informed consent.

This report uses data collected at Exam 3 of the FOS cohort between 1984 and 1987. During this period, some 1,868 men and 2,005 women participated in FOS Exam 3 (about 75% of the original FOS cohort - Exam I participants), and 1,666 (89%) men and 1,828 (91%) women completed the Framingham food frequency questionnaire (FFQ) (15). The FFQ was used to help characterize dietary patterns among the participating men and women, and was modified from a prototype antecedent to the validation of the original Willett food frequency questionnaire (16) and subsequently validated as a measure of nutrient intake within Framingham (15). The FFQ contains 145 food items and seven, non-overlapping response categories, which range from “rarely or never use” to use “four or more times each day.” Subjects were asked to estimate how often, on average, they consume a standardized portion of a given food item. The reported frequencies were used to estimate the subjects’ number of usual daily servings of each food item (3).

Summary of Previous Findings: Construction of 13 Food Group Clusters

A diagram outlining the construction of the food group clusters and resulting unique dietary patterns is provided in Figure 1. The 145 separate FFQ items were classified into 42 nutrient-based groupings of foods, which were similar to, but more detailed than, those found in the American Dietetic Association Exchange List for Meal Planning (e.g., vitamin A-rich vegetables, vitamin C-rich fruits, lower-fat meats, higher-fat meats, etc). An estimated frequency for each of the 42 groups was derived by summing the frequencies of its component foods.

Dietary pattern analyses were carried out separately for men and women, as it was assumed that dietary patterns would differ by gender (6). Indeed, early analyses in men suggested that alcohol preference (beer, wine or liquors/spirits) hindered clear distinctions in patterns of food intake so we eliminated these beverages from dietary pattern assessment in men. The determination of dietary patterns was completed in two steps. First, the 42 nutrient-based food groups were analyzed for similarities in reported intake frequency (rarely/never to 4 or more times a day). For example, food groupings consumed with rare/never to high frequency were aggregated statistically using the VARCLUS procedure in SAS (17). This method is similar to a factor analysis (18) of the 42 food categories, with the added requirement that foods are divided into non-overlapping groups. As an example, it was found that both men and women who reported relatively rare/never to the highest intakes of Vitamin A-rich vegetables also tended to consume “vitamin-C rich,” “higher in fiber” and “other” vegetables with similar patterns of frequency; these foods were statistically clustered into a “Vegetables” group based upon these similarities. It is further noted that the FFQ does not query the time of day or type of meal so these features were not considered. VARCLUS identified 13 food groupings for women and 13 for men (See Figure 2).

Summary of Previous Findings: Emergence of 5 Unique Dietary Patterns

The second step identified subgroups of men and women who had similar intakes of the 13 food groups. Ward’s clustering method (19) was used for this step, and was implemented in the Cluster procedure in SAS (17). A criterion at this step is that an individual cannot be clustered into more than one subgroup. Five subgroups of Framingham men and five subgroups of Framingham women emerged from these analyses, each of which represents a distinct dietary pattern within the cohort. They have been described in detail elsewhere (3–5). Here we give a brief summary to provide a better context for our research question. Among the 1828 women, there were notable differences in the daily intakes of all 13 food groups, except red meats, and fattier poultry and beer. Five unique patterns were identified. The first pattern,

labeled by us *higher fat* made up 20% of the women, with $n = 370$, and was characterized by higher mean daily consumption of diet beverages, vegetable fats, breads, and sweets and other fats. They also had a lower mean daily consumption of lower-fat foods than other clusters. The second pattern, *lighter eating*, was the largest cluster of women, with $n = 872$, comprising 48% of the population. This pattern consumed the lowest daily amounts of sweets, other fats, and bread and margarine. The third pattern, labeled *heart healthier* made up 20% of the women, with $n = 366$, and they consumed more vegetables, lower-fat foods, fruits, lower-fat dairy products, and soups. They also consumed fewer diet beverages and vegetable fats on a daily basis than the other patterns. The fourth pattern, *wine and moderate eating*, was the smallest group, making up only 3.5% of the population, with $n = 64$. This group consumed more high-fat dairy foods and snack items and more wine and cholesterol-rich foods, such as eggs and organ meats, on a daily basis. However, they also consumed lower amounts of desserts. Lastly, pattern five, with $n = 156$, included 8.5% of the women who consumed about ten times more sweetened beverages than the other women, daily. They also had the highest intakes of fats, oils and sweets. We labeled them *empty calorie* (3).

The original results on the 1666 men were found to be alcohol driven, so they were re-clustered, omitting alcohol. The results from this second analysis presented 5 unique patterns. The *empty calorie* group ($n=423$, 25.4% of all men) was characterized by the highest intake of sweets, salty snacks, high-fat animal protein foods, and refined grains and the lowest intake of low-fat milk. These men also had the highest level of saturated fat consumption and an overall nutritional risk score that exceeded that of the other male dietary patterns. The second pattern, which we labeled *lower variety*, had the lowest consumption of most food groups ($n=503$, 30.2%). The *higher starch* pattern consisted of 183 men (11% of the sample) who consumed higher levels of leaner protein foods (low-fat red meats and poultry), firm vegetable fats, refined grains and desserts. About one fifth of the sample ($n=341$) was classified into what we labeled the *transition to heart healthy* pattern, with the highest intakes of vegetables, fruits, whole grains, oils and lower-fat foods, soup, organ meats, various types of fish, and carbohydrates. Also, their intakes of refined grains, desserts, saturated fat, and total fat were the lowest. The *average male* pattern ($n=216$, 13%) was characterized by high intakes of diet and decaffeinated beverages and lowest consumption of leaner proteins (3,4).

Classification of new individuals

To apply this approach to other populations for applications in nutrition research, preventive interventions, and clinical treatment settings, it is imperative to be able to classify new adult males and female individuals into one of the five *a priori* Framingham Dietary patterns described above. As the process described above was based on data collected on 1828 and 1666 women and men, it is not possible to repeat it for each new individual for whom we seek to establish a dietary pattern. Hence, a simplified approach which borrows from the process described above is necessary.

We first note that an FFQ is needed for the new individual in question. Then the 145 responses can be grouped into the 42 categories. These can be further classified into the 13 clusters, following the VARCLUS results described above. However, it is not possible to repeat Ward's clustering based on one individual's data. Hence, we need a simplified method that can classify a person into one of the 5 unique dietary patterns based on the 13 clusters that can be obtained from individual data.

To do this, we apply the discriminant analysis procedure (SAS DISCRIM) to the 13 food groupings presented in Figure 2 in order to derive a statistical scoring system (algorithm) that would classify a new subject into one of five dietary patterns. The result of this procedure provides a simple linear Fisher classification function or formula for each of the five patterns. These formulas, unlike the results of Ward's clustering, can be used to score a new individual's

FFQ – the higher the score, the higher the likelihood of belonging to a particular pattern. An individual is assigned to a pattern with the highest classification function score.

Performance of Classification Algorithm

Since the new classification algorithm is intended to detect a dietary pattern membership for each new individual, it is imperative to assess its performance. The true pattern membership was based on original Ward's clustering and compared with the membership assigned based on the new algorithm. Since including a given individual in the creation of a function intended to classify this individual might yield over-optimistic results (and is not realistic in practice), we used a technique called cross-validation to improve the credibility of our results. Cross-validation omits the data from the individual of interest in the construction of this individual's functions, resulting in a more objective comparison.

The performance metrics included the simple and weighted kappa statistics calculated for the cross-tabulation of true and algorithm-based pattern memberships (20,21). The kappa statistic provides an assessment of agreement between the "true" and algorithm-based pattern memberships, based on the difference in observed and expected numbers on the diagonal of our cross-tabulation (20). Kappa ranges from -1 to 1 and values exceeding 0.75 represent excellent agreement, while values between 0.4 and 0.75 can be classified as fair to good (21). Kappa equal to zero indicates the amount of agreement that would be obtained by pure chance. If the nature of the categories or dietary patterns were such that misclassification into one category had more detrimental consequences than misclassification into another category, the weighted kappa could be used to assess the level of agreement (22).

We also computed five individual sensitivities and specificities, each measuring the classification ability related to one of the five clusters. Since the last two measures work with binary variables, we dichotomized our five cluster categories five times, at each time i defining membership or classification to cluster i as 1 and any other as 0. We also calculated the weighted average of the sensitivities with weights corresponding to the sample sizes in each pattern, which is equal to the proportion of individuals classified correctly (the ratio of the total number of people placed in the cluster to which they truly belong and the total number of people; it is also equal to one minus the misclassification fraction). Confidence intervals were obtained using asymptotic methods for the kappa statistics (23) and bootstrap for the other measures (24).

A different assessment of the classification algorithm can be obtained using the method described by Obuchowski et al. (25). They define a c-statistics-based index (and present a method for calculating its confidence interval) which calculates the probability that given two subjects, one from cluster i and the other from cluster j , the one from i has a higher probability of being classified as belonging to i .

Our simple classification algorithms are meant to be used for individuals outside of the Framingham developmental cohort. On the other hand, it is advantageous to base them on as many individuals as possible to improve their stability. That is why the algorithms described thus far were based on all individuals available. However, to detect the degree of over-optimism that exists despite cross-validation, we randomly split our sample in the 2:1 ratio, and used the larger portion to develop a classification algorithm and the remainder to evaluate its performance. All steps described above were repeated, separately for the evaluation of algorithms in women and men and degree of over-optimism was quantified as a difference between the results on developmental and validation (evaluation) sets.

RESULTS

Classification algorithm and its performance

The relationship of the 13 food groupings and 5 unique dietary patterns obtained through discriminant analysis was in good agreement with the results of Ward's clustering. The relative magnitudes of discriminant coefficients for dietary patterns for each of the 13 food groupings are depicted in Figure 3a (women) and Figure 3b (men). In particular, in women, vegetable intake had the highest coefficient for the heart healthy pattern and so did fruits and low-fat milk as well as soups and miscellaneous. Sweets and fats pointed towards the high fat pattern and so did red meats. Sweetened beverages had the highest coefficient for empty calorie and wine and cholesterol-rich foods for the wine and moderate eating pattern.

In men, vegetables characterized the transition to heart healthy pattern and so did fruits and high fat animal proteins. Desserts and whole milk, eggs and animal fats had the highest coefficients for the empty calorie pattern while firm vegetable fats were pointing towards the higher starch pattern. Diet and decaffeinated beverages pointed towards the average male pattern.

The cross-tabulations of true and algorithm based cluster memberships based on the Fisher's classification functions are presented in Tables 1a and 1b. The simple kappa statistic for the women's algorithm was 0.709 with 95% confidence interval (CI) (0.682, 0.736) indicating good agreement between our algorithm and the original clustering (21). Sensitivities and specificities with corresponding 95% confidence intervals for classification into individual clusters versus all other clusters are presented in Table 2a. Overall proportion classified correctly reached 0.798 with 95% CI (0.781, 0.821). Not surprisingly, the classification measures for individual clusters indicated better ability to identify the smaller and more uniquely defined clusters – *wine and moderate eating and empty calorie*. The pooled conditional c statistic reached 0.953 with 95% CI (0.946, 0.960) indicating very high probability of classifying a woman to the correct one of any two clusters.

The results for the men's algorithm were very similar. The simple kappa statistic almost reached excellent classification agreement with a point estimate of 0.742 and 95% CI (0.717, 0.767). Overall proportion classified correctly reached 0.800 with 95% CI (0.777, 0.819). The individual sensitivities, shown in Table 2b, were somewhat more variable than in women, with *empty calorie* and *transition* clusters being identified correctly least often. The pooled conditional c statistic reached 0.962 with 95% CI (0.956, 0.968) indicating very high probability of classifying a man to the correct one of any two clusters.

When the analyses were repeated after splitting the original sample in the 2:1 ratio and the performance metrics were re-calculated on the developmental and validation sets, the resulting magnitude of over-optimism in the kappa statistics was estimated not to exceed 0.02 in the proportion of subjects classified correctly. There is relatively more difference in the individual sensitivities, but it never exceeds 0.05. Overall, we can be quite confident that the good performance displayed by our algorithms will carry to individuals from similar populations.

Possible practical applications

Objective methods of dietary assessment, for example the FFQ, provide a valuable tool in dietetics. However, the variety of foods consumed necessitates large number of questions in these instruments. This in turn makes it difficult to summarize the food intake and pattern of an individual who completed the FFQ. The ability of classifying this person into one unique pattern offers the dietician a simple starting tool in designing the optimal diet plan for the person in question. Our algorithm facilitates such classification. For example, based on Sonnenberg et al. (7) it can be concluded that women in the wine and moderate eating tend to have higher

blood pressure levels. Thus, if the woman in question has elevated blood pressure and gets classified into the wine and moderate eating, we might be able to suggest a dietary intervention that would move this person towards a different pattern. Similar argument can be made for abdominal obesity and metabolic syndrome and the empty calorie pattern.

It is also possible to automate the classification algorithm using simple spreadsheets (for example, Microsoft Excel) and make it available to dietitians and the general public, together with explanations of what health risks and benefits can result from following a given pattern and what dietary modifications should be considered.

DISCUSSION

This paper describes the development and validation of a simple classification algorithm (scoring system) to identify the unique dietary patterns of new adult male and female individuals. Several methods of performance assessment for our algorithm and for algorithms of this type in general are suggested, provided that the goal is defined as a correct identification of membership in one of several categories.

Several points merit particular attention. The Framingham Dietary Pattern algorithms have been shown to perform very well in correctly classifying new individuals from the same population. The performance assessment is most appropriately extended to people who can be believed to follow a diet that resembles that of the FNS participants. It has been previously reported that our population characteristics and dietary behaviors and trends parallel those seen in nationally-representative samples of the US population (FNS papers). However, given that Framingham subjects are largely white, non-minority individuals, the performance of this dietary pattern approach needs to be further demonstrated in more diverse populations. Of interest are multiethnic studies which demonstrate considerable overlap in the dietary patterns of diverse populations (26). Also, our dietary patterns were created based on food frequency questionnaire data collected in the 1980s; it is possible that changes have occurred in male and female dietary patterns since then. However, it has been reported that the contrasting features of the dietary patterns and their related health characteristics of Framingham men and women have persisted over 15 or more years of follow-up (10).

Identification of an individual's dietary patterns depends on dietary data collection instrument and the statistical approach utilized to develop the referent population's dietary patterns. Hu (1) summarizes three general options for dietary pattern assessment that include statistical factor and cluster analysis (terms a posteriori approaches) and the application of a priori dietary quality indices. Researchers may choose to develop population-specific methods using the approaches outlined in this report. However, it may be possible to replicate our dietary patterns using food frequency questionnaires that are identical to ours as well as those that are similar such as the Willet questionnaire (16).

This report presents a variety of methods for evaluating the performance of classification algorithms. Even though, in our case, they all lead to similar conclusions, indicating very good general performance, in different circumstances the focus might be different and thus some, and not other of the methods presented here, will be of particular interest. Of note, there is one important difference in assumptions between the kappa statistic and the other methods discussed (sensitivities, proportion classified correctly, Obuchowski's conditional c statistic). The latter assume an existence of a gold standard; in our context it means that the clusters represent true reality. The kappa statistic, on the other hand, makes no assumptions of this kind; to the contrary, its definition does not require or imply that the original clustering procedure is closer to the truth than the discriminant analysis presented here. In our case neither set of assumptions fully represents the situation – the clusters do not represent a gold standard in the

same sense in which a cancer or cardiovascular event is diagnosed, but they are the standard on which the classification would be based. Hence, dietary patterns as a gold standard might be closer to reality in our application.

The Framingham Dietary Pattern methodology can serve as a useful aid for nutrition researchers, clinicians, and public health specialists in advancing methods for examining human nutrition behaviors and their relation to health outcomes and development of innovative strategies for population nutritional intervention and disease prevention. They may also have applications for direct consumer use in health promotion and nutrition education campaigns. But above all, the hope of the authors is, that the approach presented here will encourage nutritional experts in devising further algorithms to better characterize the dietary consumption in the general population and will help set the standards for the evaluation of their performance.

Acknowledgements

This work was supported by the National Heart, Lung, and Blood Institute's Framingham Heart Study (contract N01-HC-25195).

REFERENCES

1. Hu FB. Dietary pattern analysis: A new direction in nutritional epidemiology. *Curr Opin Lipidol* 2002;13:3–9. [PubMed: 11790957]
2. Jacques PF, Tucker KL. Are dietary patterns useful for understanding the role of diet in chronic disease. *Am J Clin Nutr* 2001;73:1–2. [PubMed: 11124739]
3. Millen BE, Quatromoni PA, Gagnon DR, Cupples LA, Franz MM, D'Agostino RB. Dietary patterns of men and women suggest targets for health promotion: the Framingham Nutrition Studies. *Am J Health Promot* 1996;11:42–52. [PubMed: 10163450]
4. Millen BE, Quatromoni PA, Pencina MJ, Kimokoti R, Nam BH, Cobain S, Kozak W, Appugliese DP, Ordovas J, D'Agostino RB. Unique Dietary Patterns and Chronic Disease Risk Profiles of Adult Men: The Framingham Nutrition Studies. *J Am Diet Assoc* 2005;105:1723–1734. [PubMed: 16256756]
5. Millen BE, Quatromoni PA, Copenhafer DL, Demissie S, O'Horo CE, D'Agostino RB. Validation of a dietary pattern approach for evaluating nutritional risk: the Framingham Nutrition Studies. *J Am Diet Assoc* 2001;101:187–194. [PubMed: 11271691]
6. Quatromoni PA, Copenhafer DL, Demissie S, D'Agostino RB, O'Horo CE, Nam BH, Millen BE. The internal validity of a dietary pattern analysis. The Framingham Nutrition Studies. *J Epidemiol Community Health* 2002;56:381–388. [PubMed: 11964437]
7. Sonnenberg L, Pencina MJ, Kimokoti R, Quatromoni PA, Nam BH, D'Agostino RB, Meigs JB, Ordovas J, Cobain MR, Millen BE. Dietary Patterns and the Metabolic Syndrome in Obese and Non-obese Framingham Women. *Obes Res* 2005;13:153–162. [PubMed: 15761175]
8. Quatromoni PA, Copenhafer DL, D'Agostino RB, Millen BE. Dietary Patterns Predict the Development of Overweight in Women: The Framingham Nutrition Studies. *J Am Diet Assoc* 2002;102:1240–1246.
9. Millen BE, Quatromoni PA, Nam BH, O'Horo CE, Polak JF, Wolf PA, D'Agostino RB. Dietary Patterns, Smoking and Subclinical Heart Disease in Women: Opportunities for Primary Prevention from the Framingham Nutrition Studies. *J Am Diet Assoc* 2004;104:208–214. [PubMed: 14760568]
10. Millen BE, Pencina MJ, Kimokoti RW, Z L, Meigs JB, Ordovas JM, D'Agostino RB. Nutritional risk and the metabolic syndrome in women, opportunities for preventive intervention from the Framingham Nutrition Study. *Am J Clin Nutr* 2006;84:434–441. [PubMed: 16895895]
11. Fisher RA. The use of multiple measurements in taxonomic problems. *Annals of Eugenics* 1936;7:179–188.
12. Armitage, P.; Colton, T., editors. *Encyclopedia of Biostatistics*. vol. 2. Wiley: Chichester; 1998.
13. D'Agostino, RB.; Kannel, WB. *Epidemiological background and design: The Framingham Study*. Proceedings of the American Statistical Association Sesquicentennial Invited Paper Sessions; American Statistical Association; Alexandria, VA. 1989.

14. Kannel WB, Feinleib M, McNamara PM, Garrison RJ, Castelli WP. An investigation of coronary heart disease in families: the Framingham Offspring Study. *Am J Epidemiol* 1979;110:281–290. [PubMed: 474565]
15. Posner BM, Martin-Munley SS, Smigelski C, Cupples LA, Cobb JL, Schaefer E, Miller DR, D'Agostino RB. Comparison of techniques for estimating nutrient intake: the Framingham Study. *Epidemiology* 1992;3:171–177. [PubMed: 1576223]
16. Willett WC, Sampson L, Stampfer MJ, Rosner B, Bain C, Witschi J, Hennekens CH, Speizer FE. Reproducibility and validity of a semiquantitative food frequency questionnaire. *Am J Epidemiol* 1985;122:51–65. [PubMed: 4014201]
17. SAS user's guide, version 6. 4th ed.. vols 1& 2. Cary (NC): SAS Institute; 1989. SAS Institute, Inc.; p. 1641-1659.
18. Cureton, EE.; D'Agostino, RB. Factor analysis: an applied approach. Hillsdale (NJ): Lawrence Erlbaum Associates; 1983. p. 1983
19. Ward JH. Hierarchical grouping to optimize an objective function. *J Am Stat Assoc* 1963;58:236–244.
20. Cohen J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 1960;20:37–46.
21. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–174. [PubMed: 843571]
22. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin* 1968;70:213–220.
23. Fleiss JL, Cohen J, Everitt BS. Large sample standard errors of kappa and weighted kappa. *Psychological Bulletin* 1969;72:323–327.
24. Carpenter J, Bithell J. Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. *Statist. Med* 2000;19:1141–1164.
25. Obuchowski NA, Goske MJ, Applegate KE. Assessing physicians' accuracy in diagnosing paediatric patients with acute abdominal pain: measuring accuracy for multiple diseases. *Statist. Med* 2001;20:3261–3278.
26. Park SY, Wilkens LR, Yamamoto JF, Sharma S, Hankin JH, Henderson BE, Kolonel LN. Dietary Patterns using the Food Guide Pyramid groups are associated with sociodemographic and lifestyle factors: the multiethnic cohort study. *J. Nutr* 2005;135(4):843–849. [PubMed: 15795445]

Figure 1.
Flow Diagram for Original Cluster Analysis

Women		Men	
1. <i>Vegetables</i>	7. <i>Fruits and Low-Fat Milk</i>	1. <i>Vegetables</i>	7. <i>Whole grains, oils and lower-fat foods</i>
Vitamin A-rich vegetables	Medium-fat dairy beverages	Vitamin A-rich vegetables	Whole grains
Vitamin C-rich vegetables	Vitamin C-rich fruits	Vitamin C-rich vegetables	Low-fat dairy products
Higher-fiber vegetables	Higher-fiber fruits	Higher-fiber vegetables	Low-fat poultry
Other vegetables	Other fruits	Other vegetables	Oils
2. <i>Diet Beverages and Vegetable Fat</i>	Liquors	2. <i>Refined grains and desserts</i>	Soft vegetable fats
Firm vegetable fats	8. <i>Sweetened Beverages</i>	Refined grains	Low-fat health foods
Noncaloric caffeinated beverages	Sweet caffeinated beverages	High-fat desserts	8. <i>Leaner proteins and diet beverages</i>
3. <i>Sweets and Fats</i>	Sweet decaffeinated beverages	Lower-fat desserts	Low-fat meats
High-fat dairy beverages	9. <i>Soups and Miscellaneous</i>	Lower-fat sweets	High-fat poultry
Butter and creams	Shellfish	3. <i>Fruits</i>	Energy-free caffeinated beverages
Lower-fat sweets	Legumes	Vitamin C-rich fruits	9. <i>Sweets and salty snacks</i>
4. <i>Desserts</i>	High-fat health foods	Higher-fiber fruits	High-fat sweets
High-fat desserts	Soups	Other fruits	Sweetened caffeinated beverages
Lower-fat desserts	10. <i>Bread and Margarine</i>	4. <i>Desserts</i>	Sweetened decaffeinated beverages
High-fat sweets	Refined grains	High-fat desserts	Salty snacks
5. <i>Other Lower-Fat Foods</i>	Soft vegetable fats	Lower-fat desserts	10. <i>Whole milk, eggs and animal fats</i>
Lower-fat poultry	11. <i>High-Fat Dairy and Snacks</i>	High-fat sweets	High-fat dairy beverages
Fish	High-fat dairy and cheese	5. <i>High-fat animal proteins</i>	Eggs
Whole grains	Salty or fatty snacks	High-fat dairy products	Animal fats
Lower-fat dairy products	12. <i>Fattier Poultry and Beer</i>	High-fat meats	11. <i>Low-fat milk</i>
Lower-fat health foods	Higher-fat poultry	Lower-fat meats	Low-fat dairy beverages
Lower-fat beverages	Beer	Mixed protein dishes	Nonfat dairy beverages
Noncaloric decaffeinated beverages	13. <i>Wine and Cholesterol-Rich Foods</i>	6. <i>Shellfish and vegetarian foods</i>	12. <i>Diet and decaffeinated beverages</i>
6. <i>Red Meats</i>	Organ meats	Shellfish	Diet and decaffeinated beverages
High-fat meats	Eggs	Legumes	13. <i>Firm vegetable fats</i>
Lower-fat meats	Wine	High-fat health foods	Firm vegetable fats
Mixed dishes			

Figure 2.
Descriptions of Food Groupings for Framingham Women and Men

Figure 3.

Figure 3a: Relative magnitude of discriminant coefficients for dietary patterns among the food groupings in Women

Abbreviations: dt = diet, bev = beverages, lf = low fat, hf = high fat, mar = margarine, misc = miscellaneous, choles = cholesterol

Figure 3b: Relative magnitude of discriminant coefficients for dietary patterns among the food groupings in Men

Abbreviations: ref = refined, des = desserts, lf = low fat, hf = high fat, gr = grain, prof = protein, dt = diet, bev = beverages, veg = vegetable

Table 1

Table 1a Cross-tabulation of True and Algorithm-based Pattern Memberships in Women									
Table of actual vs. classification-based dietary pattern									
Dietary Pattern		Classified into							
Frequency		high fat	light eating	heart healthy	wine and moderate	empty calorie	Total		
high fat		279	49	34	0	8	370		
light eating		57	691	110	2	12	872		
heart healthy		15	55	285	2	9	366		
wine and moderate		0	0	0	64	0	64		
empty calorie		8	6	3	0	139	156		
Total		359	801	432	68	168	1828		
Table 1b Cross-tabulation of True and Algorithm-based Pattern Memberships in Men									
Table of actual vs. classification-based dietary pattern									
Dietary Pattern		Classified into							
Frequency		empty calorie	lower variety	higher starch	transition	average male	Total		
empty calorie		294	56	24	34	15	423		
lower variety		38	425	6	30	4	503		
higher starch		9	4	162	3	5	183		
transition		24	37	12	248	20	341		
average male		0	4	4	5	203	216		
Total		365	526	208	320	247	1666		

Table 2

Table 2a Performance Measures for Specific Pattern vs. All Other Patterns: Women						
Pattern	high fat	light eating	heart healthy	wine/moderate	empty calorie	
Sensitivity 95% CI	0.754 (0.719, 0.800)	0.792 (0.762, 0.822)	0.779 (0.743, 0.828)	1.000 (1.000, 1.000)	0.891 (0.848, 0.937)	
Specificity 95% CI	0.945 (0.934, 0.958)	0.885 (0.871, 0.908)	0.899 (0.883, 0.915)	0.998 (0.995, 1.000)	0.983 (0.976, 0.988)	
Table 2b Performance Measures for Specific Pattern vs. All Other Patterns: Men						
Pattern	empty calorie	lower variety	higher starch	transition	average male	
Sensitivity 95% CI	0.695 (0.657, 0.749)	0.845 (0.793, 0.869)	0.885 (0.842, 0.938)	0.727 (0.678, 0.778)	0.940 (0.900, 0.969)	
Specificity 95% CI	0.943 (0.929, 0.955)	0.913 (0.895, 0.927)	0.969 (0.962, 0.977)	0.946 (0.928, 0.957)	0.970 (0.963, 0.979)	