European Journal of Clinical Nutrition (2009) 63, 707–717

0954-3007/09 $32.00
www.nature.com/ejcn

ORIGINAL ARTICLE

Associations between food patterns defined by cluster analysis and colorectal cancer incidence in the NIH–AARP diet and health study

E Wirfält[1], D Midthune[2], J Reedy[2], P Mitrou[2,3], A Flood[4], AF Subar[2], M Leitzmann[2], T Mouw[2], AR Hollenbeck[5], A Schatzkin[2] and V Kipnis[2]

[1]*Nutrition Epidemiology, Department of Clinical Sciences (Malmö), Lund University, Malmö, Sweden;* [2]*National Cancer Institute, National Institutes of Health, Bethesda, MD, USA;* [3]*Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK;* [4]*Division of Epidemiology, School of Public Health, University of Minnesota, Minneapolis, MN, USA and* [5]*AARP, Washington, DC, USA*

Background/Objectives: To examine associations between food patterns, constructed with cluster analysis, and colorectal cancer incidence within the National Institutes of Health–AARP Diet and Health Study.
Subjects/Methods: A prospective cohort, aged 50–71 years at baseline in 1995–1996, followed until the end of 2000. Food patterns were constructed, separately in men ($n = 293\,576$) and women ($n = 198\,730$), with 181 food variables (daily intake frequency per 1000 kcal) from a food frequency questionnaire. Four large clusters were identified in men and three in women. Cox proportional hazards regression examined associations between patterns and cancer incidence.
Results: In men, a vegetable and fruit pattern was associated with reduced colorectal cancer incidence (multivariate hazard ratio, HR: 0.85; 95% confidence interval, CI: 0.76, 0.94), when compared to less salutary food choices. Both the vegetable and fruit pattern and a fat-reduced foods pattern were associated with reduced rectal cancer incidence in men. In women, a similar vegetable and fruit pattern was associated with colorectal cancer protection (age-adjusted HR: 0.82; 95% CI: 0.70, 0.95), but the association was not statistically significant in multivariate analysis.
Conclusions: These results, together with findings from previous studies support the hypothesis that micronutrient dense, low-fat, high-fiber food patterns protect against colorectal cancer.
European Journal of Clinical Nutrition (2009) **63,** 707–717; doi:10.1038/ejcn.2008.40; published online 6 August 2008

Keywords: food patterns; cluster analysis; colorectal cancer; prospective cohort

Introduction

Epidemiological research suggests that dietary factors may both protect against and promote the development of colorectal cancer. High intakes of fiber, folate and calcium have been associated with reduced colorectal cancer risk (Giovannucci, 2002; Bingham *et al*., 2003; Norat and Riboli, 2003; Larsson *et al*., 2006), and high intakes of meat and fat with increased risk (Giovannucci *et al*., 1992; Norat *et al*., 2005). Experts argue that because of the multifaceted nature of diet–disease associations, traditional multivariate analysis may be an inefficient approach in nutrition epidemiology (Schatzkin *et al*., 1995; Jacques and Tucker, 2001). Because foods are consumed together, and dietary components act in synergism or are metabolized jointly, it can be argued that the true effect of diet may only be observed when all components are considered simultaneously. Also, analysis of dietary data and interpretation of diet–disease associations are hampered by the difficulties in separating out individual dietary components and adequately modeling their potential interactions (Byers and Gieseker, 1997).

Patterning methodologies, including cluster analysis (CA), factor analysis (FA) and diet quality indexes, may turn the analytical difficulties into an advantage (Hu, 2002; Kant, 2004; Newby and Tucker, 2004). CA, which aggregates individuals with similar characteristics (Aldenderfer and

Correspondence: Dr E Wirfält, Nutrition Epidemiology, Department of Clinical Sciences (Malmö), Clinical Research Centre, Lund University, Entrance 72, Malmö SE-205 02, Sweden.
E-mail: elisabet.wirfalt@med.lu.se
Received 22 February 2008; revised 17 June 2008; accepted 30 June 2008; published online 6 August 2008

Blashfield, 1984) has successfully been applied in epidemiology (Hulshof *et al*., 1992; Tucker *et al*., 1992; Greenwood *et al*., 2000; Wirfält *et al*., 2001; Engeset *et al*., 2005), but only a few CA studies have examined food patterns and colorectal health (Rouillier *et al*., 2005; Austin *et al*., 2007). This study examines associations between food pattern clusters and colorectal cancer incidence in the National Institutes of Health (NIH)–AARP (AARP is formely known as the American Association of Retired Persons) Diet and Health Study. In a series of papers, the same group of researchers is currently investigating different ways of constructing food patterns and their associations with colorectal cancer incidence (Flood *et al*., 2008; Reedy *et al*., 2008). A forthcoming paper will discuss and compare the experiences of this CA study with other approaches.

Methods

Cohort establishment and follow-up

The NIH–AARP Diet and Health Study was established in 1995–1996 (Schatzkin *et al*., 2001). A total of 340 148 men and 227 021 women above 50 years of age, residing in six states (California, Florida, Pennsylvania, New Jersey, North Carolina and Louisiana) and two metropolitan areas (Atlanta, GA and Detroit, MI), adequately completed a 16-page mailed questionnaire. The study protocol was approved by the Special Studies Institutional Review Board of the US National Cancer Institute, and all subjects provided their informed consent upon entry.

Vital status is ascertained annually through linkage of the cohort to the Social Security Administration Death Master File in the US, follow-up searches of the National Death Index Plus for participants who matched to the Social Security Administration Death Master File, cancer registry linkage, and responses to questionnaires and other mailings. The design and maintenance of this cohort have been described in detail elsewhere (Schatzkin *et al*., 2001).

Study sample

In this analysis, we excluded individuals with prevalent cancer (43 341 men and 26 048 women), end-stage renal disease (626 men and 371 women) at baseline and those reporting extreme energy intakes (2566 men and 1835 women) defined as being below the 25th percentile minus two interquartile ranges or above the 75th percentile plus two interquartile ranges of energy intake on the logarithmic scale. In preliminary CA with 100 clusters (performed twice), we also identified individuals with extreme food intakes; individuals in small clusters (less than 10 individuals) were removed (39 men and 37 women). The final sample was 293 576 men and 198 730 women.

Ascertainment of cancer cases

Incident cases of cancer were identified by linkage between the NIH–AARP cohort membership and cancer registry databases of the eight targeted states, which are estimated to be 95% complete within 2 years of cancer diagnosis and certified by the North American Association of Central Cancer Registries for meeting the highest standard of data quality (Michaud *et al*., 2005). Incident colorectal cancer cases were defined according to the *International Classification of Disease-Oncology*, 3rd edn. (codes C180–C189, C260, C199 and C209). A total of 2151 men and 959 women were diagnosed with primary incident colorectal cancer during the 4.5-year period from the baseline examinations (1995–1996) until the end of 2000. In men 631 cases were diagnosed with rectal cancer, and 1539 with colon cancer. In women 258 cases were diagnosed with rectal cancer, and 707 with colon cancer. Person years of observation accumulated from the date of study entry until the date of colorectal cancer diagnosis, or until censoring at the date of cancer diagnosis at another site, death, migration out of the study areas, or until 31 December 2000, whichever occurred first.

Baseline questionnaire

The baseline food frequency (AARP-FFQ) questionnaire was an early version of the new Diet History Questionnaire (DHQ) of the National Cancer Institute that has undergone extensive cognitive testing during development (Subar *et al*., 1995, 2001a, b). The AARP-FFQ was evaluated against two 24-h dietary recalls in the calibration substudy of 2000 men and women and demonstrated a satisfactory relative validity (Thompson *et al*., 2008). The energy-adjusted validity coefficients were in men for protein 0.43, carbohydrate 0.71, fat 0.72, and fruit and vegetables 0.72; in women for protein 0.50, carbohydrate 0.64, fat 0.62, and fruit and vegetables 0.61. The energy-adjusted attenuation factors were lowest for protein in both men (0.26) and women (0.31) and highest for saturated fat in men (0.68) and for vitamin B6 in women (0.62). The baseline questionnaire included 124 food items with 10 frequency response categories (that is, never; 1–6 times per year; 7–11 times per year; once per month; 2–3 times per month; 1–2 times per week; 3–4 times per week; 5–6 times per week; once per day; and twice or more per day) and 3 portion size alternatives. In addition, 21 questions requested frequency information on intake of low-fat and high-fiber foods and food preparation, and two crosschecking questions asked about the overall consumption of vegetables and fruits. The questionnaire, designed for the general population, includes some regional and ethnic group-specific foods, and three items on the type, frequency and dosage of supplement use. The reference period was the last 12 months. The energy and nutrient intakes were calculated by applying the food frequency and portion size information to the nutrient composition database that was newly derived from national survey data; CSFII, US Department of Agriculture (Subar *et al*., 2000). This study examined intakes of adjusted for energy using the density method.

A total of 204 food frequency variables were available in the database. We reduced these variables to 181 by collapsing those indicating different ways of eating butter and margarines into five variables (that is, butter, stick margarine, tub margarine, butter–margarine mixture and diet margarine), and noncaloric sweeteners (that is, aspartame and saccharine) into one variable. Two of the original food variables (that is, 'other fruits' and 'other vegetables') were excluded due to no reported consumption.

We used energy-adjusted food frequency variables (that is, food frequencies per 4.19 MJ and day) in order to concentrate on dietary proportions, and to reduce measurement error common in food frequency questionnaires (Willett *et al*., 1997; Kipnis *et al*., 2003). To remove the extraneous effect of variables with large variances on formation of clusters we also standardized the energy-adjusted food variables to a mean of zero and standard deviation of one.

The baseline questionnaire included questions on demographics and potential cancer risk factors. The following variables were used in this study: age; education (high school or less; completed high school; some college; college degree and higher); ethnicity (white; black; other); smoking (never; former, <20 cigarettes per day; former, ⩾20 cigarettes per day; current, ⩽20 cigarettes per day; current, >20 cigarettes per day), leisure time physical activity (never or rarely; 1–3 times per month; 1–2 times per week; 3–4 times per week; 5 or more times per week), body mass index (BMI, kg/m^2) computed from self-reported weights and heights (<18.5; 18.5<25.0; 25.0<30.0; 30.0<35.0; 35.0<40.0; ⩾40.0); and in women only menopausal hormone therapy (MHT; never use; current use; past use). An indicator variable for missing responses in each covariate was used, if applicable.

Statistical analysis

We used SAS version 8.1. (SAS Institute, Inc., Cary, NC, USA) for all statistical analyses. Statistical tests were two sided with significance levels equal to 0.05 and all analytical procedures were conducted separately for men and women. CA was performed using a *k*-means method, an iterative partitioning procedure that attempts to group the data into *k* clusters in such a way as to maximize the overall R^2 value, defined as $R^2 = 1 - W/T$, where W is the sum of squared Euclidean distances between each data point and its within-cluster mean (or center), and T is the sum of squared distances between each data point and the overall mean (Aldenderfer and Blashfield, 1984). The *k*-means methodology is recommended when working with large data sets, and have previously been used in a large number of diet-chronic

Table 1 Food and nutrient characteristics of the four largest food pattern clusters in men of the NIH–AARP Diet and Health Study 1995–2000

	Many foods (N = 176 127)	*Vegetables and fruits* (N = 81 318)	*Fatty meats* (N = 22 756)	*Fat-reduced foods* (N = 11 273)
$R^2 \geq 0.2$				Mayo, nf sand Mayo, nf salad
R^2 0.19–0.10		Apple, Banana, Broccoli, Carrot, Lettuce, Tomato, raw	Liver	Cheese, nf, salad dressing, nf, skim milk, cer
R^2 0.09–0.05		Fish, nf nfa, chicken, l ns, green beans, grapefruit, grapes, dried fruit, orange, pasta, peppers, salad dressing, lf, spinach, raw, tomato, sauce nm, vegetable medley, cauliflower and so on	Beef stew, chicken, fr d ws, chicken, fr l ws, cold cut, reg, gravy, ham, hot dog, reg, roast beef, reg, roast beef, sand, sausage, reg	Cracker, lf, frozen yogurt
Total energy (kcal)	1982	1620	2111	1704
Fat (% energy)	32.3	25.6	36.7	26.5
Protein (% energy)	14.7	15.9	16.4	16.8
Carbohydrate (% energy)	49.0	57.1	45.3	54.8
Sweets (% energy)	12.4	6.8	11.9	8.1
Alcohol (% energy)	2.1	1.70	0.86	1.27
Fiber (g per 1000 kcal)	8.8	13.3	8.7	11.7
Calcium (mg per 1000 kcal)	355	418	326	426
Folate (μg per 1000 kcal)	145	204	141	180
Vitamin C (mg per 1000 kcal)	62.5	105	60.7	81.5
ω3/ω6 ratio	0.11	0.12	0.11	0.11
P/S ratio	0.69	0.87	0.67	0.85
M/S ratio	1.22	1.29	1.23	1.32
Trans-FA (g per 1000 kcal)	2.41	1.78	2.86	2.05

Abbreviations: cer, with cereal; fa, fat added; FA, factor analysis; l, light; lf, low fat; nf, no fat; nfa, no fat added; nm, no meat; ns, no skin; reg, regular; sand, sandwich.
Food variables with R^2 values greater and equal to 0.05 are presented separately for each cluster: the highest ranking energy-adjusted food frequencies, and the median intakes of total energy and energy-adjusted nutrients.

disease studies (Kant, 2004; Newby and Tucker, 2004). Clustering was based on the 181 energy-adjusted and standardized food frequency variables for $k=3$–12 clusters. A final number of clusters was chosen based on the stability of large clusters ($n>10\,000$) that were formed, and on the overall R^2 values. When plotting the R^2 values against the number of clusters, six clusters for men and nine clusters for women accounted for most of the increase in R^2 and ensured three stable large clusters for each gender. Four clusters in men and three in women were used in subsequent analyses.

The distributions of relative food frequencies and the medians of total energy and energy-adjusted nutrient variables were examined across clusters. The distribution of common risk factors for colorectal cancer was examined by χ^2 analysis.

The Cox proportional hazards regression (Cox, 1972), with time since entry into the study as the time scale, was used to examine the association between clusters and incidence of colorectal cancer, colon cancer and rectal cancer. The largest cluster (labeled 'Many foods' in both men and women) was used as the reference category. Three models were fit for each cancer end point. The first model included only cluster (categorical) and age (continuous) as covariates. The second model also included BMI, and the third, multivariate, model adjusted in addition for education, ethnicity, smoking, leisure time physical activity and log total energy (continuous), and MHT in women. We also assessed the potential impact of dietary fiber, folic acid and calcium intakes, but as results did not change materially, these nutrients were not included in our final models.

Results

The food and nutrient characteristics of clusters are described in Tables 1 and 2 (also see Appendices 1 and 2 for detailed description of clusters). In men, four clusters, with more than 10 000 individuals, emerged. For the largest cluster 'Many foods,' the CA procedure did not indicate any specific distinguishing food, but intakes of alcohol and sweets ranked comparatively high. The second largest cluster ('Vegetable and fruit') was characterized by high intakes of vegetables, fruits, and low-fat foods like fish and lean chicken. This pattern was lowest in fat and the densest in micronutrients. The third largest cluster ('Fatty meats') was characterized by regular fat meats. The fourth largest cluster ('Fat-reduced foods') was characterized by fat-reduced foods (but not lean meats), with skim milk ranking comparatively high. Specific food items (that is, pumpkin pie, custard pie, lard, bacon and eggs) influenced the formation of the two smallest clusters.

In women, three of the nine clusters had more than 30 000 individuals, whereas six clusters had fewer than 10 000 individuals. Similar to men, no specific food emerged as the distinguishing feature for the largest cluster ('Many foods'), but sweets ranked comparatively high. Although the second largest cluster 'Vegetables and fruits' had similar characteristics to the 'Vegetables and fruits' cluster in men, skim milk with cereals and yogurt also ranked high in that cluster in women. Alcohol intakes were lower overall in women than in men, but appear to rank higher both in the 'Vegetables and fruit' and the 'Many foods' clusters. Different

Table 2 Food and nutrient characteristics of the three largest food pattern clusters in women of the NIH–AARP Diet and Health Study 1995–2000

	Many foods (N = *87 109*)	*Vegetables and fruits* (N = *64 671*)	*Diet foods, lean meats* (N = *32 426*)
$R^2>0.2$			Fr Chicken, lns
R^2 0.19–0.10		Apple, broccoli, carrot, lettuce, salad dressing, nf, tomato, raw	Mayo, sand diet
R^2 0.09–0.05		Banana, beans, green, cantaloupe, cauliflower and so on, chicken, lns, fish, nf nfa, grapefruit, orange, pasta, peppers, raw spinach, sal dressing, lf, skim milk, w cer, strawberry, tomato, sauce nm, yogurt	Cold cut, lf, diet margarine, diet mayo, sal, hamburger, lean, hotdog, lf, meatloaf, roast beef, lean
Total energy (kcal)	1541	1308	1495
Fat (% energy)	32.8	24.4	30.9
Protein (% energy)	14.3	16.1	16.9
Carbohydrate (% energy)	51.9	60.0	52.2
Sweets (% energy)	12.7	6.1	8.2
Alcohol (% energy)	0.53	0.60	0.36
Fiber (g per 1000 kcal)	9.2	14.4	11.2
Calcium (mg per 1000 kcal)	384	483	410
Folate (μg per 1000 kcal)	155	222	173
Vitamin C (mg per 1000 kcal)	74.4	122	84.8
ω3/ω6 ratio	0.12	0.13	0.11
P/S ratio	0.71	0.89	0.82
M/S ratio	1.17	1.24	1.29
Trans-FA (g per 1000 kcal)	2.38	1.63	2.37

Abbreviations: cer, with cereal; fa, fat added; FA, factor analysis; l, light; lf, low fat; nf, no fat; nfa, no fat added; nm, no meat; ns, no skin; reg, regular; sand, sandwich; sc, sauce.
Food variables, with R^2 values greater and equal than 0.05, are presented separately for each cluster: the highest ranking energy-adjusted food frequencies, and the median intakes of total energy and energy-adjusted nutrients.

Table 3 Distribution (%) of some baseline characteristics* in men by four food pattern clusters in the NIH–AARP Diet and Health Study 1995–2000

	Many foods	*Vegetables and fruits*	*Fatty meats*	*Fat-reduced foods*
		Frequency (%)		
Total	176 127 (60)	81 318 (28)	22 756 (8)	11 273 (4)
Age (years)				
Below 55	25 635 (67)	8 702 (23)	2 487 (6)	1251 (3)
55–69	144 819 (59)	69 000 (28)	19 334 (8)	9604 (4)
70 or above	5673 (53)	3616 (34)	935 (9)	418 (4)
Education				
High school or less	42 048 (67)	14 903 (24)	6112 (10)	2132 (3)
College graduate	70 968 (54)	45 454 (35)	8180 (6)	5464 (4)
Ethnicity				
Black	4573 (58)	1522 (19)	1480 (19)	171 (2)
White	163 916 (60)	75 136 (28)	19 965 (7)	10 794 (4)
Smoking				
Never	48 285(56)	28 430(33)	6486 (7)	3164 (4)
Former	98 538 (59)	47 498 (29)	11 674 (7)	7101 (4)
Current	23 194 (76)	2595 (9)	3565 (12)	642 (2)
Physical activity				
Never, rarely	30 474 (70)	7180 (16)	4380 (10)	1266 (3)
Once or more per week	117 867 (56)	66 011 (32)	14 896 (7)	8827 (4)
BMI (kg/m^2)				
Less than 25	45 249 (55)	27 860 (34)	5622 (7)	3572 (4)
25–30	86 473 (61)	38 823 (27)	10 663 (8)	5189 (4)
30 or more	40 107 (66)	12 608 (21)	5733 (9)	2237 (4)

Abbreviation: BMI, body mass index.
**P*-values <0.0001 for all examined variables in χ^2-tests comparing differences across all categories and clusters.

diet foods and lean meats characterized the third largest cluster in women ('Diet foods and lean meats'). Similar to men the formation of the smallest clusters was driven by frequent consumption of specific foods (that is, several types of pie or chicken, shortening, lard or liver).

Tables 3 and 4 show the within-cluster distributions of some potential risk factors for colorectal cancer. In men, the 'Vegetable and fruit' cluster was associated with being older, more educated, more likely to have never smoked, more physically active and less obese than the total sample, while the 'Many foods' cluster was associated with being younger, less educated, more likely to have smoked, less physically active and more obese. Similar tendencies were seen for the comparable clusters in women. The 'Diet foods and lean meats' cluster in women was associated with obesity, but the 'Fat-reduced foods' cluster in men was not. MHT use appeared more common among women of the 'Vegetable and fruit' cluster.

Hazard ratio (HR) estimates for colorectal cancer incidence are shown in Table 5 for clusters with more than 10 000 individuals. Smaller clusters had too few cases to give meaningful estimates. In men, the 'Vegetable and fruit' cluster was statistically significantly associated with reduced colorectal cancer incidence when compared to the 'Many foods' cluster; the association remained significant after multivariate adjustment (HR: 0.85; 95% confidence interval, CI: 0.76, 0.94). In women, the 'Vegetable and fruit' cluster was statistically significantly associated with reduced colorectal cancer incidence in the age- and BMI-adjusted models (HR: 0.83; 95% CI: 0.72, 0.97), but not in the multivariate model.

When analyses were repeated for colon and rectal cancer as separate end points (Table 6), both the 'Vegetable and fruits' (HR: 0.74; 95% CI: 0.60, 0.91), and the 'Fat-reduced foods' (HR: 0.56; 95% CI: 0.34, 0.95) clusters in men were inversely associated with rectal cancer after multivariate adjustment for other risk factors. The 'Vegetable and fruits' cluster was also associated with a borderline protective association for colon cancer. In women, no significant associations were observed for any food pattern when colon and rectal cancer were examined as separate end points.

Discussion

Several large clusters of diverse dietary composition were identified in the NIH–AARP cohort. A food pattern characterized by high intake of vegetables, fruits and other foods high in micronutrients and low in fat, was associated with reduced colorectal cancer incidence in men, even after

Table 4 Distribution (%) of some baseline characteristics* in women by three food pattern clusters in the the NIH–AARP Diet and Health Study 1995–2000

	Many foods	*Vegetables and fruits*	*Diet foods and lean meat*
		Frequency (%)	
Total	87 109 (44)	64 671 (32)	32 426 (16)
Age (years)			
Below 55	14 000 (49)	8643 (30)	4129 (14)
55–69	70 350 (43)	53 692 (33)	27 167 (17)
70 and above	2759 (41)	2336 (34)	1130 (17)
Education			
High school or less	30 877 (49)	14 520 (23)	10 925 (17)
College graduate	22 282 (38)	25 453 (43)	8604 (15)
Ethnicity			
Black	4822 (43)	2698 (24)	1376 (12)
White	78 243 (44)	58 349 (33)	29 808 (17)
Smoking			
Never	36 464 (41)	29 583 (33)	15 367 (17)
Former	30 847 (41)	28 676 (38)	12 165 (16)
Current	17 200 (61)	4307 (15)	3956 (14)
Physical activity			
Never, rarely	24 583 (53)	8525 (19)	6908 (16)
Once or more per week	46 789 (37)	48 444 (39)	20 404 (17)
BMI (kg/m^2)			
Less than 25	34 775 (42)	31 321 (38)	11 177 (14)
25–30	27 162 (43)	19 803 (32)	10 975 (18)
30 or more	21 044 (47)	10 666 (24)	8885 (20)
MHT			
Never user	42 767 (46)	27 613 (30)	14 434 (16)
Current user	36 445 (42)	31 097 (35)	14 964 (17)

Abbreviations: BMI, body mass index; MHT, menopausal hormone therapy.
**P*-values <0.0001 for all examined variables in χ^2-tests comparing differences across all categories and clusters.

adjusting for other known risk factors. In men, the 'Vegetable and fruits' and 'Fat-reduced foods' patterns were also associated with reduced rectal cancer incidence, although the small number of cases ($n=15$) for the 'Fat-reduced foods' pattern makes this finding somewhat tentative. In women, a similar 'Vegetable and fruit' pattern was associated with reduced colorectal cancer incidence, but that association was not independent of other risk factors.

The major advantages of the NIH–AARP Diet and Health Study are the large sample size and endpoint ascertainment from high-quality registries (Schatzkin *et al*., 2001). Further, prospective dietary data collection avoids biases associated with differential recall for cases and noncases. We kept the aggregation of the original food items to a minimum, in order to avoid the potential attenuation of food pattern–disease associations that may occur with broader food groups (McCann *et al*., 2001). The use of density variables based on consumption frequency and standardized to have the same variance, allowed food patterns characterized by low-energy foods to emerge. This may be an advantage when the diet–disease hypotheses include the health benefits of nonenergy contributing plant foods (Giovannucci, 2002; Bingham *et al*., 2003; Norat and Riboli, 2003; Larsson *et al*., 2006).

Findings of other food pattern studies (Randall *et al*., 1992; Slattery *et al*., 1998; Terry *et al*., 2001; Harnack *et al*., 2002; Fung *et al*., 2003; Dixon *et al*., 2004; Mizoue *et al*., 2005), are largely consistent with ours. Although two previous CA studies of dietary patterns and colorectal adenomas used distinct analytical approaches their findings were also consistent with ours (Rouillier *et al*., 2005; Austin *et al*., 2007). A French case–control study ($n=1372$) identified 5 clusters by first reducing the diet history data (159 food items) into 13 factors and then applying these factors to CA procedure (Rouillier *et al*., 2005). A US case–control study ($n=725$) used FFQ data converted to gram per 1000 kcal

Table 5 The hazard ratios of colorectal cancer associated with food patterns clusters in men and women of the NIH–AARP Diet and Health Study 1995–2000

Food pattern clusters	*Person years*	*Cases*	*Model 1*[a]			*Model 2*[b]			*Model 3*[c,d]		
			HR	*95%*	*CI*	*HR*	*95%*	*CI*	*HR*	*95%*	*CI*
Men											
Many foods	783 645	1372	1.00			1.00			1.00		
Vegetables and fruits	362 551	510	0.75	0.68	0.83	0.77	0.70	0.86	0.85	0.76	0.94
Fatty meats	100 515	178	0.95	0.82	1.11	0.95	0.81	1.11	0.94	0.80	1.10
Fat-reduced foods	50 049	76	0.82	0.65	1.04	0.84	0.67	1.06	0.88	0.70	1.11
Women											
Many foods	390 844	443	1.00			1.00			1.00		
Vegetables and fruits	291 021	283	0.82	0.70	0.95	0.83	0.72	0.97	0.90	0.77	1.06
Diet foods, lean meats	145 172	172	1.00	0.84	1.19	0.99	0.83	1.18	1.04	0.87	1.24

Abbreviations: CI, confidence interval; HR, hazard ratio.
HR estimates are shown with the largest cluster as the reference category.
[a]Age adjusted (continuous).
[b]Further adjusted for BMI (categories).
[c]Further adjusted for education, ethnicity, smoking, leisure time physical activity and total energy (continuous).
[d]In women also adjusted for MHT.

Table 6 The hazard ratios of colon and rectal cancer associated with food patterns clusters in men and women of the NIH–AARP Diet and Health study 1995–2000

Food pattern clusters	*Person years*	*Colon cancer*				*Rectal cancer*			
		Cases	*HR*[a,b]	*95%*	*CI*	*Cases*	*HR*[a,b]	*95%*	*CI*
Men									
Many foods	783 645	959	1.00			424	1.00		
Vegetables and fruits	362 551	381	0.89	0.79	1.01	136	0.74	0.60	0.91
Fatty meats	100 515	128	0.97	0.80	1.16	51	0.88	0.65	1.17
Fat-reduced foods	50 049	61	0.99	0.77	1.29	15	0.56	0.34	0.95
Women									
Many foods	390 844	329	1.00			116	1.00		
Vegetables and fruits	291 021	213	0.90	0.75	1.08	74	0.95	0.69	1.29
Diet foods, lean meats	145 172	118	0.95	0.77	1.18	54	1.29	0.93	1.78

Abbreviations: CI, confidence interval; HR, hazard ratio.
HR estimates are shown with the largest cluster as the reference category.
[a]Adjusted for age (continuous), BMI (categories), education, ethnicity, smoking, leisure time physical activity and total energy (continuous).
[b]In women also adjusted for MHT.

variables, but aggregated food variables into 18 food groups (Austin *et al*., 2007). The French study found that a food pattern high in animal fat, processed meat and total energy was associated with increased risk of colorectal cancer (Rouillier *et al*., 2005), whereas the US study found that a pattern high in fruit and low in meat was associated with reduced risk (Austin *et al*., 2007).

Although red and processed meats are thought to contain carcinogenic substances for large bowel cancer, and other studies have linked these foods to increased colorectal cancer risk (WCRF/AICR, 2007), comparable associations were not seen in our study. The lack of a significant association with the 'Fatty meats' cluster in men in our study was unexpected. However, the intake of alcohol, that previously has been associated with increased colorectal cancer risk (WCRF/AICR, 2007), was comparatively low in this cluster, and may have contributed to the findings. In women no cluster characterized by fatty meats emerged, instead hamburgers and meatloaf ranked comparatively high in the 'Many foods' cluster. The largest clusters in men and in women appear overall to show similar dietary characteristics. However, low-fat dairy foods ranked comparatively high in the 'Vegetable and fruit' cluster in women, and these foods ranked high in the 'Fat-reduced foods' cluster in men. Previous reports from this cohort also indicate differences in dietary heterogeneity in men and in women (Schatzkin *et al*., 2001). As we used energy-adjusted food variables, the differences cannot simply be a result of different energy intakes. These food selection differences by gender, consistent with previous research in this area (Randall *et al*., 1992; Wirfält *et al*., 2001), may influence the formation of patterns and could partly explain the observed differences in associations with colorectal cancer. Such food choice differences could depend on differences in health behavior awareness and social desirability (Hebert *et al*., 1997). A Danish review concluded that higher education in men was associated with food habits that tended to be more similar to those of women (O'Doherty Jensen and Holm, 1999). These differences could translate into actual dietary differences, or alternatively into differences in reporting of diet (measurement error; Macintyre and Anderson, 1997).

Dietary measurement error may affect the food pattern analysis in two ways. First, it may influence the formation of clusters leading to distortion of the main exposure. Although the effect of this potential distortion on the estimated HR has not been sufficiently studied, it is likely to attenuate the estimated cluster effect in a simple univariate analysis. Second, dietary measurement error may affect covariate adjustment, even for exactly measured confounders, by producing residual confounding in a multivariate model. The Observing Protein and Energy Nutrition (OPEN) study with reference biomarkers for protein and energy intake indicated that measurement error may be a greater threat to dietary assessment in women than in men (Kipnis *et al*., 2003), and could therefore contribute to the differences in associations observed in this study. The smaller sample size resulted in fewer cases and less analytical power to detect associations in women than in men, which, especially in the presence of measurement error, could have contributed to the observed differences in study outcomes by gender.

Moreover, not only diet but also lifestyle and socioeconomic factors may be imperfectly measured, so that residual confounding could affect results even when major potential confounders are included in the model. Also, as dietary patterns tend to covary with lifestyle and socioeconomic factors, both in men and women (Patterson *et al*., 1994; Greenwood *et al*., 2000; Engeset *et al*., 2005; Reedy *et al*., 2005) other unknown risk factors could, even in multivariate analysis, easily confound associations between clusters and disease risk.

To conclude, food patterns characterized by plant foods high in micronutrients and low in fat were associated with reduced colorectal cancer incidence in the NIH–AARP study. The associations were stronger in men than in women, and in men observed even in a multivariate model after adjusting for other known risk factors. Also, in men, these food patterns were more strongly associated with rectal cancer than with colon cancer. The observed gender differences may be due to actual differences in reported food choices, resulting in cluster differences or; alternatively, may be due to differences in statistical power or differences in residual confounding between men and women. Our findings are supported by previous food pattern studies.

Acknowledgements

This research was supported by the Intramural Research Program of the NIH, National Cancer Institute. Cancer incidence data from the Atlanta metropolitan area were collected by the Georgia Center for Cancer Statistics, Department of Epidemiology, Rollins School of Public Health, Emory University. Cancer incidence data from California were collected by the California Department of Health Services, Cancer Surveillance Section. Cancer incidence data from the Detroit metropolitan area were collected by the Michigan Cancer Surveillance Program, Community Health Administration, State of Michigan. The Florida cancer incidence data used in this report were collected by the Florida Cancer Data System under contract to the Department of Health (DOH). The views expressed herein are solely those of the authors and do not necessarily reflect those of the contractor or DOH. Cancer incidence data from Louisiana were collected by the Louisiana Tumor Registry, Louisiana State University Medical Center in New Orleans. Cancer incidence data from New Jersey were collected by the New Jersey State Cancer Registry, Cancer Epidemiology Services, New Jersey State Department of Health and Senior Services. Cancer incidence data from North Carolina were collected by the North Carolina Central Cancer Registry. Cancer incidence data from Pennsylvania were supplied by the Division of Health Statistics and Research, Pennsylvania Department of Health, Harrisburg, Pennsylvania. The Pennsylvania Department of Health specifically disclaims responsibility for any analyses, interpretations or conclusions. We are indebted to the participants in the NIH–AARP Diet and Health Study for their outstanding cooperation. Funding was also received from the Swedish Cancer Foundation (contract 05 0128 to EW) and the Swedish Council for Working Life and Social Research (contract 2005-1703 to EW).

References

Aldenderfer MS, Blashfield RK (1984). *Cluster Analysis. Quantitative Applications in the Social Sciences*. Sage Publications: California.

Austin GL, Adair LS, Galanko JA, Martin CF, Satia JA, Sandler RS (2007). A diet high in fruits and low in meats reduces the risk of colorectal adenomas. *J Nutr* **137**, 999–1004.

Bingham SA, Day NE, Luben R, Ferrari P, Slimani N, Norat T *et al.* (2003). Dietary fibre in food and protection against colorectal cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC): an observational study. *Lancet* **361**, 1496–1501.

Byers T, Gieseker K (1997). Issues in the design and interpretation of studies of fatty acids and cancer in humans. *Am J Clin Nutr* **66** (Suppl), 1541S–1547S.

Cox DR (1972). Regression models and life tables. *J R Stat Soc B* **34**, 187–220.

Dixon LB, Balder HF, Virtanen MJ, Rashidkhani B, Mannisto S, Krogh V *et al.* (2004). Dietary patterns associated with colon and rectal cancer: results from the Dietary Patterns and Cancer (DIETSCAN) Project. *Am J Clin Nutr* **80**, 1003–1011.

Engeset D, Alsaker E, Ciampi A, Lund E (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.

Flood A, Rastogi T, Wirfält E, Mitrou PN, Reedy J, Subar AF *et al.* (2008). Dietary patterns as identified by factor analysis and colorectal cancer among middle aged Americans. *Am J Clin Nutr* **88**, 176–184.

Fung T, Hu FB, Fuchs C, Giovannucci E, Hunter DJ, Stampfer MJ *et al.* (2003). Major dietary patterns and the risk of colorectal cancer in women. *Arch Intern Med* **163**, 309–314.

Giovannucci E (2002). Epidemiologic studies of folate and colorectal neoplasia: a review. *J Nutr* **132**, 2350S–2355S.

Giovannucci E, Stampfer MJ, Colditz G, Rimm EB, Willett WC (1992). Relationship of diet to risk of colorectal adenoma in men. *J Natl Cancer Inst* **84**, 91–98.

Greenwood DC, Cade JE, Draper A, Barrett JH, Calvert C, Greenhalgh A (2000). Seven unique food consumption patterns identified among women in the UK Women's Cohort Study. *Eur J Clin Nutr* **54**, 314–320.

Harnack L, Nicodemus K, Jacobs D, Folsom AR (2002). An evaluation of the Dietary Guidelines for Americans in relation to cancer occurrence. *Am J Clin Nutr* **76**, 889–896.

Hebert JR, Ma Y, Clemow L, Ockene IS, Saperia G, Stanek Jr E *et al.* (1997). Gender differences in social desirability and social approval bias in dietary self-report. *Am J Epidemiol* **146**, 1046–1055.

Hu FB (2002). Dietary pattern analysis: a new direction in nutritional epidemiology. *Curr Opin Lipidol* **13**, 3–9.

Hulshof KF, Wedel M, Lowik MR, Kok FJ, Kistemaker C, Hermus RJ *et al.* (1992). Clustering of dietary variables and other lifestyle factors (Dutch Nutritional Surveillance System). *J Epidemiol Commun Health* **46**, 417–424.

Jacques PF, Tucker KL (2001). Are dietary patterns useful for understanding the role of diet in chronic disease? *Am J Clin Nutr* **73**, 1–2.

Kant AK (2004). Dietary patterns and health outcomes. *J Am Diet Assoc* **104**, 615–635.

Kipnis V, Subar AF, Midthune D, Freedman LS, Ballard-Barbash R, Troiano RP *et al.* (2003). Structure of dietary measurement error: results of the OPEN biomarker study. *Am J Epidemiol* **158**, 14–21.

Larsson SC, Bergkvist L, Rutegard J, Giovannucci E, Wolk A (2006). Calcium and dairy food intakes are inversely associated with colorectal cancer risk in the Cohort of Swedish Men. *Am J Clin Nutr* **83**, 667–673.

Macintyre S, Anderson A (1997). Socio-demographic and psychosocial variables. In: Margetts BM, Nelson M (eds). *Design concepts in Nutitional Epidemiology*, 2nd edn. Oxford University Press: Oxford, pp 273–288.

McCann SE, Marshall JR, Brasure JR, Graham S, Freudenheim JL (2001). Analysis of patterns of food intake in nutritional epidemiology: food classification in principal components analysis and the subsequent impact on estimates for endometrial cancer. *Public Health Nutr* **4**, 989–997.

Michaud DS, Midthune D, Hermansen S, Leitzmann M, Harlan LC, Kipnis V *et al.* (2005). Comparison of cancer registry case ascertainment with SEER estimates and self-reporting in a subset

of the NIH–AARP Diet and Health Study. *J Registry Manag* **32**, 70–75.
Mizoue T, Yamaji T, Tabata S, Yamaguchi K, Shimizu E, Mineshita M *et al.* (2005). Dietary patterns and colorectal adenomas in Japanese men: the Self-Defense Forces Health Study. *Am J Epidemiol* **161**, 338–345.
Newby PK, Tucker KL (2004). Empirically derived eating patterns using factor or cluster analysis: a review. *Nutr Rev* **62**, 177–203.
Norat T, Bingham S, Ferrari P, Slimani N, Jenab M, Mazuir M *et al.* (2005). Meat, fish, and colorectal cancer risk: the European Prospective Investigation into cancer and nutrition. *J Natl Cancer Inst* **97**, 906–916.
Norat T, Riboli E (2003). Dairy products and colorectal cancer. A review of possible mechanisms and epidemiological evidence. *Eur J Clin Nutr* **57**, 1–17.
O'Doherty Jensen K, Holm L (1999). Preferences, quantities and concerns: socio-cultural perspectives on the gendered consumption of foods. *Eur J Clin Nutr* **53**, 351–359.
Patterson RE, Haines PS, Popkin BM (1994). Health lifestyle patterns of US adults. *Prev Med* **23**, 453–460.
Randall E, Marshall JR, Brasure J, Graham S (1992). Dietary patterns and colon cancer in Western New York. *Nutr Cancer* **18**, 265–276.
Reedy J, Haines PS, Campbell MK (2005). The influence of health behavior clusters on dietary change. *Prev Med* **41**, 268–275.
Reedy J, Mitrou PN, Krebs-Smith SM, Wirfalt E, Flood A, Kipnis V *et al.* (2008). Index-based dietary patterns and risk of colorectal cancer: the NIH–AARP Diet and Health Study. *Am J Epidemiol* **168**, 38–48.
Rouillier P, Senesse P, Cottet V, Valleau A, Faivre J, Boutron-Ruault MC (2005). Dietary patterns and the adenomacarcinoma sequence of colorectal cancer. *Eur J Nutr* **44**, 311–318.
Schatzkin A, Dorgan J, Swanson C, Potischman N (1995). Diet and cancer: future etiologic research. *Environ Health Perspect* **103**, 171–175.
Schatzkin A, Subar AF, Thompson FE, Harlan LC, Tangrea J, Hollenbeck AR *et al.* (2001). Design and serendipity in establishing a large cohort with wide dietary intake distributions: the National Institutes of Health-American Association of Retired Persons Diet and Health Study. *Am J Epidemiol* **154**, 1119–1125.
Slattery ML, Boucher KM, Caan BJ, Potter JD, Ma KN (1998). Eating patterns and colon cancer. *Am J Epidemiol* **148**, 4–16.
Subar AF, Midthune D, Kulldorff M, Brown CC, Thompson FE, Kipnis V *et al.* (2000). Evaluation of alternative approaches to assign nutrient values to food groups in food frequency questionnaires. *Am J Epidemiol* **152**, 279–286.
Subar AF, Thompson FE, Kipnis V, Midthune D, Hurwitz P, McNutt S *et al.* (2001a). Comparative validation of the Block, Willett, and National Cancer Institute food frequency questionnaires: the Eating at America's Table Study. *Am J Epidemiol* **154**, 1089–1099.
Subar AF, Thompson FE, Smith AF, Jobe JB, Ziegler RG, Potischman N *et al.* (1995). Improving food frequency questionnaires: a qualitative approach using cognitive interviewing. *J Am Diet Assoc* **95**, 781–788.
Subar AF, Ziegler RG, Thompson FE, Johnson CC, Weissfeld JL, Reding D *et al.* (2001b). Is shorter always better? Relative importance of questionnaire length and cognitive ease on response rates and data quality for two dietary questionnaires. *Am J Epidemiol* **153**, 404–409.
Terry P, Hu FB, Hansen H, Wolk A (2001). Prospective study of major dietary patterns and colorectal cancer risk in women. *Am J Epidemiol* **154**, 1143–1149.
Thompson FE, Kipnis V, Midthune D, Freedman LS, Carroll R, Subar AF *et al.* (2008). Performance of a food-frequency questionnaire in the US NIH–AARP (National Institutes of Health–American Association of Retired Persons) Diet and Health Study. *Public Health Nutr* **11**, 183–195.
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.
Willett WC, Howe GR, Kushi LH (1997). Adjustment for total energy intake in epidemiologic studies. *Am J Clin Nutr* **65** (Suppl), 1220S–1228S.
Wirfält E, Hedblad B, Gullberg B, Mattisson I, Andrén C, Rosander U *et al.* (2001). Food patterns and components of the metabolic syndrome in men and women. A cross-sectional study within the Malm' Diet and Cancer cohort. *Am J Epidemiol* **154**, 1150–1159.
World Cancer Research Fund/American Institute for Cancer Research (2007). *Food, Nutrition, Physical Activity and the prevention of Cancer: A Global Perspective*. AICR: Washington, DC.

Appendix 1

Description of six food pattern clusters in men ($n = 293\,576$) of the AARP cohort 1995–2000

	Many foods (N = 176 127)	*Vegetables and fruits (N = 81 318)*	*Fatty meats (N = 22 756)*	*Fat-reduced foods (N = 11 273)*	*Bacon and eggs (N = 1907)*	*Dessert (N = 195)*
	Energy-adjusted frequencies (frequency per 1000 kcal)					
$R^2 > 0.2$						
Lard	<0.001	<0.0001	<0.01	<0.001	0.18	<0.0001
Mayo, nf sand	<0.001	<0.01	<0.001	0.06	<0.001	<0.01
Pie, pumpkin	<0.001	<0.001	<0.01	<0.001	<0.01	0.27
Mayo, nf sal	<0.01	<0.01	<0.01	0.18	<0.01	<0.01
R^2 0.19–0.10						
Lettuce	0.15	0.32	0.16	0.26	0.12	0.13
Broccoli	0.05	0.13	0.05	0.09	0.03	0.07
Liver	<0.001	<0.001	0.02	<0.01	<0.01	<0.01
Carrot	0.06	0.19	0.08	0.13	0.04	0.09
Salad dress, nf	<0.01	0.06	<0.01	0.15	<0.01	0.01
Tomato, raw	0.12	0.25	0.14	0.21	0.12	0.11
Banana	0.18	0.35	0.15	0.29	0.13	0.22
Skim milk, w cer	0.05	0.18	0.03	0.18	0.02	0.05
Cheese, nf	<0.01	<0.01	<0.01	0.06	<0.001	<0.01
Apple	0.09	0.22	0.08	0.16	0.07	0.10
Peppers	0.04	0.12	0.05	0.08	0.04	0.04
R^2 0.09–0.05						
Fish, nf nfa	0.02	0.06	0.04	0.05	0.02	0.02
Salad dress, lf	0.05	0.13	0.04	0.04	0.02	0.04
Tomato sc, nm	0.02	0.06	0.02	0.04	<0.01	0.02
Cauliflower etc.	0.02	0.07	0.03	0.04	0.02	0.04
Orange	0.06	0.16	0.07	0.10	0.05	0.09
Fr chicken, d ws	<0.001	<0.001	0.01	<0.001	<0.01	<0.01
Bacon, reg	0.03	<0.01	0.06	0.01	0.11	0.02
Sausage, reg	0.02	<0.01	0.05	<0.01	0.05	0.02
Chicken, l ns	<0.01	0.03	<0.01	0.03	<0.01	<0.01
Beans, green	0.07	0.12	0.10	0.10	0.06	0.10
Pasta	0.05	0.10	0.05	0.07	0.03	0.04
Veg medley	0.03	0.08	0.05	0.06	0.02	0.05
Fr chicken, l ws	<0.01	<0.001	0.01	<0.001	<0.01	<0.01
Gravy	0.03	0.01	0.07	0.02	0.05	0.02
Ham	<0.01	<0.01	0.03	<0.01	<0.01	0.01
Grapes	0.03	0.09	0.04	0.05	0.02	0.05
Fruit, dried	0.03	0.11	0.03	0.07	0.02	0.06
Eggs, fa	0.09	0.03	0.12	0.03	0.25	0.05
Beef stew	<0.01	<0.01	0.03	<0.01	0.01	0.02
Cold cut, reg	0.07	0.02	0.10	0.04	0.08	0.04
Hotdog, reg	0.02	<0.01	0.04	<0.01	0.02	0.01
Roast beef, reg	<0.01	<0.001	0.02	<0.01	<0.01	<0.01
Roast beef, sand	0.01	<0.01	0.03	<0.01	0.01	0.01
Pie, custard	<0.01	<0.001	0.01	<0.01	<0.01	0.07
Grapefruit	0.04	0.11	0.04	0.07	0.03	0.06
Raw spinach	0.01	0.05	0.02	0.02	<0.01	0.02
Cracker, lf	0.02	0.06	0.02	0.08	<0.01	0.02
Frozen yogurt	0.04	0.09	0.04	0.12	0.02	0.06

Abbreviations: d, dark; dress, dressing; fa, fat added; Fr, fried; l, light; lf, low fat; nf, no fat; nfa, no fat added; nm, no meat; ns, no skin; reg, regular; sand, sandwich; sc, sauce; veg, vegetable; w cer, with cereal; ws, with skin.
The mean relative food frequencies of food variables, with R^2 values greater and equal than 0.05, are presented separately for each cluster.

Appendix 2

Description of nine food pattern clusters in women ($n = 198\,730$) of the NIH–AARP cohort 1995–2000

	Many foods (N = 87 109)	*Vegetables and fruits (N = 64 671)*	*Diet foods, lean meats (N = 32 426)*	*Desserts (N = 6480)*	*Chicken and dessert (N = 2932)*	*Shortening (N = 2106)*	*Chicken with skin (N = 1833)*	*Lard, bacon and eggs (N = 933)*	*Liver, vegetables and fruits (N = 240)*
	Energy-adjusted frequencies (frequency per 1000 kcal)								
$R^2 > 0.2$									
Lard	<0.01	<0.0001	<0.001	<0.01	<0.01	0	<0.01	0.24	<0.001
Fr chicken, d ws	<0.001	<0.0001	<0.001	<0.01	0.01	<0.01	0.09	<0.01	<0.01
Fr chicken, d ns	<0.001	<0.001	<0.001	<0.01	0.05	<0.001	0	<0.01	<0.01
Shortening	<0.01	<0.01	<0.01	0.01	0.02	0.38	0.03	0.05	0.02
Pie, custard	<0.001	<0.001	<0.001	0.06	<0.01	<0.01	<0.01	<0.01	<0.01
Fr chicken, lns	0.01	0.05	0.04	0.02	0.02	<0.01	0	<0.01	0.03
Liver	<0.01	<0.001	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.31
R^2 0.19–0.10									
Pie, fruit	<0.01	<0.01	0.01	0.08	0.01	0.01	0.01	<0.01	0.01
Carrots	0.09	0.27	0.15	0.11	0.11	0.08	0.07	0.06	0.13
Lettuce	0.20	0.43	0.27	0.19	0.20	0.16	0.15	0.16	0.22
Broccoli	0.08	0.20	0.11	0.08	0.10	0.07	0.07	0.06	0.12
Fr chicken, l ws	<0.001	<0.0001	<0.001	<0.01	0.02	<0.01	0.04	<0.01	<0.01
Pie, pumpkin	<0.001	<0.0001	<0.001	0.02	<0.01	<0.01	<0.01	<0.01	<0.01
Apple	0.10	0.28	0.14	0.11	0.12	0.11	0.09	0.08	0.14
Tomato, raw	0.16	0.31	0.22	0.16	0.18	0.15	0.14	0.15	0.21
Mayo, sand, diet	<0.01	0.02	0.05	0.02	0.02	<0.01	0.01	<0.01	0.02
Salad, dress nf	<0.01	0.11	0.04	0.02	0.02	<0.01	0.02	<0.01	0.03
Peppers	0.06	0.17	0.09	0.07	0.07	0.06	0.06	0.06	0.08
R^2 0.09–0.05									
Margarine, diet	0.08	0.19	0.37	0.13	0.16	0.08	0.07	0.06	0.12
Mayo, sal diet	0.02	0.03	0.11	0.04	0.04	0.02	0.02	0.02	0.03
Skim milk, cer	0.06	0.19	0.11	0.06	0.06	0.03	0.03	0.02	0.05
Orange	0.09	0.22	0.11	0.10	0.10	0.08	0.08	0.06	0.13
Meatloaf	0.04	0.02	0.07	0.05	0.05	0.04	0.05	0.05	0.04
Cold cut, lf	<0.01	0.01	0.05	0.01	0.02	<0.01	<0.01	<0.01	0.02
Chicken, l ns	0.01	0.05	0.04	0.02	0.02	<0.01	0	<0.01	0.03
Fish, nf nfa	0.03	0.07	0.04	0.04	0.03	0.02	0.02	0.02	0.06
Tomato, sc nm	0.02	0.06	0.03	0.03	0.02	<0.01	0.01	<0.01	0.02
Yogurt	0.06	0.17	0.07	0.05	0.05	0.03	0.03	0.03	0.07
Banana	0.22	0.39	0.26	0.22	0.23	0.21	0.17	0.16	0.23
Cauliflower etc	0.03	0.09	0.06	0.04	0.05	0.03	0.03	0.03	0.07
Salad, dress lf	0.05	0.13	0.12	0.06	0.07	0.04	0.04	0.02	0.05
Egg, fa	0.08	0.02	0.04	0.06	0.07	0.15	0.11	0.18	0.07
Mayo, reg sand	0.09	0.02	0.06	0.08	0.09	0.10	0.11	0.12	0.09
Hamburger, lean	0.02	<0.01	0.03	0.02	0.03	0.02	0.02	0.02	0.02
Hot dog, lf	<0.01	<0.01	0.02	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
Roast beef, lean	<0.01	<0.01	0.02	0.01	<0.01	<0.01	<0.01	<0.01	<0.01
Green beans	0.08	0.15	0.14	0.11	0.12	0.10	0.09	0.08	0.14
Cantaloupe	0.01	0.04	0.02	0.02	0.02	0.01	0.01	0.01	0.02
Strawberry	0.01	0.04	0.02	0.02	0.02	<0.01	<0.01	<0.01	0.02
Spinach, raw	0.02	0.07	0.03	0.03	0.02	0.02	0.02	0.01	0.04
Grapefruit	0.05	0.14	0.07	0.06	0.06	0.05	0.04	0.04	0.10
Pasta	0.07	0.12	0.08	0.07	0.06	0.04	0.05	0.04	0.06
Bacon, reg	0.03	<0.01	0.02	0.03	0.04	0.07	0.07	0.10	0.04
Gravy	0.03	<0.01	0.04	0.05	0.06	0.05	0.06	0.05	0.06

Abbreviations: cer, with cereal; dress, dressing; fa, fat added; l, light; lf, low fat; nf, no fat; nfa, no fat added; nm, no meat; ns, no skin; reg, regular; sand, sandwich; sc, sauce; veg, vegetable.
The mean relative food frequencies of food variables, with R^2 values greater and equal than 0.05, are presented separately for each cluster.