Generalizability of dietary patterns associated with incidence of type 2 diabetes mellitus[1–4]

Fumiaki Imamura, Alice H Lichtenstein, Gerard E Dallal, James B Meigs, and Paul F Jacques

ABSTRACT

Background: Reduced rank regression (RRR) has been used to derive dietary pattern scores that predict linear combinations of disease biomarkers. The generalizability of these patterns to independent populations remains unknown.

Objective: The goal was to examine the generalizability of dietary patterns from the following prior studies using RRR to predict type 2 diabetes mellitus (T2DM): the Nurses' Health Study (NHS), European Prospective Investigation into Cancer and Nutrition Germany (EPIC), and Whitehall II Study (WS).

Design: The relative weights of food groups of each dietary pattern were used to generate each dietary pattern score in the Framingham Offspring Study (n = 2879). Each of the external scores (confirmatory scores) was examined to determine whether it could predict incident T2DM during 7 y of follow-up as well as scores developed internally in the Framingham Offspring Study using a Cox-proportional hazard model adjusted for T2DM risk factors.

Results: Intakes of meat products, refined grains, and soft drinks (caloric and noncaloric) were found to be common predictive components of all confirmatory scores, but fried foods, eggs, and alcoholic beverages were predictive in some, but not in all, confirmatory scores. On the basis of a continuous increase in the score by 1 SD, the NHS-based confirmatory score predicted T2DM risk (hazard ratio: 1.44; 95% CI: 1.25, 1.66). However, T2DM risk was only weakly predicted by the EPIC-based score (hazard ratio: 1.14; 95% CI: 0.99, 1.32) and the WS-based score (hazard ratio: 1.16; 95% CI: 1.00, 1.35).

Conclusions: The study suggested that dietary patterns that predict T2DM risk in different populations may not be generalizable to different populations. Additional dietary pattern studies should be conducted with regard to generalizability. *Am J Clin Nutr* 2009;90:1075–83.

INTRODUCTION

Reduced rank regression (RRR) has been increasingly applied in nutritional epidemiology as one approach for dietary pattern analyses (1, 2). This technique is a dimension reduction technique modeling linear combinations of predictor variables, in our case food groups, that explain the maximum variation in a set of response variables, typically intermediate metabolic biomarkers of disease outcomes. The resulting linear combination of predictor variables is subsequently related directly to the disease outcome of interest. Because dietary patterns are derived with reference to disease risk factors, the approach has an etiologic advantage over other dietary pattern methods, such as principal component analysis and dietary indexes, that is either totally data-derived without reference to disease risk or based on prior knowledge not specific to disease outcomes (2).

Past studies have used RRR to derive dietary patterns to predict variation of biomarkers such as inflammatory biomarkers, homocysteine, lipoproteins, glycated proteins, and markers related to glucose homeostasis (2–7). Although there is a growing number of studies using pattern-deriving exploratory RRR, evidence of the generalizability of these derived patterns is limited. This lack of evidence can be addressed by pattern-testing confirmatory RRR in which established dietary patterns are applied to other independent populations. Therefore, with the use of the Framingham Offspring Study (FOS) cohort as the independent cohort, we conducted confirmatory RRR to test the generalizability of 3 published RRR-derived dietary patterns associated with incidence of type 2 diabetes mellitus (T2DM) (5–7). We also conducted exploratory RRR to derive dietary patterns from the FOS dietary data for comparison with the established patterns.

SUBJECTS AND METHODS

Study population

The FOS is community-based, prospective, observational study initiated in 1971 among the offspring of the original participants of the Framingham Heart Study (8, 9). During the fifth examination cycle (1991–1995) of the FOS (the baseline visit for these analyses), 3799 participants underwent a stan-

[1] From the Jean Mayer US Department of Agriculture, Human Nutrition Research Center on Aging at Tufts University, Boston, MA (FI, AHL, GED, and PFJ), and the General Medicine Division and Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA (JBM).

[2] Any opinions, findings, conclusions or recommendations expressed in this publication are those of the authors, and do not necessarily reflect the view of the US Department of Agriculture.

[3] Supported in part by the US Department of Agriculture, Agricultural Research Service, under agreement no. 58-1950-7-707 and the Framingham Heart Study of the National Heart Lung and Blood Institute of the National Institutes of Health (contract no. N01-HC-25195). JBM was supported by NIDDK K24 DK080140.

[4] Address correspondence to PF Jacques, Jean Meyer USDA Human Nutrition Research Center on Aging, 711 Washington Street, Boston, MA, 02111. E-mail: paul.jacques@tufts.edu.

Received April 30, 2009. Accepted for publication July 23, 2009.

First published online August 26, 2009; doi: 10.3945/ajcn.2009.28009.

dardized medical examination. Participants were followed up for 7 y on average from baseline to the sixth (1995–1998) and seventh (1998–2001) examinations.

To derive dietary pattern scores from external cohorts, publications before August 2008 were searched for the terms *reduced rank regression* and *type 2 diabetes*. Three prospective cohort studies of RRR-derived dietary patterns related to incident T2DM were identified: the Nurses' Health Study (NHS) in the United States (5), European Prospective Investigation into Cancer and Nutrition Potsdam Study (EPIC) in Germany (6), and Whitehall II Study (WS) in the United Kingdom (7).

Dietary and outcome variables

At the fifth examination, dietary assessment was implemented to measure habitual dietary consumption during the previous year with a 126-item semiquantitative food-frequency questionnaire (FFQ) (10, 11). Participants were asked to choose from 1 of 9 categories to indicate how often, on average, they had consumed given amounts of various specified foods during the past year. The external reproducibility and validity of food consumption, nutrient intakes and dietary patterns are described elsewhere (11–14).

In this longitudinal setting, follow-up time of each individual was determined by exact dates of the baseline and either the sixth or seventh examination, depending on whether T2DM onset was before the sixth or seventh examination and availability of the follow-up data. T2DM in the present analysis was defined as oral hypoglycemic drug or insulin use or a fasting glucose concentration of $\geq$126 mg/dL ($\geq$7.0 mmol/L) at the time of the examinations. The cases developing T2DM between the sixth and seventh examinations showed no missing information to ascertain T2DM at the sixth examination. Sensitivity analyses simulating time of T2DM onset and follow-up duration indicated little bias due to potential misclassification (data not shown).

Exclusion criteria

We excluded 193 individuals without follow-up information and 298 individuals who had been diagnosed with T2DM at baseline. We excluded 381 individuals because their estimated daily caloric intake was $<$600 kcal/d (2.51 MJ/d), $\geq$4000 kcal/d for women, $\geq$4200 kcal/d for men, or had $\geq$12 items left blank on their FFQ (15). We excluded additional individuals (n = 48) without data on the following covariates: body mass index (BMI; in kg/m^2), weight change during follow-up, HDL cholesterol, systolic or diastolic blood pressure, and a fasting glucose concentration. Using the available data from 2879 participants, we conducted complete-case analyses for the sake of simplicity. Imputation for the missing variables showed no appreciable difference in main results and in precision estimates of interests.

Statistical analyses

Population characteristics

Population characteristics of the FOS were described by means and SDs for continuous variables and percentages for categorical variables. The characteristics of the 3 external cohorts providing RRR-derived dietary patterns were described from their previous reports (5–7, 14, 16–19). With the use of the descriptive statistics, proportions of categorical variables (sex, smoking, and race-ethnicity status) and means of continuous variables (age, energy intake, and alcohol consumption) were compared between the FOS and each of the other cohorts by chi-square tests and independent t tests, respectively. The study characteristics of dietary assessment methods were also described.

Confirmatory dietary pattern scores

Three dietary pattern scores were derived by using the FOS dietary data with reference to previous RRR studies from the NHS (5, 14), EPIC (6, 17), and WS (7, 19). Two types of information were retrieved from each cohort: food grouping and contribution of each food group to each RRR-derived dietary pattern, and then they were applied to the FOS data. First, because each study used unique food grouping from a different FFQ (5–7, 14, 17, 19), we applied each food grouping to the FOS dietary data as closely as possible and created 3 different sets of food groups. Details of the food groupings are presented in our Supplemental Section I under "Supplemental data" in the online issue. Second, multivariate regression analyses were performed to obtain residuals of food group variables adjusted for caloric intake (quintiles), age ($<$50, 50–64, or $\geq$65 y), sex, parental history of diabetes (yes or no), and hypertension treatment (yes or no) to reduce potential bias from the dependency of food groups on caloric intake and the other nondietary factors in the FOS population. Other factors, including physical activity and multivitamin use, did not alter the results. The same approach was taken in the past studies, and no difference in results was stated (5–7). Third, the residuals were standardized to variables with the mean equal to zero and the SD equal to 1; this step is embedded in the RRR procedure. Fourth, the sum of the products of the food group variable and dietary pattern coefficients were calculated and considered as confirmatory dietary pattern scores. As the dietary pattern coefficients, Pearson correlation coefficients were used for the NHS-based scores, because the original NHS study presented the Pearson correlation coefficients representing contributions of food groups to the NHS dietary pattern (5). For EPIC- and WS-based dietary patterns, factor loading values were used as the dietary pattern coefficients, because the EPIC and WS reported the factor loading values (6, 7). Either use of Pearson correlation coefficients or factor loading values did not alter our conclusions, because they were equivalent with different scales (*see* Supplemental Section II under "Supplemental data" in the online issue).

Exploratory dietary pattern scores

Because the predictability of incident T2DM could be influenced solely by manners of food groupings, we additionally performed 3 RRR analyses within the FOS cohort (*exploratory analyses* in which each food grouping of the NHS, EPIC, and WS cohorts was applied. The exploratory and confirmatory analyses allowed together the comparison between internally and externally derived scores in the predictability of incident T2DM given the same manner of food grouping. In the FOS cohort, RRR was used to fit multiple risk factors for T2DM to food group variables. From each fitted model of the 3 RRR analyses, the exploratory dietary pattern score was derived. The

scoring coefficients derived from each RRR analysis are provided in the online Supplement Section I under "Supplemental data" in the online issue.

The response variables of the exploratory RRR were risk factors for T2DM in the FOS cohort (20): BMI, fasting glucose concentration, triglyceride concentration, HDL-cholesterol concentration, and hypertension (based on elevated systolic and/or diastolic blood pressure or hypertension treatment). Because RRR analysis generally uses continuous variables, hypertension was replaced with residuals of systolic and diastolic blood pressure variables after being regressed on hypertension treatment. The contribution of each food group to each exploratory score was expressed by Pearson correlation coefficients (NHS-based exploratory score) or factor loading values (EPIC- and WS-based scores), so that the statistics could be comparable with the reports from the external cohorts.

Prediction of T2DM risk

Each of the confirmatory and exploratory scores was tested for prediction of T2DM incidence. After categorizing the FOS population into 4 groups by quartiles of each score, Cox-proportional hazard regression model was used to estimate hazard ratios (HRs) with the first quartile category of each score as a reference category. Ties of failure and censoring times were handled by Efron adjustment to yield valid parameter estimates (21). The regression analyses were performed with and without the following covariates: sex, age ($<$50, 50–64, or $\geq$65 y), parental history of T2DM (yes or no), and blood pressure treatment (yes or no) and weight change (quintiles). Caloric intake (quintiles) was also included in the model to allow for an isocaloric interpretation (22). Other variables, such as smoking status, menopausal status for women, multivitamin use, and physical activity were not included in the model because they had little influence on prediction of T2DM risk as confounders. The regression analysis was also performed by including each of the scores as a continuous variable, where one unit change of each score corresponded to 1 SD of the study population.

Furthermore, we calculated the ratio of HR based on confirmatory analysis to HR based on exploratory analysis (HRc/HRe) (23). This measure addressed the question of whether the prediction of T2DM risk was different between the confirmatory and exploratory dietary pattern scores. We additionally performed permutation analysis (20,000 samples) to derive P values adjusted for multiple testing for HRc/HRe to examine the generalizability of the 3 dietary patterns of the NHS, EPIC, and WS cohorts (24, 25). By the same permutation analysis, we also computed P values to draw simultaneous inference based on $\geq$2 null hypotheses of the ratios of the HRs.

Stratified analyses were performed to assess the sensitivity of the findings to different strata of age ($<$54 y of age or not), sex, and dichotomy of the 6 response variables. Also, exploratory analyses were repeated 6 times, excluding 1 of 6 response variables in RRR in each repeat. This demonstrated the robustness of the predictability of T2DM risk with respect to the selection of biomarkers in our exploratory RRR analyses (6). The details of these analyses are presented in the supplementary section III under "Supplemental data" in the online issue. All statistical analyses were performed by using SAS version 9.13. P values were based on 2-sided tests.

RESULTS

Characteristics of the 4 cohorts are presented in **Table 1**: the FOS as the main cohort for this study and the other 3 reporting RRR-derived dietary patterns previously. The NHS and EPIC had nested-case-control designs. The proportions of men and whites and the mean age of the FOS were statistically significantly different from those of the NHS and WS populations ($P <$ 0.0001), but not from the EPIC populations ($P > 0.05$). Mean differences in BMI, energy intake, and alcohol consumption and the proportion of smokers were statistically significantly different between the FOS and each of the 3 cohorts ($P < 0.0001$). Compared with the FOS population, the NHS participants showed less smoking prevalence, energy intake, and alcohol drinking, but higher BMI. The EPIC population showed higher smoking prevalence, higher BMI, higher energy intake, and higher alcohol intake than the FOS population. The WS population was younger, smoked less, and was less obese, but had higher energy and alcohol intakes than the FOS cohort. Hypertension status and family history were not comparable because of different definitions for ascertainment.

Differences in study characteristics are also presented in Table 1: dietary assessment, food grouping, and response variables used in RRR analyses. The 3 external cohorts and the FOS had different response variables. The NHS used biomarkers of inflammation, the EPIC used biomarkers of lipid and glucose homeostasis and inflammation, and the WS used biomarkers of glucose homeostasis. In the FOS, biomarkers of lipid and glucose homeostasis and measures of blood pressure and adiposity were selected. There was no common response variable in the FOS and the NHS, whereas the FOS and the EPIC or WS used at least one common response variable.

Correlations of selected food groups with the dietary pattern scores assessed by Pearson correlation coefficients (NHS) or factor loadings (EPIC and WS) are presented in the **Table 2**. They are arranged by major categories (vegetables, fruit, meat, dairy products, grains, beverages, and others) and represent either those having high contributions or those previously reported to be associated with T2DM risk, such as nuts and dairy products (26). All 3 exploratory scores had similar positive contributions (ie, indication of greater risk) by food groups: meat, processed meat, eggs, margarine, fried products, refined grain, and caloric and noncaloric soft drinks. However, the negative contributors varied substantially from one score to the next, with few exceptions of tea and whole-grain consumption. There was also good agreement between the exploratory and confirmatory RRR scores for NHS food groupings, although the original NHS report presented only selected food groups. There was less agreement between the exploratory and confirmatory dietary scores for the EPIC and WS food groupings. The contributions also differed to some extent for margarine, pizza, fruit juice, oil and vinegar salad dressing, garlic, dark-yellow vegetables, and alcoholic beverages.

Pearson correlation coefficients among the dietary different dietary pattern scores and those between the scores and response variables showed large variability (**Table 3**). Three scores based on exploratory RRR correlated well ($r > 0.7$), with only differences in food grouping. The correlation between the confirmatory and exploratory scores based on NHS food grouping was high ($r = 0.63$) relative to the other pairs based on EPIC or WS

TABLE 1
Population characteristics of the Framingham Offspring Study and the 3 cohorts conducting dietary pattern analyses by using reduced rank regression (RRR) for type 2 diabetes mellitus (T2DM) risk prediction[1]

	Framingham Offspring Study	NHS[2]		EPIC Study[2]		Whitehall II Study[2]
		Cases	Controls	Cases	Controls	
Subject characteristics[3]						
No. of subjects	2879	656	694	192	382	7339
Men (%)	54.5	0	0	58.9	58.9	69.6
White (%)	≈100	94.6	94.6	≈100	≈100	90.4
Age (y)	54.2 ± 9.7[4]	56.3 ± 6.9	56.2 ± 6.9	55.5 ± 6.8	55.5 ± 6.8	49.5 ± 8.6
Current smoking (%)[3]	18.6	13.4	13.2	19.8	20.7	13.5
Total energy (kcal/d)	1871 ± 621	1817 ± 561	1769 ± 512	2211 ± 727	2177 ± 652	2160 ± 661
BMI (kg/m^2)[3]	27.1 ± 4.7	30.2 ± 5.6	26.1 ± 5.1	30.8 ± 4.8	26.7 ± 3.6	25.3 ± 3.4
Hypertension (%)[5]	19.8	—	—	78.8	51.7	16.7
Family history of T2DM (%)[6]	17.2	47.6	20.8	—	—	—
Alcohol (g/d)[3]	10.7 ± 16.2	3.7 ± 7.1	6.4 ± 9.0	18.5 ± 28.2	16.1 ± 16.4	12.1 ± 17.1
Study characteristcs[7]						
Years of FFQ collection	1991–1995	1986 and 1990		1994–1998		1985–1988
No. of items on the FFQ	126	131		148		127
No. of food groups[7]	—	39		48		71
RRR response variables	BMI, glucose, HDL, triglycerides, systolic and diastolic blood pressures	Inflammatory cytokines[8]		HDL cholesterol, glycated hemoglobin, C-reactive protein, adiponectin		Homeostasis model assessment of insulin resistance

[1] Values for the external cohorts were retrieved from previous publications (5–7, 14, 16–19). NHS, Nurses' Health Study; EPIC, European Prospective Investigation into Cancer Potsdam Study; FFQ, food-frequency questionnaire.

[2] The NHS and EPIC had a prospective nested case-control design (5, 6), whereas the Framingham Offspring Study (FOS) and the Whitehall II Study (WS) had a prospective cohort design (7, 8). The NHS applied matching by year of birth, date of blood draw, race-ethnicity, and time of the blood draw. EPIC applied matching by age.

[3] The statistics of the FOS were different from those of the 3 cohorts ($P < 0.0001$) according to a chi-square test for categorical variables and an independent t test for continuous variables. Proportions of men and whites and the mean age of the FOS were different from those of the NHS and WS cohorts ($P < 0.0001$) but not from those of the EPIC cohort ($P > 0.05$).

[4] Mean ± SD (all such values).

[5] On the basis of self-report of prior diagnosis or hypertension treatment in EPIC and systolic blood pressure >140 mm Hg or diastolic blood pressure >90 mm Hg in the WS and FOS. With the definition used in the EPIC, 30.4% of the FOS population had hypertension.

[6] Parental history for the FOS cohort and family history of first-order relatives of the NHS cohort.

[7] All studies used FFQs. The NHS had repeated measurements at the 2 time points and took the average of each food group, whereas the other studies had single measurements. The FOS applied all 3 food groupings reported in the 3 external studies.

[8] Tumor necrosis factor α receptor 2, interleukin-6, C-reactive protein, E-selectin, and intracellular and vascular cell adhesion molecules.

food groups (0.12 and 0.27, respectively). The 3 confirmatory scores were poorly correlated ($r < 0.3$). None of exploratory or confirmatory pattern scores were highly correlated with the response variables used for the exploratory RRR analyses. The strongest correlations were observed between the exploratory scores and BMI ($r = 0.24$–0.28) and HDL cholesterol ($r = -0.12$ to -0.23). The scores derived from confirmatory analyses were less correlated with the response variables.

Multivariate HRs (95% CIs) for T2DM were 1.58 (1.37, 1.83) and 1.44 (1.25, 1.66) according to continuous increase by 1 SD of exploratory and confirmatory scores using NHS food groups, respectively; 1.60 (1.39, 1.83) and 1.14 (0.99, 1.32) for the EPIC food groups; and 1.60 (1.39, 1.83) and 1.16 (1.00, 1.35) for the WS food groups (**Table 4**). The ratios of the confirmatory and exploratory HRs were statistically significantly different from 1.0 for analyses based on EPIC (HRc/HRe: 0.76; 95% CI: 0.64, 0.90) and WS (HRc/HRe: 0.75; 95% CI: 0.62, 0.90) ($P = 0.027$ and 0.021, respectively, adjusted for multiple testing). This concomitant observation of rejecting the 2 hypotheses was statistically significant ($P = 0.0016$) according to permutation analysis involving 20,000-times iteration. The exploratory and confirmatory scores based on the NHS were not significantly different in predication of incident T2DM (HRc/HRe: 0.91; 95% CI: 0.82, 1.01; $P = 0.16$ adjusted for multiple testing) (**Table 5**).

DISCUSSION

We examined whether T2DM risk in the FOS could be predicted from dietary patterns originating from 3 external cohort studies: NHS (5), EPIC-Potsdam Study (6), and WS (7). The dietary pattern derived from the NHS score was as predictive of T2DM risk as was the dietary pattern derived from the FOS cohort itself. However, the dietary patterns derived from the EPIC and WS were significantly less predictive. Regarding the consistency and inconsistency of the predictive capability of T2DM risk within the FOS population, the finding suggest that the T2DM-related dietary pattern of the NHS population was generalizable to the FOS population and that those of the EPIC and WS populations were not generalizable to the FOS population. The study indicates that overall diet related to T2DM risk can be different across population.

More consistency across dietary patterns was observed for those food groups that were positive contributors to the dietary scores (ie, food groups that were associated with greater risk of

TABLE 2
Pearson correlation coefficients or factor loadings for contributions of food groups to dietary pattern scores reported from 3 cohort studies and derived from reduced rank regression (RRR) analyses in the Framingham Offspring Study cohort 1991–2001 (n = 2879)[1]

	NHS		EPIC Study		Whitehall II Study	
Food group	Explore[2]	Confirm[3]	Explore[2]	Confirm[3]	Explore[2]	Confirm[3]
Vegetables						
Cooked vegetables[4]			−0.02	0.12		
Raw vegetables			0.01	−0.03		
Dark-yellow vegetables[5]	−0.20	−0.21			−0.17	0.03
Green leafy vegetables[4]	−0.22					
Green salad, tomatoes					−0.10	−0.04
Spinach, kale					−0.07	0.09
Cruciferous vegetables[4]	−0.10	−0.21				
Cabbage, Brussels sprouts, cauliflower					−0.13	0.05
Broccoli					−0.11	−0.13
Garlic	−0.12		−0.11	0.10	−0.07	0.25
Fruit						
Fruit[4]	−0.08		−0.09	0.09		
Dried fruit (raisins)					−0.20	−0.15
Meat, eggs, nuts						
Red meats	0.28		0.26	0.36	0.19	0.15
Eggs	0.16		0.15	0.18	0.17	0.02
Nuts	−0.08		−0.08	0.04	0.08	−0.03
Processed meats[4]	0.21	0.39	0.20	0.26	0.17	0.12
Beef burgers and sausages					0.25	0.22
Dairy products[5]						
Margarine	0.27		0.24	0.01	0.19	0.04
Polyunsaturated margarine						−0.10
High-fat dairy	−0.11		−0.14	0.08	−0.10	0.09
Low-fat dairy	0.08		0.05	0.00	0.02	0.07
High-fat cheese			0.06	0.08	0.06	−0.07
Low-fat cheese			0.01	−0.01	0.02	−0.04
Yogurt					−0.16	−0.13
Grain products						
Refined grains	0.26	0.46	0.21	0.24	0.22	0.20
Whole grains[4]	−0.22		−0.13	−0.13	−0.11	−0.05
Pasta, rice (whole or refined)			−0.04	0.03		
Pasta (whole or refined)					−0.05	−0.16
Rice (whole or refined)					−0.03	0.18
Beverages						
Low-calorie soft drinks	0.45		0.42	0.07	0.39	0.27
Caloric soft drinks	0.20	0.47	0.18	0.27	0.14	0.23
Fruit juice	−0.10		−0.09	0.09	−0.15	−0.03
Wine	−0.44	−0.43	−0.40	−0.08		
Beer	−0.32		−0.29	0.35		
Other alcoholic beverages	−0.24		−0.22	0.18		
Coffee	−0.16	−0.29	−0.10	−0.10	−0.08	−0.10
Tea	−0.12		−0.11	−0.13	−0.15	−0.11
Other groups						
French fries	0.22		0.20	0.15	0.19	0.13
Fried foods	0.29				0.24	0.19
Snacks	−0.05		−0.04	−0.01	0.09	0.22
Pizza	0.25		0.23	0.04	0.21	−0.07
Oil and vinegar dressing	−0.12		−0.11	0.06	−0.05	−0.21
Jams, honey, marmalade	0.01		−0.13	−0.16	−0.07	−0.21

[1] Pearson correlation coefficients are presented for the scores based on the Nurses' Health Study (NHS). Factor loadings are presented for the scores based on European Prospective Investigation into Cancer and Nutrition Potsdam (EPIC) Study and Whitehall II Study (WS) cohorts.

[2] Defined as the exploratory dietary pattern score derived from RRR analyses within the Framingham Offspring Study on the basis of food groupings of 1 of the 3 external cohorts.

[3] Defined as the dietary pattern reported in 1 of the 3 external cohorts, and the values were obtained from each of the original dietary pattern reports.

[4] Grain, fruit, and vegetable groups had substantial differences in the manner of grouping between the 3 cohorts; overall, WS groups were aggregated to the least extent (*see* the number of food groups in Table 1 and in Supplementary Section I under "Supplemental data" in the online issue).

[5] In the WS, factor loading values for yogurt and carrots are presented for the foods groups high-fat dairy products and dark-yellow vegetables, respectively. Of the dairy products, only the NHS collapsed cheese and milk, and only the WS separated yogurt from the other dairy food groups.

TABLE 3
Pearson correlation coefficients between the exploratory and confirmatory dietary pattern scores and between the dietary pattern scores and type 2 diabetes risk factors in the Framingham Offspring Study Cohort 1991–2001 (n = 2879)[1]

	Dietary pattern scores					
	Exploratory[2]			Confirmatory[3]		
Dietary pattern scores	NHS	EPIC	WS	NHS	EPIC	WS
Exploratory						
NHS	1.00	0.94	0.75	0.63	0.12	0.23
EPIC		1.00	0.75	0.60	0.12	0.25
WS			1.00	0.50	0.21	0.27
Confirmatory						
NHS				1.00	0.08	0.28
EPIC					1.00	0.17
WS						1.00
Response variables						
BMI	0.20	0.24	0.24	0.28	0.14	0.07
Fasting glucose	0.11	0.06	0.05	0.11	0.03	0.04
HDL	−0.09	−0.23	−0.22	−0.12	−0.14	0.05
Triglyceride	0.11	0.10	0.08	0.12	0.08	0.01
Systolic blood pressure	0.10	0.05	0.05	0.13	0.07	0.04
Diastolic blood pressure	0.09	0.09	0.09	0.13	0.06	0.05

[1] EPIC, European Prospective Investigation into Cancer and Nutrition Potsdam Study; NHS, Nurses' Health Study; WS, Whitehall II Study. $P < 0.05$ for $r > 0.04$ or $r < -0.04$, $P < 0.01$ for $r > 0.06$ or $r < -0.06$, and $P < 0.001$ for $r > 0.07$ or $r < 0.07$. The 95% CIs surrounding all correlation statistics ranged from ±0.033 to ±0.037.

[2] Defined as the exploratory dietary pattern score derived from reduced rank regression analyses within the Framingham Offspring Study on the basis of food groupings of 1 of the 3 external cohorts.

[3] Defined as the confirmatory dietary pattern score derived from applying the dietary patterns of the 3 external cohorts to the Framingham Offspring Study population on the basis of food groupings of 1 of the 3 external cohorts.

T2DM in each cohort), rather than strong negative contributors to the scores. Consistent across the different cohorts, soft drinks, particularly caloric soft drinks, meat and processed meat, eggs, refined grains, and French fried potatoes were positive contributors to the dietary patterns associated with T2DM risk. These food groups were previously found as those associated with incident T2DM (27–37), with the exception of French fried potatoes. French fried potatoes and other fried foods are potentially a source of *trans*-fatty acids, at least have been in the past, and have in some observational studies been positively associated with T2DM risk or insulin resistance (37–39). Further work in this area is necessary to assess individual food groups within the context of different dietary patterns.

Contributions of some food groups to dietary patterns were not consistent across the populations, including alcoholic beverages, margarine, and different types of vegetables and fruit. One notable example was alcoholic beverages. The discrepancies are likely due to differences in the sex proportion and the range of alcohol consumption among the study populations and possibly cultural patterns of intake (Table 1). Alcohol consumption in our study (FOS) was relatively similar to that in NHS which recruited only women. The number of participants consuming >30 g/d in our study was 351 (12.2%), a percentage similar to that of the full NHS population (12.0%, 5624 of 46,892) (40). On the contrary, alcohol consumption in EPIC and WS was higher than that in FOS. Past observational studies suggest that the association between alcohol and T2DM risk is U-shaped (41, 42). In the FOS, a linear inverse association between alcohol and T2DM risk was observed, probably because of the limited proportion of our population with higher intakes (43). Therefore, populations with a narrow range of alcohol consumption and few heavy consumers are likely to yield a favorable result for alcohol consumption in relation to T2DM risk, whereas other populations with a wide range of consumption are more likely to yield an unfavorable result. These characteristics may mask dietary benefit or harm due to U-, J-, or S-shaped associations and yield inconsistent results among dietary pattern studies in different populations. Therefore, automation of standardizing each food variable and assumption of monotonic associations among food variables should be considered as potential limitations of dietary pattern analyses.

In addition, discordance of dietary patterns could be observed in case one population had a narrow distribution of consumption of a food group relative to other food groups, because such a food group cannot be captured by data-driven dietary patterns. This is independent of dietary pattern techniques and shapes of diet-disease associations. This may explain, for example, the difference in margarine consumption as dietary pattern components of the FOS and the 2 European cohorts (Table 2). Since the late 1980s, European populations may have had a limited consumption of margarine rich in *trans*-fatty acids but not the US population (44). Thus, it is likely that European cohorts had a narrow distribution of margarine consumption, and such a condition would contribute to the difference in the contribution of margarine to T2DM-related dietary patterns.

Discordance of dietary patterns may have been driven by different dietary instruments (FFQs). For example, polyunsaturated margarine was ascertained in the WS, but not in the FOS, and it had a relatively high contribution in the WS cohort (7). The difference in dietary instruments may drive the difference in dietary patterns and further predictability of T2DM risk. This is supported by the observation that the FFQ of the FOS was not

TABLE 4
Results from Cox proportional hazard regression analyses estimating hazard ratios for incident type 2 diabetes with the dietary pattern score as the main independent variable in the Framingham Offspring Cohort 1991–2001 (n = 2879)[1]

	HRs and 95% CIs of quartile categories of the dietary pattern score				HRs for continuous increase of the dietary pattern score[2]
	First	Second	Third	Fourth	
Nurses' Health Study					
Exploratory					
Case (%)	19 (2.6)	27 (3.8)	47 (6.5)	65 (9.0)	
PY	4874	4905	4818	4912	
Crude	1.0	1.46	2.53	3.33	1.57
Adjusted[3]	1.0	1.49	2.55	3.22	1.58
95% CI		0.82, 2.68	1.48, 4.38	1.93, 5.38	1.37, 1.83
Confirmatory					
Case (%)	18 (2.5)	38 (5.3)	36 (5.0)	66 (9.2)	
PY	4904	4868	4931	4805	
Crude	1.0	2.16	2.03	3.88	1.39
Adjusted[3]	1.0	2.21	2.07	4.14	1.44
95% CI		1.25, 3.88	1.16, 3.67	2.45, 6.99	1.25, 1.66
EPIC Study					
Exploratory					
Case (%)	13 (1.8)	33 (4.6)	40 (5.6)	72 (10.0)	
PY	4906	4879	4852	4872	
Crude	1.0	2.58	3.21	5.50	1.57
Adjusted[3]	1.0	2.73	3.41	5.46	1.60
95% CI		1.43, 5.23	1.81, 6.41	3.02, 9.87	1.39, 1.83
Confirmatory					
Case (%)	31 (4.3)	38 (5.3)	44 (6.1)	45 (6.3)	
PY	4909	4848	4896	4856	
Crude	1.0	1.26	1.45	1.53	1.14
Adjusted[3]	1.0	1.40	1.45	1.55	1.14
95% CI		0.86, 2.26	0.91, 2.30	0.98, 2.46	0.99, 1.32
Whitehall II Study					
Exploratory					
Case (%)	18 (2.5)	28 (3.9)	40 (5.6)	72 (10.0)	
PY	4906	4857	4878	4868	
Crude	1.0	1.61	2.27	4.03	1.55
Adjusted[3]	1.0	1.60	2.27	4.02	1.60
95% CI		0.88, 2.90	1.30, 3.97	2.39, 6.75	1.39, 1.83
Confirmatory					
Case (%)	29 (4.0)	35 (4.9)	45 (6.3)	49 (6.8)	
PY	4900	4901	4860	4848	
Crude	1.0	1.17	1.58	1.74	1.16
Adjusted[3]	1.0	1.29	1.63	1.81	1.16
95% CI		0.78, 2.11	1.01, 2.61	1.14, 2.87	1.00, 1.35

[1] EPIC, European Prospective Investigation into Cancer and Nutrition Potsdam Study; HR, hazard ratio; PY, person-years.
[2] HRs according to linear increase by 1 SD of each dietary pattern score in the Cox proportional hazard regression model.
[3] Adjusted model included sex, age (<50, 50–64, or ≥65 y), parental history of diabetes (yes or no), blood pressure treatment (yes or no), caloric intake (quintiles), and weight change (quintiles).

substantially different from that of the NHS, but different from those of the EPIC and WS. In addition to the difference in dietary instruments, food grouping may be the source of difference in dietary patterns. However, we observed that food grouping was not important, because the predictability of T2DM risk was not substantially different between the 3 exploratory dietary pattern scores.

Both the exploratory and confirmatory dietary patterns indicate that beverages are important in dietary patterns related to T2DM risk. The top positive and negative contributors included consumption of wine, tea, and both sugar-sweetened and low-calorie soft drinks. Beverage consumption, including dairy products and fruit juice, needs further evaluation with regard to dietary patterns and careful interpretation of past epidemiologic studies of beverages and T2DM risk (41, 42, 45–47).

Dietary patterns based on the 2 European studies, but not those based on the NHS, more weakly predicted T2DM than did those based on the FOS. The NHS-based dietary pattern, although derived in American women, was applicable to American men for the prediction of T2DM risk, according to sex-stratified analyses (Supplementary Section III under "Supplemental data" in the online issue). The difference in dietary patterns for prediction of T2DM risk could be because of regional, social, or cultural differences, because influences by these factors on dietary

TABLE 5
Results from multivariate Cox proportional hazard regression analyses estimating the ratio of the 2 hazard ratios (HRs) on the basis of confirmatory and exploratory reduced rank regression (HRc/HRe) in the Framingham Offspring Cohort 1991–2001 (n = 2879)[1]

	HRc/HRe and 95% CIs of each of the quartile categories of dietary pattern scores				HRc/HRe for continuous increase in the dietary pattern score[2]
	First	Second	Third	Fourth	
Nurses' Health Study					
Crude	1.0	1.48	0.80	1.17	0.89
Adjusted[3]	1.0	1.49	0.81	1.29	0.91
95% CI		0.69, 3.19	0.42, 1.58	0.74, 2.25	0.82, 1.01
EPIC Study					
Crude	1.0	0.49	0.45	0.28	0.76
Adjusted[3]	1.0	0.51	0.43	0.29	0.76
95% CI		0.23, 1.13	0.20, 0.92	0.14, 0.59	0.64, 0.90
Whitehall II Study					
Crude	1.0	0.73	0.70	0.43	0.76
Adjusted[3]	1.0	0.79	0.71	0.46	0.75
95% CI		0.37, 1.72	0.37, 1.36	0.24, 0.88	0.62, 0.90

[1] EPIC, European Prospective Investigation into Cancer and Nutrition Potsdam Study.
[2] The ratio of 2 HRs according to linear increase by 1 SD in each dietary pattern score.
[3] Adjusted model included sex, age (<50, 50–64, or ≥65 y), parental history of diabetes (yes or no), blood pressure treatment (yes or no), caloric intake (quintiles), and weight change (quintiles).

patterns were reported previously (48–50). Furthermore, these sociodemographic characteristics should be reflected to each dietary instrument (51). Although the cohorts we used in this report were mainly white, other differences in the study populations possibly led to the observed differences in RRR-based dietary patterns and the predictability of T2DM risk. The results indicate the difficulty in drawing a single dietary pattern or even a single set of dietary recommendations for the prevention and treatment of chronic diseases.

The strengths of our study include its longitudinal analyses, which used data from a well-characterized, community-based, prospective cohort study, and the availability of high-quality data for both diet and standard risk factors for T2DM. One of the limitations of our study was the issue of multiple testing due to the multiple scores statistically examined in the single cohort (25). However, our permutation analysis indicated statistical significance of our simultaneous inference that dietary patterns of the both European cohorts were not as predictive of incident T2DM as were those of the FOS. Therefore, multiple testing may not matter in our overall conclusion. Another limitation was that the interpretation of our findings related to the European-derived scores was complicated by the aforementioned differences in FFQs and food groupings. Therefore, additional studies are needed to evaluate the generalizability of dietary patterns by confirmatory analyses with similar instruments or by cross-confirmatory analyses in which both exploratory and confirmatory evaluations were performed collaboratively in multiple studies. Similarly to the food groups, the selection of multiple response variables in RRR analyses may also affect generalizability. Each of the RRR-derived dietary patterns in this report used different markers of T2DM risk (Table 1). However, in the NHS and FOS cohorts, despite no common response variables, the exploratory and confirmatory scores based on NHS food grouping showed similar overall results. Therefore, based on this observation as well as the sensitivity analyses in our study (*see* Supplementary Table 5 under "Supplemental data" in the online issue) and in the EPIC study (6), variation in the response variables may not substantively affect the derived dietary pattern and the predictability of incident T2DM. Nevertheless, additional studies are needed to test the generalizability of RRR-based dietary patterns in different cohorts with the use of the same metabolic biomarkers.

In conclusion, our confirmatory and exploratory dietary pattern analyses using data from the US and the 2 European studies identified differences in dietary patterns between the study populations that significantly influenced T2DM risk prediction. Dietary patterns related to T2DM risk were characterized by some food groups common across the cohorts, such as refined grains, beverages (except alcohol), and animal products. The other food groups, such as alcoholic beverages and specific types of fruit and vegetables, contributed differentially to dietary patterns derived from the different cohorts. Further research should exploit confirmatory analyses of dietary patterns among study populations and dietary variables. Such research will enhance our understanding of dietary pattern methods and strengthen the inference from studies using dietary patterns.

The authors' responsibilities were as follows—FI: conducted all of the statistical analyses and drafted the manuscript. All authors designed the project, prepared the statistical strategies and discussion materials, and participated in editing the manuscript. None of the authors had a conflict of interest.

REFERENCES

1. Hoffmann K, Schulze MB, Schienkiewitz A, Nothlings U, Boeing H. Application of a new statistical method to derive dietary patterns in nutritional epidemiology. Am J Epidemiol 2004;159:935–44.
2. Schulze MB, Hoffmann K. Methodological approaches to study dietary patterns in relation to risk of coronary heart disease and stroke. Br J Nutr 2006;95:860–9.
3. Hoffmann K, Zyriax B-C, Boeing H, Windler E. A dietary pattern derived to explain biomarker variation is strongly associated with the risk of coronary artery disease. Am J Clin Nutr 2004;80:633–40.
4. Weikert C, Hoffmann K, Dierkes J, et al. A homocysteine metabolism-related dietary pattern and the risk of coronary heart disease in two independent German study populations. J Nutr 2005;135:1981–8.

5. Schulze MB, Hoffmann K, Manson JE, et al. Dietary pattern, inflammation, and incidence of type 2 diabetes in women. Am J Clin Nutr 2005;82:675–84.
6. Heidemann C, Hoffmann K, Spranger J, et al. A dietary pattern protective against type 2 diabetes in the European Prospective Investigation into Cancer and Nutrition (EPIC)—Potsdam Study cohort. Diabetologia 2005;48:1126–34.
7. McNaughton SA, Mishra GD, Brunner EJ. Dietary patterns, insulin resistance, and incidence of type 2 diabetes in the Whitehall II Study. Diabetes Care 2008;31:1343–8.
8. Feinleib M, Kannel WB, Garrison RJ, McNamara PM, Castelli WP. The Framingham Offspring Study. Design and preliminary data. Prev Med 1975;4:518–25.
9. Dawber TR, Kannel WB. An epidemologic study of heart disease: the Framingham Study. Nutr Rev 1958;16:1–4.
10. Willett WC, Lenart E. Reproducibility and validity of food-frequency questionnaires. In: Willett WC, ed. Nutritional epidemiology. 2nd ed. New York, NY: Oxford University Press, 1998:101–47.
11. Willett WC, Sampson L, Stampfer MJ, et al. Reproducibility and validity of a semiquantitative food frequency questionnaire. Am J Epidemiol 1985;122:51–65.
12. Feskanich D, Rimm EB, Giovannucci EL, et al. Reproducibility and validity of food intake measurements from a semiquantitative food frequency questionnaire. J Am Diet Assoc 1993;93:790–6.
13. Giovannucci E, Colditz G, Stampfer MJ, et al. The assessment of alcohol consumption by a simple self-administered questionnaire. Am J Epidemiol 1991;133:810–7.
14. 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.
15. Willett WC. Issues in analysis and presentation of dietary data. In: Willett WC, ed. Nutritional epidemiology. 2nd ed. New York, NY: Oxford University Press, 1998:321–46.
16. Meigs JB, Hu FB, Rifai N, Manson JE. Biomarkers of endothelial dysfunction and risk of type 2 diabetes mellitus. JAMA 2004;291:1978–86.
17. Schulze MB, Hoffmann K, Kroke A, Boeing H. Risk of hypertension among women in the EPIC-Potsdam Study: comparison of relative risk estimates for exploratory and hypothesis-oriented dietary patterns. Am J Epidemiol 2003;158:365–73.
18. Spranger J, Kroke A, Mohlig M, et al. Inflammatory cytokines and the risk to develop type 2 diabetes: results of the prospective population-based European Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam Study. Diabetes 2003;52:812–7.
19. Brunner E, Stallone D, Juneja M, Bingham S, Marmot M. Dietary assessment in Whitehall II: comparison of 7 d diet diary and food-frequency questionnaire and validity against biomarkers. Br J Nutr 2001; 86:405–14.
20. Wilson PW, Meigs JB, Sullivan L, Fox CS, Nathan DM, D'Agostino RB Sr. Prediction of incident diabetes mellitus in middle-aged adults: the Framingham Offspring Study. Arch Intern Med 2007;167:1068–74.
21. Hertz-Picciotto I, Rockhill B. Validity and efficiency of approximation methods for tied survival times in Cox regression. Biometrics 1997;53: 1151–6.
22. Willett WC, Stampfer MJ. Implications of total energy intake for epidemiologic analyses. In: Willett WC, ed. Nutritional epidemiology. 2nd ed. New York, NY: Oxford University Press, 1998:273–301.
23. Hoffmann K, Pischon T, Schulz M, Schulze MB, Ray J, Boeing H. A statistical test for the equality of differently adjusted incidence rate ratios. Am J Epidemiol 2008;167:517–22.
24. Westfall PH, Young SS. P value adjustments for multiple tests in multivariate binomial models. J Am Stat Assoc 1989;84:780–6.
25. Bender R, Lange S. Adjusting for multiple testing—when and how? J Clin Epidemiol 2001;54:343–9.
26. Jenkins DJ, Hu FB, Tapsell LC, Josse AR, Kendall CW. Possible benefit of nuts in type 2 diabetes. J Nutr 2008;138:1752S–6S.
27. Dhingra R, Sullivan L, Jacques PF, et al. Soft drink consumption and risk of developing cardiometabolic risk factors and the metabolic syndrome in middle-aged adults in the community. Circulation 2007;116:480–8.
28. Palmer JR, Boggs DA, Krishnan S, Hu FB, Singer M, Rosenberg L. Sugar-sweetened beverages and incidence of type 2 diabetes mellitus in African American women. Arch Intern Med 2008;168:1487–92.
29. Schulze MB, Manson JE, Willett WC, Hu FB. Processed meat intake and incidence of type 2 diabetes in younger and middle-aged women. Diabetologia 2003;46:1465–73.
30. Fung TT, Schulze M, Manson JE, Willett WC, Hu FB. Dietary patterns, meat intake, and the risk of type 2 diabetes in women. Arch Intern Med 2004;164:2235–40.
31. Song Y, Manson JE, Buring JE, Liu S. A prospective study of red meat consumption and type 2 diabetes in middle-aged and elderly women: the Women's Health Study. Diabetes Care 2004;27:2108–15.
32. van Dam RM, Willett WC, Rimm EB, Stampfer MJ, Hu FB. Dietary fat and meat intake in relation to risk of type 2 diabetes in men. Diabetes Care 2002;25:417–24.
33. Gross LS, Li L, Ford ES, Liu S. Increased consumption of refined carbohydrates and the epidemic of type 2 diabetes in the United States: an ecologic assessment. Am J Clin Nutr 2004;79:774–9.
34. Liu S. Intake of refined carbohydrates and whole grain foods in relation to risk of type 2 diabetes mellitus and coronary heart disease. J Am Coll Nutr 2002;21:298–306.
35. Fung TT, Hu FB, Pereira MA, et al. Whole-grain intake and the risk of type 2 diabetes: a prospective study in men. Am J Clin Nutr 2002;76: 535–40.
36. Djousse L, Gaziano JM, Buring JE, Lee IM. Egg consumption and risk of type 2 diabetes in men and women. Diabetes Care 2009;32:295–300.
37. Riserus U, Willett WC, Hu FB. Dietary fats and prevention of type 2 diabetes. Prog Lipid Res 2009;48:44–51.
38. Mozaffarian D, Katan MB, Ascherio A, Stampfer MJ, Willett WC. Trans fatty acids and cardiovascular disease. N Engl J Med 2006;354:1601–13.
39. Odegaard AO, Pereira MA. Trans fatty acids, insulin resistance, and type 2 diabetes. Nutr Rev 2006;64:364–72.
40. Conigrave KM, Hu BF, Camargo CA, Stampfer MJ, Willett WC, Rimm EB. A prospective study of drinking patterns in relation to risk of type 2 diabetes among men. Diabetes 2001;50:2390–5.
41. Koppes LLJ, Dekker JM, Hendriks HFJ, Bouter LM, Heine RJ. Moderate alcohol consumption lowers the risk of type 2 diabetes: a meta-analysis of prospective observational studies. Diabetes Care 2005;28: 719–25.
42. Howard AA, Arnsten JH, Gourevitch MN. Effect of alcohol consumption on diabetes mellitus: a systematic review. Ann Intern Med 2004; 140:211–9.
43. Imamura F, Lichtenstein AH, Dallal GE, Meigs JB, Jacques PF. Confounding by dietary patterns of the inverse association between alcohol consumption and type 2 diabetes risk. Am J Epidemiol 2009;170:37–45.
44. Michels K, Sacks F. Trans fatty acids in european margarines. N Engl J Med 1995;332:541–2.
45. Schulze MB, Manson JE, Ludwig DS, et al. Sugar-sweetened beverages, weight gain, and incidence of type 2 diabetes in young and middle-aged women. JAMA 2004;292:927–34.
46. van Dam RM, Hu FB. Coffee consumption and risk of type 2 diabetes: a systematic review. JAMA 2005;294:97–104.
47. Apovian CM. Sugar-sweetened soft drinks, obesity, and type 2 diabetes. JAMA 2004;292:978–9.
48. Moore LV, Diez Roux AV, Nettleton JA, Jacobs DR Jr. Associations of the local food environment with diet quality—a comparison of assessments based on surveys and geographic information systems: the Multi-Ethnic Study of Atherosclerosis. Am J Epidemiol 2008;167:917–24.
49. Bermudez OI, Ribaya-Mercado JD, Talegawkar SA, Tucker KL. Hispanic and non-Hispanic white elders from Massachusetts have different patterns of carotenoid intake and plasma concentrations. J Nutr 2005; 135:1496–502.
50. Galobardes B, Morabia A, Bernstein MS. Diet and socioeconomic position: does the use of different indicators matter? Int J Epidemiol 2001; 30:334–40.
51. Tucker KL, Bianchi LA, Maras J, Bermudez OI. Adaptation of a food frequency questionnaire to assess diets of Puerto Rican and non-Hispanic adults. Am J Epidemiol 1998;148:507–18.