American Journal of Epidemiology

Vol. 170, No. 1
DOI: 10.1093/aje/kwp096
Advance Access publication May 8, 2009

Original Contribution

Confounding by Dietary Patterns of the Inverse Association Between Alcohol Consumption and Type 2 Diabetes Risk

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

Initially submitted November 14, 2008; accepted for publication March 25, 2009.

The ability to interpret epidemiologic observations is limited because of potential residual confounding by correlated dietary components. Dietary pattern analyses by factor analysis or partial least squares may overcome the limitation. To examine confounding by dietary pattern as well as standard risk factors and selected nutrients, the authors modeled the longitudinal association between alcohol consumption and 7-year risk of type 2 diabetes mellitus in 2,879 healthy adults enrolled in the Framingham Offspring Study (1991–2001) by Cox proportional hazard models. After adjustment for standard risk factors, consumers of $\geq$9.0 drinks/week had a significantly lower risk of type 2 diabetes mellitus compared with abstainers (hazard ratio = 0.47, 95% confidence interval (CI): 0.27, 0.81). Adjustment for selected nutrients had little effect on the hazard ratio, whereas adjustment for dietary pattern variables by factor analysis significantly shifted the hazard ratio away from null (hazard ratio = 0.33, 95% CI: 0.17, 0.64) by 40.0% (95% CI: 16.8, 57.0; $P = 0.002$). Dietary pattern variables by partial least squares showed similar results. Therefore, the observed inverse association, consistent with past studies, was confounded by dietary patterns, and this confounding was not captured by individual nutrient adjustment. The data suggest that alcohol intake, not dietary patterns associated with alcohol intake, is responsible for the observed inverse association with type 2 diabetes mellitus risk.

alcohol drinking; bias (epidemiology); confounding factors (epidemiology); diabetes mellitus, type 2; diet; factor analysis, statistical; least-squares analysis; proportional hazards models

Abbreviations: CI, confidence interval; HR_{adj}, hazard ratio adjusted for potential confounders; HR_{unadj}, hazard ratio unadjusted for potential confounders.

Residual confounding is an important concern in observational studies focusing on individual dietary factors and health outcomes. For example, habitual intake of alcohol is associated with lower incidence of cardiovascular diseases (1) and type 2 diabetes mellitus (2) and higher incidence of some types of cancer (3). However, it remains unclear whether the observed associations are fully attributable to alcohol intake itself, because of potential residual confounding by unadjusted dietary factors that covary with alcohol intake.

In observational studies relating individual dietary factors to health outcomes, confounding by dietary components is conventionally controlled by a limited number of dietary factors. Selection of the covariates is often subjective, and it is unclear whether the selected covariates minimize confounding and whether the observed association is independent of dietary patterns. To our knowledge, no study relating a single food or nutrient to a health outcome has addressed the issue of residual confounding by correlated foods and dietary patterns.

Various dietary pattern approaches have been used in the field of nutritional epidemiology (4–8). The advantage of this approach includes aggregation of the small effects of individual foods and the feasibility to examine protective or detrimental associations between overall diet and heath outcomes. This advantage raises the possibility that the dietary pattern approach can aggregate small confounding by dietary factors. Although confounding by individual foods

Correspondence to Dr. Paul F. Jacques, Jean Mayer US Department of Agriculture Human Nutrition Research Center on Aging, 711 Washington Street, Boston, MA 02111 (e-mail: paul.jacques@tufts.edu).

or nutrients may be negligible, net confounding by correlated foods or patterns could result in biased associations between individual foods or nutrients and health outcomes (9). Therefore, current dietary pattern techniques may be useful to aggregate and adjust for the net confounding by multiple correlated foods.

We used the dietary patterns approach to examine potential residual confounding of the association between habitual alcohol intake and incident type 2 diabetes mellitus. An association between alcohol and a disease is particularly susceptible to confounding by dietary patterns, because many earlier studies demonstrated that alcoholic beverages tend to be consumed with certain foods or specific dietary patterns (5, 6). In this report, we quantified the potential confounding of dietary patterns on the association between alcohol consumption and incident type 2 diabetes mellitus and determined whether the association was attributable to alcohol itself or dependent on dietary patterns related to alcohol consumption.

MATERIALS AND METHODS

Study population

The Framingham Offspring Study is a community-based, prospective, observational study initiated in 1971 among the offspring generation of the original Framingham Heart Study (10, 11). During the fifth examination cycle (1991–1995) of the Framingham Offspring Study (the baseline visit for these analyses), 3,799 participants underwent a standardized medical examination. The participants were followed up for 7 years on average from baseline to the sixth (1995–1998) and seventh (1998–2001) examinations. The times of failure and censoring for type 2 diabetes mellitus cases and noncases were determined by the baseline and follow-up examination dates. Type 2 diabetes mellitus in this study was defined as being on an oral hypoglycemic drug or insulin use or having a fasting glucose level of $\geq$126 mg/dL (7.0 mmol/L).

Dietary assessment

At the fifth examination, we assessed habitual dietary consumption using a 126-item semiquantitative food frequency questionnaire (12, 13). Participants were asked to choose 1 of 9 categories to indicate how often, on average, they had consumed given amounts of various specified foods during the past year. Nutrient intakes were calculated by multiplying the frequency of a food item with each prespecified portion size and the nutrient composition for that item. The external reproducibility and validity of the food frequency questionnaire are described elsewhere (13–16). Correlation coefficients of repeated measures of dietary pattern scores and those by different dietary assessment tools ranged from 0.45 to 0.74 (14). The correlation coefficients of alcohol intake estimates in different studies were highly consistent in various cohort studies, ranging from 0.65 to 0.88 (15–17). Habitual alcohol intake was estimated as the sum of consumption frequencies of 4 types of beverages: red wine, white wine, beer, and liquor. For these analyses, alcohol consumption was classified into 5 categories: abstainers and approximate quartile categories based on the frequency of alcoholic beverage consumption among drinkers. Other measures of alcohol intake, such as grams per day, yielded similar results.

Exclusion criteria

Individuals who had been diagnosed with type 2 diabetes mellitus at baseline ($n = 298$) were ineligible for these analyses. We excluded 381 individuals because their dietary data were deemed to be invalid on the basis of the following criteria: estimated daily caloric intake <600 kcal/day (2.51 MJ/day), $\geq$4,000 kcal/day for women, $\geq$4,200 kcal/day for men, or $\geq$12 blank items on their food frequency questionnaire (18). We excluded an additional 193 individuals without follow-up information on outcome measures and 48 individuals without information on covariates: body mass index (kg/m^2), weight change during follow-up, high density lipoprotein cholesterol, systolic or diastolic blood pressure, and a fasting glucose concentration. Those with missing variables were not associated with alcohol consumption ($P = 0.6$) and were more likely to be old, men, and current smokers and less likely to be physically active ($P < 0.05$). Multiple imputations (100 imputations) conditional on the predictors of missingness showed no differences in results and no appreciable gain in precision. For the sake of simplicity, we present only the complete-case analysis.

Statistical analyses

Descriptive statistics. Descriptive statistics were obtained for 3 of the 5 alcohol consumption groups: abstainers and the second and fourth quartile groups of drinkers. For continuous variables, means and standard deviations were estimated. For categorical variables, frequencies and percentage across groups determined by alcohol consumption were calculated. Bivariate associations were tested by analysis of variance for continuous variables or chi-square tests for categorical variables.

Cox proportional hazard model. To test the association between alcohol consumption and type 2 diabetes mellitus risk and confounding of the association by covariates, we used a Cox proportional hazard regression model to estimate the hazard ratios of 4 quartile groups of alcohol drinkers relative to abstainers. The results did not alter materially when abstainers were excluded, and the lowest drinking frequency group among drinkers was used as the referent group (data not shown), indicating no serious bias due to inclusion of abstainers or former drinkers as reviewed previously (2). Ties of the failure and censoring time were corrected by Efron approximation to yield valid hazard ratios (19). The regression model included the following covariates for statistical adjustment: age (<50, 50–64, or $\geq$65 years of age), parental history of diabetes (yes/no), body mass index (<25.0, 25.0–29.9, or $\geq$30.0 kg/m^2), hypertension (blood pressure >130/85 mm Hg or receiving therapy), hyperglycemia (fasting blood glucose 100–126 mg/dL), triglyceride concentration ($\geq$150 mg/dL), low high-density

lipoprotein cholesterol concentration (<40 mg/dL for men and <50 mg/dL for women), and sex and weight change over the follow-up (quintiles); these standard factors were shown to predict diabetes risk in the same study cohort (20). Caloric intake (quintiles) was also included in the model to allow isocaloric interpretation; a residual technique was also performed (21) but not presented because it had no impact on the results. Other variables, such as smoking status, menopausal status for women, multivitamin use, and physical activity, were considered as covariates, but those results were not presented because the inclusion of these additional covariates did not affect our estimates.

Statistical tests for confounding. Statistical tests for confounding were conducted by estimating a ratio of the hazard ratio adjusted for potential confounders (HR_{adj}) and the hazard ratio unadjusted for these factors (HR_{unadj}) as a measure of confounding, that is, HR_{adj}/HR_{unadj} ratio (22). For example, if the ratio was 0.9, adjustment for the tested covariates would reduce the hazard ratio by 10%. Comparing the hazard ratio of those with the highest drinking frequencies with that of abstainers, we tested for confounding by the following nutrients: saturated fatty acids, polyunsaturated fatty acids, *trans*-fatty acids, glycemic index, and dietary fiber. These were used as covariates in the previous studies relating alcohol consumption to type 2 diabetes mellitus risk (23, 24). As in the past studies, these dietary covariates were residualized separately by total caloric intake and then converted to categorical variables (quintile) (21).

Dietary pattern variables. To address our main hypotheses, we tested confounding on the hazard ratio estimates by dietary patterns. Variables representing dietary patterns were generated by maximum likelihood factor analysis and partial least-square analysis (6, 8, 25, 26). Both methods derived uncorrelated latent variables; in factor analysis, variables were generated to linearly predict covariance among food groups including alcoholic beverages; in partial least-square analysis, variables were generated to predict variation of 4 alcoholic beverages (beer, red wine, white wine, and liquor) by food groups. Coefficients for the combination were interpreted as characteristics of dietary patterns derived from factor analysis and partial least-square analysis. Other latent variable techniques were considered but not presented. Principal component analysis using the same food groups would derive components that are functions of alcoholic beverages, and therefore including the components as covariates in the regression model with the alcohol variable would be overadjustment (25). Reduced rank regression was used to derive variables predicting alcoholic beverages, and the use of these variables yielded results similar to those from partial least-square analysis (8, 25). The further methodological comparison is not presented because it is beyond the scope of this report.

For both factor analysis and partial least-square analysis, 40 food groups were created by grouping the 126 food items of the food frequency questionnaire, including alcoholic beverages (27). The food group variables were logarithmically transformed to improve normality, after adding 1 to all variables to avert 0 for the log-transformation.

Factor analysis was used to derive 3 factors predicting the 40 food group variables including the 4 alcoholic beverages, involving orthogonal rotation to balance total eigenvalues. Different numbers of factors were considered but are not presented. The scree plot and eigenvalue <1.0 of prerotated factors both supported the 3-factor solution; when 1 or 2 factors were selected, the derived factors were little correlated with alcoholic beverage consumption and therefore considered inappropriate to test confounding; when ≥4 factors were selected, the overall conclusion was not substantially different from the presented results (data not shown). Latent variables were also derived from factor analysis not including alcoholic beverages as input variables, and use of the derived variables as covariates did not change our conclusion. These results are not shown, as this approach was considered inappropriate to capture correlations among alcoholic beverages and other foods.

For partial least-square analysis, the 3 latent variables were derived as a linear combination of the 36 food group variables predicting the 4 variables of alcoholic beverages: beer, white wine, red wine, and liquor. More variables could be derived, but not used, because of little impact on our results.

After derivation of the latent variables from factor analysis or partial least-square analysis, we used each or all of the latent variables as covariates in Cox proportional hazard models. Because the adjustment is analogous to adjustment for dietary factors correlated with alcoholic beverages, we also tested confounding by 36 food groups not including alcoholic beverages; the 36 log-transformed food group variables were simultaneously included in the regression model. Statistical adjustment for neither the latent variables nor the food group variables indicated issues of multicollinearity or overadjustment, according to variance inflation factors (<2.5) for variables of alcohol consumption in any models (28).

All statistical analyses were performed by SAS, version 9.1, software (SAS Institute, Inc., Cary, North Carolina). Results were considered to be statistically significant if the associated 2-sided *P* values were less than 0.05 given the null hypothesis of HR_{adj}/HR_{unadj} equal to 1.0. Excluding heavy drinkers (i.e., those drinking 4 or more drinks per day, $N = 96$) from the analyses as outliers did not alter our conclusions; hence, they were included in the highest frequency category of alcohol consumption.

Alternatively, using the same data from the Framingham Offspring Study cohort as presented herein, we also performed Cox proportional hazard regression analyses to test whether dietary patterns were associated with incident type 2 diabetes mellitus. These analyses are presented in supplementary material posted to the *Journal*'s website (http://aje.oxfordjournals.org/).

RESULTS

Compared with abstainers, those individuals with the highest frequencies of alcoholic beverage consumption were less likely to be women or to have a parental history of diabetes mellitus (Table 1). They had a significantly lower mean body mass index, dietary glycemic index, and intakes of dietary fiber, saturated fatty acids, and *trans*-fatty acids.

Table 1. Descriptive Statistics of Participants in the Framingham Offspring Cohort ($N = 2,879$) According to Alcoholic Beverage Consumption Category, 1991–2001

Characteristics of Participants[a]	Alcoholic Beverage Consumption, drinks/week[b]			*P* Value[c]
	0	1.1–3.4	≥9.0	
Alcohol, median g/day (range)	0	3.6 (1.5–7.6)	33.9 (14.5–160.8)	
No. of subjects	754	574	582	
Age, years	55.3 (10.0)	52.8 (9.5)	53.9 (9.4)	0.23
Women, %	62.6	59.1	33.3	<0.001
Current smoking, %	20.3	15.5	24.1	<0.001
Parental history of diabetes, %	17.6	20.7	12.2	<0.001
Body mass index, kg/m^2	27.5 (5.0)	27.3 (5.1)	26.9 (4.1)	0.04
Weight change, kg	1.8 (7.5)	2.4 (6.5)	1.2 (5.8)	0.02
Systolic blood pressure, mm Hg	125.1 (19.2)	122.3 (17.5)	127.0 (17.8)	<0.001
Diastolic blood pressure, mm Hg	73.7 (10.2)	74.1 (9.9)	75.4 (9.9)	<0.001
Hypertension (>130/85 mm Hg or receiving therapy), %	39.7	34.0	39.9	0.05
Total cholesterol, mg/dL	202.6 (36.8)	202.6 (35.6)	210.3 (36.2)	0.003
Triglyceride, mg/dL	147.7 (91.9)	128.5 (74.7)	143.6 (107.5)	<0.001
HDL cholesterol, mg/dL	47.3 (13.8)	51.6 (14.5)	54.0 (15.7)	0.002
LDL cholesterol, mg/dL	125.8 (32.8)	125.3 (32.7)	127.6 (34.1)	<0.001
Fasting glucose, mg/dL	94.2 (9.6)	94.5 (9.4)	96.7 (9.4)	<0.001
Total energy, kcal/day	1,844.7 (639.7)	1,821.2 (618.5)	2,081.0 (610.2)	0.29
Glycemic index	55.3 (3.6)	54.8 (3.2)	53.6 (3.5)	<0.001
Dietary fiber, g/1,000 kcal/day	9.9 (3.4)	10.3 (3.1)	8.5 (2.5)	0.02
Saturated fatty acids, % energy	10.8 (3.0)	10.6 (2.9)	9.8 (2.6)	<0.001
trans-Fatty acids, % energy	1.6 (0.9)	1.5 (0.7)	1.3 (0.6)	<0.001
PUFA, % energy	5.9 (1.7)	5.8 (1.5)	5.5 (1.7)	0.42

Abbreviations: HDL, high density lipoprotein; LDL, low density lipoprotein; PUFA, polyunsaturated fatty acids.

[a] For continuous variables, means and standard deviations are presented.

[b] Drinkers were divided into 4 groups by approximate quartile values of alcoholic beverage consumption among drinkers; the abstainers and the second and fourth quartile groups of alcoholic beverage consumption are presented for simplicity.

[c] *P* values for the association of alcohol consumption with covariates were tested by analysis of variance for age for continuous variables and by chi-square tests for categorical variables.

Those with the highest frequency of alcohol consumption had significantly higher mean blood concentrations of total cholesterol, high density lipoprotein cholesterol, low density lipoprotein cholesterol, and fasting glucose, as well as a higher mean diastolic blood pressure. There were significant nonlinear associations for smoking status, hypertension, weight change, systolic blood pressure, and plasma triglycerides across the alcohol consumption groups. Of these variables, the lowest or the highest prevalence or mean value was observed in the light drinker group.

Loadings obtained from factor analysis and partial least-square analysis are depicted in Table 2. The matrix from factor analysis indicates how consumption of alcoholic beverages was correlated with other food groups and that 3 patterns were typical in this population. For example, according to the factor we named ''Western,'' consumption of meat and processed meat was positively associated with beer consumption, but not with red or white wine. Partial least-square analysis indicated food groups predicting the variability of different types of alcoholic beverages. Pattern solutions were similar to those from factor analysis with respect to correlations between food and alcoholic beverage consumption.

Results from Cox proportional hazard regression analyses (Table 3) show that individuals with a higher frequency of alcoholic beverage consumption had a lower risk of diabetes. The crude model and the model adjusted for diabetes risk factors showed hazard ratios of 0.45 (95% confidence interval (CI): 0.27, 0.75) and 0.47 (95% CI: 0.27, 0.81), respectively. Adjustment for individual nutrients shifted the hazard ratio slightly; including all of the selected nutrients resulted in a hazard ratio of 0.53 (95% CI: 0.29, 0.94).

Results from the tests of confounding by the dietary pattern variables are also presented in Table 3. Adjustment for Western and prudent patterns had little impact on the hazard ratios of interest, whereas the alcohol pattern shifted the adjusted hazard ratios away from the null more than 35%.

Table 2. Factor Loadings From Factor Analysis and Partial Least Squares Using 40 Food Groups in the Framingham Offspring Cohort ($N = 2,879$), 1991–2001

Food Group	Factor Analysis[a]			Partial Least Squares[a]		
	Western	Prudent	Alcohol	Beer Drinkers	Wine Drinkers	Liquor Drinkers
Beer[a]	0.21	−0.29	0.09	0.87	−0.01	0.30
Red wine[a]	−0.01	−0.16	0.22	0.19	0.47	0.15
White wine[a]	−0.06	−0.09	0.32	0.05	0.81	0.09
Liquor[a]	0.07	−0.17	0.15	0.46	0.36	0.94
High-fat dairy	0.32	0.01	0.14	0.13	−0.08	−0.28
Reduced-fat dairy	−0.08	0.39	0.00	−0.25	−0.06	−0.01
High-fat dairy desserts	0.35	0.12	−0.06	0.05	−0.20	−0.11
Low-fat dairy desserts	−0.01	0.17	0.05	−0.13	0.01	−0.05
Margarine	0.18	0.25	−0.01	−0.11	−0.07	0.06
Nondairy creamers	0.06	−0.02	0.01	0.02	−0.06	−0.07
Fruit juices	0.07	0.26	0.14	0.07	0.21	0.17
Fruits	−0.18	0.50	0.35	−0.34	0.13	−0.32
Fruit drinks	0.18	0.15	0.08	0.03	−0.06	−0.22
Tofu and beans	−0.01	0.18	0.30	−0.12	0.08	−0.15
Nuts and seeds	0.30	0.24	0.18	0.27	0.06	0.10
Vegetables	−0.11	0.22	0.77	0.04	0.48	−0.18
Starchy vegetables	0.26	0.23	0.27	−0.01	−0.01	−0.18
Eggs	0.39	−0.01	0.05	0.26	−0.09	−0.04
Poultry	0.05	0.14	0.35	−0.03	0.15	−0.30
Processed meat	0.61	−0.09	−0.05	0.39	−0.26	−0.03
Liver	0.14	0.00	0.10	0.07	0.06	0.04
Meat	0.55	−0.06	0.12	0.28	−0.05	−0.04
Fish and other seafood	0.00	0.19	0.41	0.10	0.33	0.01
Whole-grain cereal	−0.06	0.29	0.07	−0.14	0.04	−0.09
Refined-grain cereal	−0.06	0.33	0.06	−0.29	−0.05	−0.14
Refined grains	0.34	0.17	0.13	0.02	0.02	−0.35
Whole grains	−0.02	0.35	0.29	−0.07	0.22	−0.10
Pasta	0.12	0.09	0.29	0.10	0.18	−0.44
Chocolate	0.38	0.13	−0.11	−0.07	−0.25	−0.33
Candy without chocolate	0.12	0.20	−0.05	−0.26	−0.11	−0.14
Sweet baked goods	0.46	0.37	−0.07	−0.13	−0.20	−0.26
Miscellaneous sweets	0.12	0.42	0.06	−0.16	0.07	0.00
Vegetable oils	0.19	0.04	0.47	0.18	0.39	−0.18
Chowder/cream soup	0.32	0.05	0.14	0.26	0.14	0.13
Soda	0.38	−0.06	−0.16	0.15	−0.30	−0.09
Low-calorie soda	0.04	0.00	0.10	0.08	0.06	0.11
Coffee and tea	0.12	−0.04	0.11	0.14	0.17	−0.02
Pizza, sandwich, casserole	0.36	−0.01	0.09	0.21	0.00	−0.32
Potato or corn chips	0.48	−0.08	0.05	0.30	0.00	−0.23
Fried foods	0.48	−0.14	−0.03	0.24	−0.21	−0.20

[a] In factor analysis, 4 alcoholic beverages were included as input variables in addition to the other 36 food group variables. In partial least squares, 4 alcoholic beverages were used as the outcome set, whereas the other 36 food group variables were used as the predictor set.

Table 3. Results From Cox Proportional Hazard Regression Analyses to Estimate Hazard Ratios of Alcohol Consumers to Abstainers and the Ratio of 2 Hazard Ratios (HR_{adj}/HR_{unadj}) as an Indicator of Confounding in the Framingham Offspring Cohort ($N = 2,879$), 1991–2001

	Alcoholic Beverage Consumption Frequency, drinks/week									HR_{adj}/HR_{unadj} Ratio[a]	
	0	0.1–1.0		1.1–3.4		3.5–8.9		≥9.0			
		Hazard Ratio	95% Confidence Interval	Hazard Ratio	95% Confidence Interval	Hazard Ratio	95% Confidence Interval	Hazard Ratio	95% Confidence Interval	Hazard Ratio	95% Confidence Interval
No. of cases	54	23		37		24		20			
%	7.2	5.8		6.5		4.2		3.4			
Person-years, no.	5,063.2	2,698.4		3,881.9		3,880.4		3,984.8			
Crude model	1.0	0.78	0.48, 1.27	0.91	0.60, 1.38	0.58	0.36, 0.94	0.45	0.27, 0.75		
Adjusted model[b]	1.0	0.97	0.59, 1.60	0.99	0.65, 1.53	0.68	0.41, 1.11	0.47	0.27, 0.81	Referent	
+ SFA[c]	1.0	0.99	0.60, 1.63	0.99	0.64, 1.52	0.68	0.41, 1.12	0.47	0.27, 0.83	1.01	0.93, 1.10
+ PUFA	1.0	0.97	0.59, 1.58	1.05	0.68, 1.62	0.72	0.44, 1.18	0.50	0.29, 0.87	1.06	0.96, 1.17
+ *trans*-Fatty acids	1.0	0.98	0.59, 1.61	1.01	0.66, 1.57	0.68	0.42, 1.13	0.47	0.27, 0.82	1.02	0.94, 1.10
+ Glycemic index	1.0	0.94	0.57, 1.55	0.98	0.64, 1.51	0.67	0.41, 1.10	0.47	0.27, 0.83	1.01	0.94, 1.09
+ Dietary fiber	1.0	0.99	0.60, 1.63	0.99	0.64, 1.52	0.68	0.41, 1.12	0.47	0.27, 0.82	1.02	0.96, 1.08
+ All of the above	1.0	0.96	0.58, 1.59	1.06	0.68, 1.65	0.71	0.43, 1.17	0.53	0.29, 0.94	1.13	0.93, 1.37
Variables derived from factor analysis											
+ Western pattern	1.0	0.97	0.59, 1.58	1.00	0.65, 1.54	0.69	0.42, 1.12	0.49	0.28, 0.84	0.95	0.90, 1.01
+ Prudent pattern	1.0	0.98	0.60, 1.60	0.98	0.63, 1.51	0.65	0.39, 1.07	0.41	0.22, 0.76	0.92	0.78, 1.09
+ Alcohol pattern	1.0	0.96	0.59, 1.58	0.89	0.57, 1.38	0.56	0.34, 0.93	0.34	0.19, 0.62	0.74	0.61, 0.92
+ All of the above	1.0	0.96	0.58, 1.57	0.87	0.56, 1.35	0.53	0.32, 0.90	0.33	0.17, 0.64	0.60	0.43, 0.83
Variables derived from partial least squares											
+ Beer pattern	1.0	0.98	0.60, 1.61	0.98	0.64, 1.52	0.65	0.40, 1.07	0.43	0.25, 0.74	0.86	0.76, 0.96
+ Wine pattern	1.0	0.97	0.59, 1.59	0.93	0.60, 1.44	0.61	0.37, 1.01	0.39	0.22, 0.69	0.84	0.72, 0.97
+ Liquor pattern	1.0	0.97	0.59, 1.60	1.00	0.65, 1.53	0.66	0.40, 1.09	0.41	0.23, 0.74	0.96	0.89, 1.03
+ All of the above	1.0	0.98	0.60, 1.61	0.93	0.60, 1.43	0.59	0.36, 0.97	0.35	0.19, 0.63	0.70	0.56, 0.88
36 food groups	1.0	0.93	0.56, 1.56	0.95	0.60, 1.50	0.62	0.37, 1.04	0.35	0.19, 0.66	0.63	0.48, 0.84

Abbreviations: HR_{adj}, hazard ratio adjusted for potential confounders; HR_{unadj}, hazard ratio unadjusted for potential confounders; PUFA, polyunsaturated fatty acids; SFA, saturated fatty acids.

[a] The confounding test ratio is the ratio of the hazard ratio adjusted for the indicated covariates in the first column over the hazard ratio adjusted for only the standard risk variables.

[b] Covariates: age (<50, 50–64, or ≥65 years of age), parental history of diabetes (yes/no), body mass index (<25.0, 25.0–29.9, or ≥30.0 kg/m^2), hypertension (yes/no, yes if >130/85 mm Hg or receiving therapy), hyperglycemia (<100 or 100–126 mg/dL), triglyceride concentration (≥150 mg/dL or else), low high-density lipoprotein cholesterol concentration (yes/no, yes if <40 mg/dL for men and <50 mg/dL for women), sex, weight change over the follow-up period (quintiles), and total caloric intake (quintiles).

[c] All nutrients are quintiles of residuals derived from regressing on caloric intake.

Adjustment for all the dietary pattern variables from factor analysis significantly shifted the adjusted hazard ratios away from the null by 40% (HR_{adj}/HR_{unadj} ratio = 0.60, 95% CI: 0.43, 0.83; $P = 0.002$). Similarly, adjustment for the variables derived from partial least-square analysis strengthened the inverse associations between alcohol consumption and type 2 diabetes mellitus risk, particularly adjustment for the dietary patterns predicting beer drinkers and wine drinkers, but not those predicting liquor drinkers. Adjustment for all of the partial least-square analysis variables shifted the adjusted hazard ratios by 30% (HR_{adj}/HR_{unadj} ratio = 0.70, 95% CI: 0.56, 0.88; $P = 0.002$). Comparing the adjustment for these dietary pattern variables with the adjustment for 36 food groups, adjustment for partial least-square analysis variables showed a narrower 95% confidence interval for the HR_{adj}/HR_{unadj} ratio = 0.32 (0.88 − 0.56), whereas adjustment for factors from factor analysis showed a wider 95% confidence interval = 0.43; the adjustment for 36 food groups resulted in a 95% confidence interval width of 0.40. The hazard ratios adjusted for dietary pattern variables were not affected by additional adjustment for selected nutrients (data not presented).

DISCUSSION

Using dietary pattern analyses and capturing correlations between alcoholic beverages and other food groups, we found that the association between alcohol consumption

and type 2 diabetes mellitus risk was significantly negatively confounded by dietary patterns. Adjustment for dietary pattern variables showed stronger inverse associations rather than attenuation. This indicates that consumption of alcoholic beverages was correlated with dietary patterns that were positively associated with type 2 diabetes mellitus risk. Moreover, dietary pattern analyses did not capture the benefit of drinking for reduced risk of type 2 diabetes mellitus in the study population, indicating that this association was independent of dietary patterns.

The observed inverse association between alcohol consumption and type 2 diabetes mellitus risk was stronger than the one based on meta-analysis of past observational studies; the pooled risk ratio was estimated to be 0.72 (95% CI: 0.62, 0.84) comparing moderate alcohol drinkers with abstainers (2), whereas we showed hazard ratios of 0.47 after adjustment for the standard type 2 diabetes mellitus risk factors and 0.33 after additional adjustment for dietary patterns. The difference may be partially attributable to the lack of adjustment for dietary covariates in the past observational studies, under the assumption of negligible difference in study bases between studies, and therefore past studies likely underestimated the inverse association. Therefore, our study suggests that epidemiologic findings of this association were confounded by intakes of other foods, because alcoholic beverages are typically consumed with certain types of foods. Indeed, a dietary pattern characterized by high alcohol consumption was often identified in prior studies (5, 6).

We demonstrated that dietary pattern analyses can aggregate small confounding effects by correlated foods and that derived pattern variables were useful for confounding adjustment. Using a dietary pattern technique is a novel approach to address the issue of confounding by multiple correlated food groups (9). Importantly, adjustment for dietary pattern variables did not result in attenuation of the association between alcohol consumption and type 2 diabetes mellitus risk. This indicates that the dietary pattern variables failed to capture the benefit of alcohol consumption for type 2 diabetes mellitus risk reduction. This is plausible because factor and partial least-square analyses derive variables solely by capturing correlations among food groups (6–8). Dietary pattern analyses may succeed in describing certain correlations among food groups, but not in characterizing the benefit of individual food groups. On the basis of our findings, dietary pattern analyses appear to be useful to characterize how food consumption covaries and to use the correlations for analyses and interpretation in studies of individual foods or nutrients.

To obtain dietary pattern variables, we used factor analysis and partial least-square analysis (6–8). Factor analysis derived the variables from correlations between food groups including alcoholic beverages, whereas partial least-square analysis derived the variables from regression between alcoholic beverages and other food groups. The 2 sets of dietary pattern variables resulted in similar outcomes when used for statistical adjustment in examining the association between alcohol consumption and type 2 diabetes mellitus. As expected, the 2 techniques were informative about dietary patterns; factor analysis showed dietary patterns including alcoholic beverages, and partial least-square analysis showed the food groups predicting alcoholic beverages consumption. This information provided by the techniques and the usefulness of the derived variables as covariates may offer advantages over simultaneous statistical adjustment for food groups used as independent exploratory covariates. A slight difference appeared in the precision estimates from these techniques; in the present example, partial least-square analysis provided narrower confidence intervals. Wider confidence intervals, by using variables derived from factor analysis, are likely because the derivation of variables captured the variation unrelated to confounding from the correlations between alcohol and certain food groups. With regard to the information and usefulness of dietary patterns, the choice of partial least-square analysis, factor analysis, or other latent variable techniques, such as principal component analysis, needs further research. Use of these techniques should also be investigated regarding several assumptions, such as distribution normality and linear association among dietary variables.

It is noteworthy that our approach is similar to the propensity score approach with respect to application of a latent variable (29, 30). A propensity score is generated by scoring the estimated probability of having a certain exposure level, and it is used as a covariate in a regression analysis, a matching factor, a stratification variable, or an inverse probability weight to control for confounding. We generated dietary pattern variables to collapse between-foods correlations and used the variables as covariates in regression analyses. In contrast to the propensity score technique, dietary pattern techniques allow practical interpretation of diet. Compared with studies of the propensity score technique with real and simulated data (31, 32), adopting a dietary pattern approach for statistical adjustment may be premature. However, this technique shows potential to deal with multiple correlated confounders, and more discussion is warranted.

Strengths of our study include the longitudinal analyses based on a well-characterized, community-based cohort and the availability of high-quality data for both standard risk factors for type 2 diabetes mellitus and diet. One potential limitation of this study is that we could not test the previously reported hypothesis of *U*-shaped association between alcohol and type 2 diabetes mellitus risk (2), because too few subjects in our cohort reported heavy drinking, defined as $\geq$5 drinks per day. Hence, this limits the generalizability of our results regarding heavy drinkers. Another limitation is the use of frequency data to assess habitual alcoholic intake. We could not examine potential associations between type 2 diabetes mellitus risk and consumption patterns, such as variability by day of the week, binge drinking, and temporal relations of alcohol and food intake.

In summary, our findings from the Framingham Offspring Study demonstrated that light to moderate alcohol consumption was inversely associated with type 2 diabetes mellitus risk, and that this association was independent of differences in diet between abstainers and drinkers. We demonstrated that the association was negatively confounded by

correlations among food groups, and that adjustment for the correlated food groups by a dietary pattern approach strengthened the inverse association. In addition, we have shown that this dietary pattern approach to adjust for confounding can be useful in examining potential associations between individual foods or nutrients and disease outcomes. In conclusion, our findings add further evidence to support the hypothesis that moderate alcohol consumption is associated with a decreased risk for type 2 diabetes mellitus; however, we cannot recommend moderate alcohol consumption solely for the prevention of type 2 diabetes mellitus because of the potential adverse health effects associated with alcohol consumption (33, 34).

ACKNOWLEDGMENTS

Author affiliations: Jean Mayer US Department of Agriculture Human Nutrition Research Center on Aging and the Friedman School of Nutrition Science and Policy, Tufts University, Boston, Massachusetts (Fumiaki Imamura, Alice H. Lichtenstein, Gerard E. Dallal, Paul F. Jacques); and General Medicine Division and Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts (James B. Meigs).

This work was supported in part by the US Department of Agriculture (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). J. B. M. is supported by the National Institute of Diabetes and Digestive and Kidney Diseases (K24-DK090140).

The authors' opinions in this publication do not necessarily reflect the view of the US Department of Agriculture.

Conflict of interest: none declared.

REFERENCES

1. Rimm EB, Klatsky A, Grobbee D, et al. Review of moderate alcohol consumption and reduced risk of coronary heart disease: is the effect due to beer, wine, or spirits? *BMJ*. 1996; 312(7033):731–736.
2. Koppes LL, Dekker JM, Hendriks HF, et al. Moderate alcohol consumption lowers the risk of type 2 diabetes: a meta-analysis of prospective observational studies. *Diabetes Care*. 2005; 28(3):719–725.
3. Boffetta P, Hashibe M. Alcohol and cancer. *Lancet Oncol*. 2006;7(2):149–156.
4. Jacobs DR Jr, Steffen LM. Nutrients, foods, and dietary patterns as exposures in research: a framework for food synergy. *Am J Clin Nutr.* 2003;78(3 suppl):508S–513S.
5. Togo P, Osler M, Sørensen TI, et al. Food intake patterns and body mass index in observational studies. *Int J Obes Relat Metab Disord*. 2001;25(12):1741–1751.
6. Newby PK, Tucker KL. Empirically derived eating patterns using factor or cluster analysis: a review. *Nutr Rev.* 2004; 62(5):177–203.
7. Hu FB. Dietary pattern analysis: a new direction in nutritional epidemiology. *Curr Opin Lipidol*. 2002;13(1):3–9.
8. Hoffmann K, Schulze MB, Schienkiewitz A, et al. Application of a new statistical method to derive dietary patterns in nutritional epidemiology. *Am J Epidemiol*. 2004;159(10): 935–944.
9. Greenland S. Invited commentary: variable selection versus shrinkage in the control of multiple confounders. *Am J Epidemiol*. 2008;167(5):523–529; discussion 530–531.
10. Feinleib M, Kannel WB, Garrison RJ, et al. The Framingham Offspring Study. Design and preliminary data. *Prev Med.* 1975;4(4):518–525.
11. Dawber TR, Kannel WB. An epidemiologic study of heart disease: the Framingham study. *Nutr Rev.* 1958;16(1):1–4.
12. Willett WC, Lenart E. Reproducibility and validity of food-frequency questionnaires. In: Willett WC, ed. *Nutritional Epidemiology.* New York, NY: Oxford University Press; 1998:101–147.
13. Willett WC, Sampson L, Stampfer MJ, et al. Reproducibility and validity of a semiquantitative food frequency questionnaire. *Am J Epidemiol*. 1985;122(1):51–65.
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(2):243–249.
15. 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(7):790–796.
16. Giovannucci E, Colditz G, Stampfer MJ, et al. The assessment of alcohol consumption by a simple self-administered questionnaire. *Am J Epidemiol*. 1991;133(8):810–817.
17. Feunekes GI, van't Veer P, van Staveren WA, et al. Alcohol intake assessment: the sober facts. *Am J Epidemiol*. 1999; 150(1):105–112.
18. Willett WC. Issues in analysis and presentation of dietary data. In: Willett WC, ed. *Nutritional Epidemiology.* New York, NY: Oxford University Press; 1998:321–346.
19. Hertz-Picciotto I, Rockhill B. Validity and efficiency of approximation methods for tied survival times in Cox regression. *Biometrics*. 1997;53(3):1151–1156.
20. Wilson PW, Meigs JB, Sullivan L, et al. Prediction of incident diabetes mellitus in middle-aged adults: the Framingham Offspring Study. *Arch Intern Med*. 2007;167(10):1068–1074.
21. Willett WC, Stampfer MJ. Implications of total energy intake for epidemiologic analyses. In: Willett WC, ed. *Nutritional Epidemiology.* New York, NY: Oxford University Press; 1998: 273–301.
22. Hoffmann K, Pischon T, Schulz M, et al. A statistical test for the equality of differently adjusted incidence rate ratios. *Am J Epidemiol*. 2008;167(5):517–522.
23. Conigrave KM, Hu BF, Camargo CA Jr, et al. A prospective study of drinking patterns in relation to risk of type 2 diabetes among men. *Diabetes*. 2001;50(10):2390–2395.
24. Wannamethee SG, Camargo CA Jr, Manson JE, et al. Alcohol drinking patterns and risk of type 2 diabetes mellitus among younger women. *Arch Intern Med*. 2003;163(11): 1329–1336.
25. Yung YF, Sall JP, Sarle WS. The FACTOR procedure. In: Jones A, Huddleston E, eds. *SAS/STAT 9.2 User's Guide*. Cary, NC: SAS Institute, Inc; 2008:1533–1620.
26. Tobias RD. The PLS procedure. In: Jones A, Huddleston E, eds. *SAS/STAT 9.2 User's Guide*. Cary, NC: SAS Institute, Inc; 2008:4759–4808.
27. Liu E, McKeown NM, Newby PK, et al. Cross-sectional association of dietary patterns with insulin-resistant phenotypes among adults without diabetes in the Framingham Offspring Study. *Br J Nutr.* 2009;16:1–8.

28. Mansfield ER, Helms BP. Detecting multicollinearity. *Am Stat*. 1982;36:158–160.
29. Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. *J Am Stat Assoc.* 1984;79:516–524.
30. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983;70(1):41–55.
31. D'Agostino RB Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Stat Med*. 1998;17(19):2265–2281.
32. Stürmer T, Schneeweiss S, Rothman KJ, et al. Performance of propensity score calibration—a simulation study. *Am J Epidemiol*. 2007;165(10):1110–1118.
33. Ellison RC. Importance of pattern of alcohol consumption. *Circulation*. 2005;112(25):3818–3819.
34. *Dietary Guidelines for Americans*. Washington, DC: US Department of Agriculture and Department of Health and Human Services; 2005.