Reproducibility and validity of dietary patterns assessed with a food-frequency questionnaire[1–3]

Frank B Hu, Eric Rimm, Stephanie A Smith-Warner, Diane Feskanich, Meir J Stampfer, Albert Ascherio, Laura Sampson, and Walter C Willett

ABSTRACT

Background: Recently, the analysis of dietary patterns has emerged as a possible approach to examining diet-disease relations.

Objective: We examined the reproducibility and validity of dietary patterns defined by factor analysis using dietary data collected with a food-frequency questionnaire (FFQ).

Design: We enrolled a subsample of men (n = 127) from the Health Professionals Follow-up Study in a diet-validation study in 1986. A 131-item FFQ was administered twice, 1 y apart, and two 1-wk diet records and blood samples were collected during this 1-y interval.

Results: Using factor analysis, we identified 2 major eating patterns, which were qualitatively similar across the 2 FFQs and the diet records. The first factor, the prudent dietary pattern, was characterized by a high intake of vegetables, fruit, legumes, whole grains, and fish and other seafood, whereas the second factor, the Western pattern, was characterized by a high intake of processed meat, red meat, butter, high-fat dairy products, eggs, and refined grains. The reliability correlations for the factor scores between the 2 FFQs were 0.70 for the prudent pattern and 0.67 for the Western pattern. The correlations (corrected for week-to-week variation in diet records) between the 2 FFQs and diet records ranged from 0.45 to 0.74 for the 2 patterns. In addition, the correlations between the factor scores and nutrient intakes and plasma concentrations of biomarkers were in the expected direction.

Conclusions: These data indicate reasonable reproducibility and validity of the major dietary patterns defined by factor analysis with data from an FFQ. *Am J Clin Nutr* 1999;69:243–9.

KEY WORDS Diet, dietary pattern, factor analysis, biomarker, reproducibility, validity, men, Health Professionals Follow-up Study, food-frequency questionnaire

INTRODUCTION

Traditional analyses in nutritional epidemiology typically examine diseases in relation to a single or a few nutrients or foods. However, people do not eat isolated nutrients. Instead, they eat meals consisting of a variety of foods with complex combinations of nutrients. The single-nutrient approach may be inadequate for taking into account complicated interactions among nutrients in studies of free-living people (eg, enhanced iron absorption in the presence of vitamin C) (1). Also, the high level of intercorrelation among some nutrients (such as potassium and magnesium) makes it difficult to examine their effects separately (2). Moreover, because nutrient intakes are commonly associated with certain dietary patterns (3, 4), single-nutrient analysis may be confounded by the effect of dietary patterns (5).

To overcome these limitations, several authors recently proposed to study overall dietary patterns by considering how foods and nutrients are consumed in combination (4, 6–13). In a dietary pattern analysis, the collinearity of nutrients and foods can be used to advantage because patterns are characterized on the basis of habitual food consumption. Examination of dietary patterns would more closely parallel real-world conditions, under which dietary intakes consist of nutrients that occur together in common foods (14).

Although the concept of studying dietary patterns has elicited considerable interest, no study has been conducted to examine the reproducibility and validity of these methods. We studied the reproducibility and validity of dietary patterns defined by factor analysis using dietary data collected with a food-frequency questionnaire (FFQ) and diet records among participants in the Health Professionals Follow-up Study, who were enrolled in a nutrient validation study in 1986.

SUBJECTS AND METHODS

Study population

Men in this study were participants in the Health Professionals Follow-up Study, a prospective study of risk factors for cancer and

[1] From the Departments of Nutrition and Epidemiology, Harvard School of Public Health, Boston; the Channing Laboratory, Boston; and the Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston.

[2] Supported by research grants CA55075, HL60712, and Nutrition Training Grant T32DK07703 from the National Institutes of Health. FBH was the recipient of the Charles A King Trust Research Fellowship from The Medical Foundation, Boston.

[3] Address reprint requests to FB Hu, Department of Nutrition, Harvard School of Public Health, 665 Huntington Avenue, Boston, MA 02115. E-mail: frank.hu@channing.harvard.edu.

Received February 25, 1998.

Accepted for publication July 15, 1998.

heart disease among 51 529 men aged 40–75 y at baseline in 1986 (15). Cohort members completed a mailed self-administered FFQ at baseline. During the following year, a random sample of 323 cohort members living in the Boston area were asked to participate in a dietary assessment validation study and provided blood samples; 157 men agreed to participate. We excluded men who had left >70 items blank on the FFQ, who reported a total daily energy intake outside the range of 3.3–17.6 MJ (800–4200 kcal) on either of the 2 questionnaires, or who did not provide a blood sample. A total of 127 men were included in the analyses.

Dietary assessment

The participants completed the same FFQ twice, 1 y apart (FFQ1 and FFQ2). The reproducibility and validity of nutrient and food intake measurements from the FFQ used in this study were described in detail elsewhere (16, 17). The FFQ includes 131 food items with specified serving sizes described by using natural portions (eg, 1 banana, 2 slices of pizza) or standard weight and volume measures of the servings commonly consumed in this study population. For each food item, participants indicated their average frequency of consumption over the past year in terms of the specified serving size by checking 1 of 9 frequency categories ranging from "almost never" to "≥6 times/d." The selected frequency category for each food item was converted to a daily intake. For example, a response of "2–4/wk" was converted to 0.43 servings/d (3 servings/wk).

Participants also completed two 1-wk diet records ≈6–7 mo apart. The first week's record began ≈3 mo after administration of FFQ1 and the second week's record began 2–3 mo before administration of FFQ2. Each subject was given a dietetic scale and was trained by the study dietitian to weigh and record all food consumed. The study dietitian reviewed the returned diet records and resolved questions or discrepancies with the participant. Foods reported in the diet records were coded by using the CBORD DIET ANALYZER SYSTEM (version 3.0.3, 1998; The CBORD Group, Ithaca, NY). Mixed dishes reported on the records that were not included in the CBORD system were coded by their component ingredients as described by the participant.

To obtain daily food intake measurements from the diet records that were comparable with those from the FFQs, an attempt was made to match each of the 1565 unique diet-record food codes to one or more items on the questionnaire. A total of 963 diet-record food codes were matched to a single food item on the questionnaire; multiple codes were frequently matched to the same questionnaire food item. For example, the diet-record codes for "avocado, whole" and "avocado, mashed" were the only 2 codes that were matched to the avocado food item on the questionnaire, whereas 56 diet-record codes for various brands of breakfast cereals were all matched to the same questionnaire item for cold breakfast cereal. For 254 diet-record foods that could not be matched to a single food item on the questionnaire, recipes were created and ingredients were assigned to separate food items. The remaining 348 diet-record foods (eg, asparagus, gravy, and clams) that did not match any of the questionnaire items were eliminated, usually because they were not consumed frequently in this population.

Food groupings

Because of the small number of subjects (n = 127) relative to the number of food items, we collapsed the individual food items into 40 predefined food groups (**Table 1**). The grouping scheme was based on the similarity of nutrient profiles or culinary usage among the foods and was somewhat similar to that used in other studies (13). Note that other criteria could have been used to define the number and types of food groups to be included in the analyses. Some individual food items were preserved either because it was inappropriate to incorporate them into a certain food group (eg, eggs, margarine, pizza, soup, coffee, and tea) or they were suspected to represent distinct dietary patterns (eg, garlic, liquor, wine, beer, and French fries).

Laboratory analyses

Participants provided blood samples shortly before completing FFQ2. Blood specimens from nonfasting participants were collected into EDTA-treated tubes between 0800 and 1200. The tubes were immediately covered with aluminum foil and stored in the dark on ice for up to 3 h until the plasma was separated. Plasma was stored at −70°C for up to 15 mo until analyzed. Plasma carotenoids and retinol were measured by reversed-phase HPLC in the laboratory of Hoffmann-La Roche (Basel, Switzerland) (18). Plasma cholesterol and triacylglycerol concentrations were determined according to the methods of Richmond (19) and Bucolo and David (20), respectively, by using kits from Hoffmann-La Roche.

Statistical analysis

Factor analysis (principal component) was used to derive food patterns based on the 40 food groups for each of the FFQs and the diet records. We conducted the analyses using the FACTOR PROCEDURE in SAS (21). The factors were rotated by an orthogonal transformation (Varimax rotation function in SAS) to achieve simpler structure with greater interpretability. In determining the number of factors to retain, we considered eigenvalues (>1), the Scree test (22), and the interpretability of the factors. We did not use the percentage of variance explained by each factor because this criterion depends largely on the total number of variables included in the analyses. The substantive meanings of the rotated factors were considered in conjunction with the above empirical criteria and the derived factors were labeled on the basis of our interpretation of the data as well as on prior literature.

The factor score for each pattern was computed by combining the observed variables with weights that were proportional to their component (factor) loadings (22).

$$\text{Component or factor score for pattern } i = \Sigma_j [(b_{ij}/\lambda_i)X_j] \quad (1)$$

where b_{ij} is the loading for the jth food item or group on the ith pattern, λ_i is the associated eigenvalue, and X_j is the standardized value of jth food item or group.

Pearson correlation coefficients were used to evaluate the consistency of dietary patterns derived from dietary data collected with the 2 FFQs and diet records. To reduce the within-person variation in food intake derived from the diet records, we conducted factor analysis using the average consumption for each food group across two 1-wk diet records. We also calculated deattenuated correlation coefficients for the dietary patterns between the 2 FFQs and the diet records, corrected for week-to-week variation in diet records by using the following formula (23):

$$r_t = r_o \sqrt{(1 + \gamma/k)} \quad (2)$$

where r_t is the corrected correlation between the dietary pattern scores derived from the FFQ and diet records, r_o is the observed

TABLE 1
Food groupings used in the dietary pattern analyses

Foods or food groups	Food items
Processed meats	Processed meats, bacon, hot dogs
Red meats	Beef, pork, lamb, hamburger
Organ meats	Beef, calf, and pork liver, chicken and turkey liver
Fish and other seafood	Canned tuna fish, dark-meat fish, other fish, shrimp, lobster, scallops
Poultry	Chicken or turkey with or without skin
Eggs	Eggs
Butter	Butter
Margarine	Margarine
Low-fat dairy products	Skim or low-fat milk, sherbet or ice milk, yogurt
High-fat dairy products	Whole milk, cream, sour cream, ice cream, cream cheese, other cheese
Liquor	Liquor
Wine	Red wine, white wine
Beer	Beer
Tea	Tea
Coffee	Coffee
Fruit	Raisins or grapes, avocado, bananas, cantaloupe, watermelon, fresh apples or pears, oranges, grapefruit, strawberries, blueberries, peaches, apricots, plums
Fruit juices	Apple juice or cider, orange juice, grapefruit juice, other fruit juice
Cruciferous vegetables	Broccoli; coleslaw and uncooked cabbage; cooked cabbage; cauliflower; Brussels sprouts; kale, mustard, and chard greens; sauerkraut
Dark-yellow vegetables	Carrots, yellow (winter) squash, yams
Tomatoes	Tomatoes, tomato juice, tomato sauce
Green, leafy vegetables	Spinach, iceberg or head lettuce, romaine or leaf lettuce
Legumes	String beans, peas or lima beans, beans or lentils, tofu or soybeans, alfalfa sprouts
Other vegetables	Celery, mushrooms, green pepper, corn, mixed vegetables, eggplant, summer squash
Garlic	Garlic
Potatoes	Potatoes
French fries	French fries
Whole grains	Cooked oatmeal, other cooked breakfast cereal, dark bread, brown rice, other grains, bran added to food, wheat germ
Cold breakfast cereal	Cold breakfast cereal
Refined grains	White bread, English muffins, bagels or rolls, muffins or biscuits, white rice, pasta, pancakes or waffles
Pizza	Pizza
Snacks	Potato chips or corn chips, crackers, popcorn
Nuts	Peanuts, other nuts, peanut butter
High-energy drinks	Cola with sugar, other carbonated beverages with sugar, fruit drinks
Low-energy drinks	Low-energy cola, other low-energy carbonated beverage
Oil and vinegar salad dressing	Oil and vinegar salad dressing
Mayonnaise and other creamy salad dressings	Mayonnaise and other creamy salad dressings
Chowder or cream soup	Chowder or cream soup
Other soup	Homemade soup, ready-made soup
Sweets and desserts	Chocolate bars or pieces, candy bars, cookies, brownies, doughnuts, cake, pie, sweet roll, coffee cake, pastry
Condiments	Red chili sauce (dry or prepared), mustard, pepper, soy or Worcestershire sauce, jam, jelly, syrup, honey

correlation, γ is the ratio of estimated within-person and between-person variation in dietary pattern scores derived from the two 1-wk diet records, and k is the number of repeated observations of diet records (in this study $k = 2$).

Pearson correlation coefficients were also used to assess the relation between dietary patterns and computed nutrients from diet records and plasma biochemical measurements. The nutrients for the diet records were energy adjusted by using the regression method (24). We adjusted plasma concentrations of carotenoids, tocopherols, and retinol for age, plasma cholesterol, plasma triacylglycerol, and body mass index. All biochemical measurements were $\log_e$ transformed to achieve normality.

RESULTS

Mean daily intakes, expressed in serving-size units, of the 40 foods and food groups determined from the 2 FFQs and from the average of the two 1-wk diet records for the 127 study participants are shown in **Table 2**. Foods underestimated by the FFQs compared with the diet records (ie, the gold standard) included processed meats, eggs, butter, high-fat dairy products, mayonnaise and creamy salad dressings, refined grains, and sweets and desserts, whereas most of the vegetable and fruit groups, nuts, high-energy and low-energy drinks, and condiments were overestimated by the FFQs. Pearson correlations comparing daily intakes of the food groups derived from the 2 FFQs and the diet records are also listed in Table 2. Reproducibility correlations for the comparison of the 2 FFQs ranged from 0.36 for legumes to 0.92 for coffee ($\bar{x}$: 0.70). Pearson correlation coefficients ranged from 0.09 for other vegetables to 0.83 for coffee ($\bar{x}$: 0.38) for the comparison between FFQ1 and the diet records and from 0.07 for other vegetables to 0.90 for coffee ($\bar{x}$: 0.42) for the comparison between FFQ2 and the diet records.

The factor analysis identified 2 major factors, the Western and prudent dietary patterns, which explained 20% of the vari-

TABLE 2
Daily intake of 40 foods or food groups assessed with diet records (DRs) and 2 food-frequency questionnaires completed 1 y apart (FFQ1 and FFQ2) by 127 participants in the Health Professional Follow-up Study (1986)

Foods or food groups	Serving size[1]	Servings			Pearson correlations[2]		
		FFQ1	FFQ2	DR	DR vs Q1	DR vs Q2	Q1 vs Q2
	g	*no. of servings*					
Processed meats	27 (1 oz)	0.26 ± 0.27[3]	0.22 ± 0.27	0.53 ± 0.60[4]	0.52	0.52	0.70
Red meats	112–168 (4–6 oz)	0.47 ± 0.38	0.45 ± 0.40	0.49 ± 0.28	0.50	0.59	0.88
Organ meats	28–112 (1–4 oz)	0.05 ± 0.08	0.04 ± 0.07	0.01 ± 0.06	0.21	0.26	0.64
Fish and other seafood	84–140 (3–5 oz)	0.50 ± 0.32	0.54 ± 0.34	0.52 ± 0.35	0.51	0.74	0.61
Poultry	112–168 (4–6 oz)	0.37 ± 0.20	0.36 ± 0.27	0.34 ± 0.22	0.54	0.48	0.74
Eggs	50 (one)	0.25 ± 0.26	0.22 ± 0.30	0.30 ± 0.33[4]	0.60	0.56	0.68
Butter	5 (1 pat)	0.31 ± 0.77	0.36 ± 0.92	0.74 ± 1.19[4]	0.55	0.53	0.73
Margarine	5 (1 pat)	0.62 ± 0.84	0.73 ± 0.95	0.87 ± 1.15[4]	0.61	0.50	0.62
Low-fat dairy products	245 (8 fl oz) or 122 (0.5 cup)	0.81 ± 0.94	0.82 ± 0.88	0.60 ± 0.51[4]	0.54	0.62	0.69
High-fat dairy products	244 (8 fl oz), 15 (1 Tbsp), 28 (1 oz)	1.04 ± 0.95	1.09 ± 1.09	1.63 ± 1.28[4]	0.62	0.62	0.58
Liquor	42 (1 drink)	0.37 ± 0.78	0.37 ± 0.79	0.44 ± 0.74	0.72	0.78	0.90
Wine	118 (4 fl oz)	0.41 ± 0.66	0.36 ± 0.64	0.41 ± 0.67[4]	0.66	0.68	0.82
Beer	356 (1 can)	0.22 ± 0.43	0.25 ± 0.53	0.26 ± 0.54	0.70	0.74	0.84
Tea	237 (1 cup)	0.48 ± 0.83	0.51 ± 0.91	0.52 ± 0.74	0.61	0.69	0.71
Coffee	237 (1 cup)	1.82 ± 1.60	1.76 ± 1.49	1.56 ± 1.28[4]	0.83	0.90	0.92
Fruit	128 (0.5 oz), 138 (1 oz), 28 (1 oz), 164 (1 slice)	1.71 ± 1.26	1.67 ± 1.28	1.34 ± 1.11[4]	0.50	0.67	0.71
Fruit juice	186 (small glass)	1.20 ± 1.21	1.13 ± 1.22	0.88 ± 0.73[4]	0.74	0.76	0.83
Cruciferous vegetables	78 (0.5 cup)	0.40 ± 0.38	0.39 ± 0.33	0.26 ± 0.31[4]	0.25	0.26	0.61
Dark-yellow vegetables	103 (0.5 cup)	0.42 ± 0.43	0.36 ± 0.31	0.22 ± 0.35[4]	0.33	0.28	0.64
Tomatoes	123 (one)	0.69 ± 0.41	0.66 ± 0.42	0.36 ± 0.26[4]	0.28	0.38	0.57
Green, leafy vegetables	90 (0.5 cup)	0.75 ± 0.50	0.79 ± 0.51	0.64 ± 0.45	0.51	0.55	0.40
Legumes	131 (0.5 cup)	0.43 ± 0.32	0.42 ± 0.32	0.26 ± 0.28[4]	0.30	0.46	0.36
Other vegetables	131 (0.5 cup)	0.82 ± 0.64	0.84 ± 0.61	0.60 ± 0.48[4]	0.09	0.07	0.58
Garlic	2.8 (1 clove, 1 shake)	0.17 ± 0.26	0.17 ± 0.24	0.06 ± 0.28[4]	0.13	0.14	0.74
Potatoes	202 (1 whole), 210 (1 cup)	0.35 ± 0.31	0.36 ± 0.32	0.25 ± 0.18[4]	0.48	0.45	0.67
French fries	112 (4 oz)	0.10 ± 0.17	0.08 ± 0.14	0.06 ± 0.08[4]	0.30	0.41	0.67
Whole grains	25 (1 slice), 158 (1 cup)	1.11 ± 1.17	1.08 ± 1.23	1.14 ± 1.08	0.31	0.27	0.57
Cold breakfast cereal	30 (1 cup)	0.41 ± 0.41	0.40 ± 0.39	0.46 ± 0.53	0.56	0.77	0.69
Refined grains	25 (1 slice), 158 (1 cup)	1.68 ± 1.28	1.71 ± 1.25	2.19 ± 0.98[4]	0.23	0.30	0.47
Pizza	240 (2 slices)	0.08 ± 0.08	0.09 ± 0.09	0.07 ± 0.10[4]	0.26	0.29	0.60
Snacks	28 (one or 1 oz)	0.41 ± 0.37	0.46 ± 0.43	0.84 ± 0.94[4]	0.41	0.36	0.44
Nuts	28 (1 oz)	0.45 ± 0.56	0.43 ± 0.62	0.32 ± 0.50[4]	0.38	0.45	0.53
High-energy drinks	370 (1 can)	0.41 ± 0.72	0.41 ± 0.75	0.23 ± 0.31[4]	0.46	0.51	0.54
Low-energy drinks	355 (1 can)	0.42 ± 0.84	0.39 ± 0.73	0.18 ± 0.35[4]	0.74	0.66	0.78
Oil and vinegar salad dressing	15 (1 tbsp)	0.34 ± 0.34	0.37 ± 0.38	0.30 ± 0.33	0.22	0.21	0.54
Mayonnaise and creamy salad dressings	14 (1 tbsp)	0.24 ± 0.33	0.25 ± 0.26	0.64 ± 0.49[4]	0.31	0.41	0.49
Chowder or cream soup	248 (1 cup)	0.21 ± 0.23	0.20 ± 0.28	0.21 ± 0.30	0.20	0.41	0.69
Other soup	240 (1 cup)	0.07 ± 0.11	0.07 ± 0.10	0.07 ± 0.10	0.26	0.44	0.33
Sweets or desserts	28 (1 oz), 64 (1 slice)	1.29 ± 1.40	1.17 ± 0.99	1.58 ± 1.49[4]	0.38	0.51	0.54
Condiments	18 (1 Tbsp), 0.3 (1 shake)	1.14 ± 1.24	1.36 ± 1.37	0.57 ± 0.73[4]	0.23	0.32	0.69

[1] Serving sizes differed from food to food even within one food group. For detailed serving size information for each food, *see* reference 17.
[2] Food intakes were log-transformed to achieve more normality in the distributions before the correlation coefficients were calculated.
[3] $\bar{x}$ ± SD.
[4] Significantly different from the average of the 2 FFQs, $P < 0.05$.

ance. Factor-loading matrixes for the 2 factors are listed in **Table 3**. A positive loading indicates a positive association with the factor, whereas a negative loading indicates an inverse association with the factor. The larger the loading of a given food item or group to the factor, the greater the contribution of that food item or group to a specific factor. The 2 major patterns identified from the 3 sources of dietary data were similar. The first factor was loaded heavily with the following foods or food groups: vegetable groups, legumes, whole grains, fruit, oil and vinegar salad dressings, and fish and other seafood, and the second factor was loaded heavily with processed meat, red meat, butter, high-fat dairy products, refined grains, eggs, and French fries. Following the method of Slattery et al (13), we labeled the first factor as the prudent pattern and the second factor as the Western pattern. The factor analyses also identified other minor patterns. However, because they were inconsistent across the 3 sources of data and explained a small amount of variance, we did not include them in the subsequent correlation analyses.

TABLE 3
Factor-loading matrix for the 2 major dietary patterns identified from diet records and 2 food-frequency questionnaires completed 1 y apart (FFQ1 and FFQ2) by 127 participants in the Health Professionals Follow-up Study (1986)

Foods or food groups	FFQ1		FFQ2		Diet records	
	Prudent pattern	Western pattern	Prudent pattern	Western pattern	Prudent pattern	Western pattern
Other vegetables	0.71	—[1]	0.64	—	0.32	—
Tomatoes	0.66	—	0.65	0.28	0.31	—
Green, leafy vegetables	0.64	—	0.69	0.17	0.58	—
Oil and vinegar salad dressings	0.61	—	0.57	—	0.36	—
Cruciferous vegetables	0.62	—	0.59	—	0.47	—
Legumes	0.55	—	0.47	—	0.50	—
Fish and other seafood	0.53	−0.22	0.52	—	0.16	—
Other soup	0.53	0.20	0.19	—	0.43	—
Dark-yellow vegetables	0.49	—	0.48	—	0.31	−0.36
Whole grains	0.49	—	0.49	−0.27	0.46	—
Condiments	0.41	0.23	0.21	0.53	0.35	0.31
Fruit	0.37	−0.30	0.65	−0.23	0.49	−0.49
Garlic	0.35	—	0.34	—	0.19	—
Poultry	0.30	—	0.43	—	0.29	−0.29
Chowder or cream soup	0.20	0.26	—	—	—	—
Butter	−0.22	0.67	−0.30	0.58	—	0.70
Processed meat	—	0.65	—	0.73	—	0.66
Red meat	—	0.64	—	0.65	−0.27	0.54
Potatoes	—	0.55	—	0.49	0.40	0.27
High-fat dairy products	—	0.49	—	0.24	—	0.38
Eggs	—	0.49	−0.26	0.35	—	0.62
Refined grains	—	0.41	—	0.53	0.25	0.28
High-energy drinks	—	0.34	—	0.31	—	0.26
Sweets and desserts	—	0.31	−0.16	—	—	—
Mayonnaise	—	0.31	0.23	0.39	—	0.28
Tea	—	0.21	0.17	0.29	—	−0.17
Beer	—	0.20	—	0.24	0.17	—
Coffee	—	—	—	0.37	—	0.57
Snack	—	—	—	0.18	—	0.18
Liquor	—	—	—	—	−0.22	0.38
French fries	—	0.36	—	0.45	—	0.56
Low-energy drinks	—	−0.25	—	0.23	—	—
Wine	0.20	—	0.18	—	—	—
Pizza	—	—	—	0.32	—	0.25
Margarine	—	—	—	—	—	—
Organ meat	—	—	—	—	—	—
Nuts	—	—	—	—	—	—
Fruit juice	—	—	0.29	—	0.37	−0.27
Cereal	—	—	0.22	−0.36	0.20	−0.36
Low-fat dairy products	0.22	—	0.35	−0.26	0.36	−0.47

[1] Absolute values <0.15 were excluded from the table for simplicity.

The correlations between the 2 FFQs were 0.70 for the prudent pattern and 0.67 for the Western pattern (**Table 4**), indicating good reproducibility. The 2 factors identified from the FFQs were reasonably correlated with those from the diet records: correlation coefficients corrected for week-to-week variation in diet records ranged from 0.45 to 0.74. Correlations between FFQ2 and the diet records were higher than those between FFQ1 and the diet records. In addition, those patterns were reasonably correlated with nutrient intakes calculated from diet records (**Table 5**). In particular, the prudent pattern was positively correlated with intakes of fiber, magnesium, potassium, folate, vitamin B-6, and carotenes, and negatively correlated with intakes of total and saturated fat. In contrast, the Western pattern was positively correlated with intakes of total and saturated fat and negatively correlated with intakes of fiber, magnesium, potassium, folate, vitamin B-6, and carotenes. The correlations between the factor scores and plasma concentrations of biomarkers were in the expected direction. The analyses eliminating current smokers (n = 11) yielded nearly identical results (data not shown).

DISCUSSION

Recently, dietary pattern analysis has emerged as a possible approach to examining diet-disease relations. In contrast with the conventional approach, which focuses on a single nutrient or a few nutrients or foods, this approach considers overall eating patterns. Few data, however, are available on the validity of this approach. In this study, we assessed the reproducibility and validity of 2 major dietary patterns (the prudent and Western patterns) derived from data collected with comprehensive, semiquantitative FFQs. Using factor analysis, we derived 2 major patterns, which were qualitatively similar across 3 sources of

TABLE 4
Pearson correlation coefficients for the prudent and Western dietary pattern scores between the 2 food-frequency questionnaires completed 1 y apart (FFQ1 and FFQ2) and the diet records

	Prudent pattern		Western pattern	
	Crude	Corrected[1]	Crude	Corrected[1]
FFQ1 vs FFQ2	0.70	—	0.67	—
FFQ1 vs diet records	0.34	0.45	0.51	0.58
FFQ2 vs diet records	0.41	0.52	0.64	0.74

[1] Corrected for week-to-week variation in diet records.

dietary data, ie, the 2 FFQs and the diet records. The first factor, the prudent dietary pattern, was characterized by a high intake of vegetables, legumes, whole grains, fruit, oil and vinegar salad dressings, and fish and other seafood. In contrast, the second factor, the Western dietary pattern, was characterized by a high intake of processed meat, red meat, butter, high-fat dairy products, refined grains, eggs, and French fries. The correlations (corrected for week-to-week variation in diet records) between each of the FFQs and the diet records ranged from 0.45 to 0.74 for the 2 patterns, suggesting reasonable comparability between the FFQs and the diet records in characterizing dietary patterns. In addition, the correlations between the factor scores and plasma concentrations of biomarkers were in the expected direction and were comparable with those between intakes of specific nutrients and plasma concentrations of these nutrients (25).

For both patterns, there were some differences in the factor loadings for the food items between the FFQs and diet records, probably because of methodologic differences between the dietary assessment methods (24) and random statistical variations. However, the major patterns generated from the FFQ and diet records were similar, and the correlations of the dietary patterns between each FFQ and the diet records ranged from 0.45 to 0.74, suggesting the usefulness of an FFQ in assessing dietary patterns relative to diet records. Correlations between dietary patterns derived from the diet records and blood measurements were generally higher than those between dietary patterns derived from the FFQ and the blood measurements. This may be in part because dietary intakes assessed by diet records are more accurate. The temporal relation may also have played a part because the FFQ asked about diet over the past year, whereas the biochemical measures reflect relatively shorter-term intakes (24).

The dietary patterns derived from our data were qualitatively similar to those from previous studies using the factor analytic approach. Using dietary data collected by a diet-history questionnaire, Slattery et al (13) grouped >800 food items from the diet-history questionnaire into 35 separate food groups. Two major eating patterns were identified: the Western pattern and the prudent pattern. They found that the prudent pattern was associated with a lower risk of colon cancer, whereas the Western pattern was associated with a higher risk of colon cancer. They also identified several minor patterns (high-fat, high-sugar, high-dairy products; drinkers; and substitutors), which were not significantly

TABLE 5
Pearson correlation coefficients between the prudent and Western dietary pattern scores and nutrients from two 1-wk diet records and plasma biochemical measurements

	Prudent pattern			Western pattern		
	FFQ1	FFQ2	Diet record	FFQ1	FFQ2	Diet record
Energy-adjusted nutrients from diet records[1]						
Total fat	−0.32	−0.41	−0.33	0.28	0.32	0.60
Saturated fat	−0.37	−0.45	−0.39	0.41	0.43	0.65
Fiber	0.30	0.41	0.52	−0.37	−0.41	−0.53
Calcium	0.03	0.14	0.16	−0.03	−0.16	−0.27
Magnesium	0.23	0.40	0.36	−0.31	−0.26	−0.57
Potassium	0.32	0.47	0.49	−0.29	−0.30	−0.51
Folate	0.12	0.28	0.38	−0.16	−0.09	−0.49
Vitamin B-6	0.22	0.39	0.36	−0.37	−0.36	−0.47
Retinol	0.01	−0.08	0.06	0.03	0.08	0
Carotene	0.17	0.18	0.35	−0.24	−0.21	−0.39
α-Carotene	0.07	0.11	0.31	−0.25	−0.22	−0.34
β-Carotene	0.18	0.23	0.41	−0.26	−0.24	−0.40
Lycopene	0.05	0.17	0.14	−0.16	−0.06	−0.24
Lutein and zeaxanthin	0.29	0.12	0.38	−0.10	−0.07	−0.02
β-Cryptoxanthin	0.03	0.26	0.33	−0.17	−0.16	−0.35
α-Tocopherol	0.06	0.08	0.13	−0.24	−0.18	−0.09
Biochemical measurements						
Total cholesterol	−0.08	−0.12	−0.25	0.18	0.12	0.18
Triacylglycerols	−0.03	−0.03	−0.17	0.23	0.24	0.10
Retinol	0.09	0.11	0.11	0.02	0.05	−0.05
Total carotene	0.27	0.40	0.31	−0.34	−0.26	−0.40
α-Carotene	0.29	0.39	0.24	−0.36	−0.28	−0.46
β-Carotene	0.23	0.37	0.32	−0.33	−0.28	−0.37
Lycopene	0.28	0.31	0.25	−0.12	0.03	−0.04
Lutein	0.33	0.33	0.38	−0.17	−0.15	−0.17
Zeaxanthin	0.08	0.11	0.27	−0.04	−0.08	−0.01
α-Tocopherol	0.12	0.15	0.12	−0.15	−0.14	−0.18

[1] All nutrients were calculated from food intakes derived from diet records. Supplementary intakes were not included.

associated with the risk of colon cancer. In a study of 939 Swiss adults, Gex-Fabry et al (26) found 2 similar major patterns: one was associated with a high intake of pork meat and sausages, pasta, and potatoes (satiating-capacity pattern), whereas the other was associated with a high intake of fresh fruit and vegetables, fish and other seafood, and poultry (healthfulness pattern).

One limitation of our study was that it included men only. Eating patterns may differ for women, even though a prior study suggested that major eating patterns apply to both men and women (13). Also, eating patterns are likely to vary with different socioeconomic statuses, ethnic groups, and cultures. Thus, it is necessary to replicate our study in other populations. In addition, because of changes in food preferences and food availability, the meaning of a dietary pattern could change over time. Finally, the 2 major patterns derived from our data explained only 20% of total variance, suggesting the existence of other eating patterns. However, in our study, dietary patterns other than the Western and prudent patterns were highly variable across the different dietary assessment methods, and may not be reproducible across studies. In a previous study (13), the Western and prudent dietary patterns explained 19% of the variance in men and 15% of the variance in women. However, these values should be interpreted with caution because they depend heavily on the total number of variables used in the factor analysis.

Because there are many potential differences in nutrient contents between dietary patterns, dietary pattern analysis cannot be specific about the particular nutrients responsible for the observed differences in disease risk, and thus may not be useful for assessing the biological relations between dietary components and disease risk. In particular, this approach would not be optimal if the effect was due to a specific nutrient (eg, neural tube defects resulting from a folic acid deficiency) because the effect of the nutrient would be diluted. Therefore, the dietary pattern approach may be more useful when traditional nutrient analyses have identified few dietary associations for the disease (eg, breast cancer). On the other hand, when many dietary associations have been shown for the disease (eg, coronary artery disease), dietary pattern analysis may also be useful because it examines not only nutrients and foods but the effects of overall diet as well. In addition, a dietary pattern can be used as a covariate when examining a specific nutrient to know whether the effect of the nutrient is independent of the overall dietary pattern.

In conclusion, our data indicate reasonable reproducibility and validity of the major dietary patterns defined by factor analysis using data from the FFQs. These findings suggest the potential use of the dietary pattern approach for studying diet-disease relations.

We thank the participants of the Health Professionals Follow-up Study and Matthew Gillman and Martha Slattery for helpful comments.

REFERENCES

1. National Research Council, Committee on Diet and Health. Diet and health: implications for reducing chronic disease risk. Washington, DC: National Academy Press, 1989.
2. Lee CN, Reed DM, MacLean CJ, Yano K, Chiu D. Dietary potassium and stroke. N Engl J Med 1988;318:995 (letter).
3. Kant AK, Schatzkin A, Block G, Ziegler RG, Nestle M. Food group intake patterns and associated nutrient profiles of the US population. J Am Diet Assoc 1991;91:1532–7.
4. Randall E, Marshall JR, Graham S, Brasure J. Patterns in food use and their associations with nutrient intakes. Am J Clin Nutr 1990; 52:739–45.
5. Ursin G, Ziegler RG, Subar AF, Graubard BI, Haile RW, Hoover R. Dietary patterns associated with a low-fat diet in the National Health Examination Follow-up Study: identification of potential confounders for epidemiologic analyses. Am J Epidemiol 1993; 137:916–27.
6. Jacobson HN, Stanton JL. Pattern analysis in nutrition. Clin Nutr 1986;5:249–53.
7. Huijbregts PPCW, Fesjens EJM, Kromhout D. Dietary patterns and cardiovascular risk factors in elderly men: the Zutphen Elderly Study. Int J Epidemiol 1995;24:313–20.
8. Huijbregts P, Feskens E, Rasanen L, et al. Dietary pattern and 20 year mortality in elderly men in Finland, Italy, and the Netherlands: longitudinal cohort study. BMJ 1997;315:13–7.
9. Kant AK. Indexes of overall diet quality: a review. J Am Diet Assoc 1996;96:785–91.
10. Millen BE, Quatromoni PA, Gagnon DR, Cupples LA, Franz MM, D'Agostino RB. Dietary patterns of men and women suggest targets for health promotion: the Framingham Nutrition Studies. Am J Health Promot 1996;11:42–53.
11. Nicklas TA, Webber LS, Thompson B, Berenson GS. A multivariate model for assessing eating patterns and their relationship to cardiovascular risk factors: the Bogalusa Heart Study. Am J Clin Nutr 1989;49:1320–7.
12. Herbert JR, Kabat GC. Implications for cancer epidemiology of differences in dietary intake associated with alcohol consumption. Nutr Cancer 1991;15:107–19.
13. Slattery ML, Boucher KM, Caan BJ, Potter JD, Ma K-N. Eating patterns and risk of colon cancer. Am J Epidemiol 1998;148:4–16.
14. Sacks FM, Obarzanek E, Windhauser MM, for the DASH investigators. Rational and design of the Dietary Approaches to Stop Hypertension Trial (DASH): a multicenter controlled-feeding study of dietary patterns to lower blood pressure. Ann Epidemiol 1994;5:108–18.
15. Rimm EB, Giovannucci EL, Willett WC, et al. A prospective study of alcohol consumption and the risk of coronary disease in men. Lancet 1991;338:464–8.
16. Rimm EB, Giovannucci EL, Stampfer MJ, Colditz GA, Litin LB, Willett WC. Reproducibility and validity of an expanded self-administered semiquantitative food frequency questionnaire among male health professionals. Am J Epidemiol 1992;135:1114–26.
17. 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.
18. Hess D, Keller HE, Oberlin B, Bonfanti R, Schuep W. Simultaneous detection of retinol, tocopherols, carotenes and lycopene in plasma by means of high-performance liquid chromatography on reversed phase. Int J Vitam Nutr Res 1991;61:232–8.
19. Richmond W. Preparation and properties of a cholesterol oxidase from *Nocarida* sp. and its application to the enzymatic assay of total cholesterol in serum. Clin Chem 1973;19:1350–6.
20. Bucolo G, David H. Quantitative determination of serum triglycerides by the use of enzymes. Clin Chem 1973;19:476–82.
21. SAS Institute Inc. SAS/STAT user's guide, version 6. 4th ed. Vol 2. Cary, NC: SAS Institute Inc, 1989.
22. Kim J-O, Mueller CW. Factor analysis: statistical methods and practical issues. Thousand Oaks, CA: Sage Publications, Inc, 1978.
23. Rosner B, Willett WC. Interval estimates for correlation coefficients corrected for within-person variation: implications for study design and hypothesis testing. Am J Epidemiol 1988;127:377–86.
24. Willett WC. Nutritional epidemiology. 2nd ed. New York: Oxford University Press, 1998.
25. Ascherio A, Stampfer MJ, Colditz GA, Rimm EB, Litin L, Willett WC. Correlations of vitamin A and E intake with the plasma concentrations of carotenoids and tocopherols among American men and women. J Nutr 1992;122:1792–801.
26. Gex-Fabry M, Raymond L, Jeanneret O. Multivariate analysis of dietary patterns in 939 Swiss adults: sociodemographic parameters and alcohol consumption profiles. Int J Epidemiol 1988;17:548–55.