Dietary Patterns Associated with Male Lung Cancer Risk in the Netherlands Cohort Study

Helena F. Balder,[1,2] R. Alexandra Goldbohm,[1] and Piet A. van den Brandt[2]

[1]Department of Food and Chemical Risk Analysis, TNO Nutrition and Food Research, Zeist, and [2]Department of Epidemiology, Maastricht University, Maastricht, the Netherlands

Abstract

The objective of this article was to study the association between dietary patterns and lung cancer incidence in the Netherlands Cohort Study on Diet and Cancer. The baseline measurement of this prospective case cohort study that was completed by 58,279 men in 1986 included a self-administered questionnaire on dietary intake, smoking habits, and other covariates. Follow-up was established by computerized record linkage to cancer registries and a pathology register. After 9.3 years of follow-up, 1,426 confirmed cases of incident male lung cancer were detected. Five dietary patterns were identified by exploratory factor analysis in a randomly sampled subcohort (*n* = 2,190). The dietary pattern labeled ''salad vegetables'' was associated with decreased risk of lung cancer [rate ratios $(RR)_{Q5}$, 0.75; 95% confidence interval (CI), 0.55-1.01], after multivariate adjustment. This inverse association was most evident among current and former smokers. A dietary pattern labeled ''sweet foods'' was also inversely associated with lung cancer risk (RR_{Q5}, 0.62; 95% CI, 0.43-0.89). However, the higher intake of monosaccharides and disaccharides, fruits, and lower consumption of alcohol associated with this pattern could not account for its full protective effect. The ''pork, processed meat, and potatoes'' pattern was nonsignificantly associated with increased risk (RR_{Q5}, 1.44; 95% CI, 0.99-2.09), and this positive association was most evident among current smokers. The other dietary patterns characterized by brown/white bread substitution and by consumption of cooked vegetables were not associated with lung cancer risk. These results show how studying both single factors and dietary patterns gives more insight into the complex, and often seemingly inconsistent, associations between diet and cancer. (Cancer Epidemiol Biomarkers Prev 2005;14(2):483–90)

Introduction

Lung cancer is among the most frequently occurring cancers in developed countries, particularly in men. Smoking is by far the most important cause of lung cancer, but lung cancer risk may also be influenced by diet. The most consistent associations with lung cancer have been shown with the consumption of fruits and vegetables (inverse; refs. 1, 2) and total and saturated fat (positive; ref. 3).

Traditional analyses in nutritional epidemiology have handled nutritional data either as isolated items or as a small number of items simultaneously. Relatively few studies have addressed the broader eating patterns that reflect many dietary exposures together. These dietary patterns may explain disease occurrence better than individual dietary exposures, in particular, when not just one dietary risk factor is responsible. One method of identifying and examining broader dietary patterns that may be associated with disease is factor analysis. This approach has been previously used in studies of colon or colorectal cancer (4-8), stomach cancer (9, 10), breast cancer (11), endometrial cancer (12), renal cell cancer (13), and thyroid cancer (14).

The only study relating lung cancer to dietary patterns used cluster analysis to identify two groups of individuals with distinct dietary patterns among 254 lung cancer cases and 184 healthy controls (15). A dietary pattern labeled ''healthy'', characterized by high intake of fiber and carbohydrate, and low intake of protein and animal fat, had a protective effect against lung cancer, however, after adjustment for smoking, the association was no longer statistically significant [odds ratio, 0.93; 95% confidence intervals (CI), 0.59-1.44].

To our knowledge, factor analysis has not previously been applied to dietary associations with lung cancer. As part of the DIETSCAN project, a collaboration with large prospective cohort studies from Sweden, Finland, and Italy, we recently identified dietary patterns using factor analysis (16). This gave us the opportunity to prospectively study the association between detailed dietary patterns and lung cancer incidence in the Netherlands Cohort Study on Diet and Cancer. Because of the relatively small number of lung cancer cases among the mainly nonsmoking women, only the results for men are presented. In addition, these results were, where appropriate, compared with food-based and nutrients-based analyses.

With 9.3 years of follow-up and more than a thousand incident lung cancer cases, the analyses could also be done for separate strata of smoking status and tumor histology.

Materials and Methods

Cohort. The Netherlands Cohort Study on Diet and Cancer was approved by the institutional review boards of the TNO Nutrition and Food Research (Zeist) and Maastricht University (Maastricht, the Netherlands). The Netherlands Cohort Study on Diet and Cancer started in September 1986 when men and women (ages 55 to 69 years) from 204 municipalities were enrolled in the cohort using computerized population registries. The case-cohort study design has been reported in detail elsewhere (17). In total, 58,279 men completed a self-administered mailed questionnaire on habitual dietary intake, smoking, lifestyle characteristics, medical history, and other potential risk factors for cancer. The dietary part of this questionnaire consisted of a 150-item semiquantitative food frequency questionnaire on the usual intake of food and beverages in the year preceding the start of the study, which was validated against a 9-day diet record (18). Questionnaire data of all cases and subcohort members have been key-entered twice and blinded with respect to case/subcohort status to

Received 5/7/04; revised 7/29/04; accepted 9/9/04.

Grant support: This work was financially supported by a grant from the Dutch Cancer Society.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Requests for reprints: Helena F. Balder, Department of Food and Chemical Risk Analysis, TNO Nutrition and Food Research, P.O. Box 360, 3700 AJ Zeist, the Netherlands. Phone: 31-30-694-4794; Fax: 31-30-695-7952. E-mail: balder@voeding.tno.nl

Table 1. Factor loadings and mean daily consumption (grams per day) of food groups, within the highest and the lowest quintile of the dietary pattern scores identified among male subcohort members of the Netherlands Cohort Study

Food groups	Salad vegetables pattern			Cooked vegetables pattern		
	Load	Q1	Q5	Load	Q1	Q5
Legumes	0.11	36	46	**0.58**	**22**	**63**
Cabbages	0.13	36	44	**0.73**	**21**	**67**
Leaf vegetables, cooked	0.01	22	22	**0.70**	**9.5**	**41**
Leaf vegetables, raw	**0.40**	**5.2**	**16**	0.13	8.5	13
Allium vegetables	**0.51**	**14**	**53**	**0.37**	**20**	**47**
Carrots	0.20	9.3	17	**0.46**	**5.9**	**21**
Tomatoes	**0.45**	**9.1**	34	0.26	15	28
Mushrooms	**0.52**	**1.1**	6.9	0.08	3.0	4.2
Citrus fruit	0.21	35	67	0.18	37	64
Bananas	0.14	8.4	17	0.05	11	15
Apples and pears	−0.01	70	68	0.08	54	74
(Straw)berries	0.14	5.0	8.4	0.06	5.4	7.5
Potatoes and potato products	−0.18	167	128	0.31	116	192
Rice	**0.43**	**3.3**	**33**	−0.04	16	14
Pasta	**0.53**	**2.3**	**14**	−0.10	8.8	6.1
Dry cereals	0.15	2.2	5.4	−0.04	4.0	3.3
Bread, crackers (etc., white types)	−0.16	50	24	0.06	31	41
Bread, crackers (etc., brown/whole meal types)	−0.01	146	145	−0.03	151	144
Fermented whole milk and milk products	−0.01	17	15	−0.01	15	16
Fermented medium, low-fat and skimmed milk products	0.04	60	72	0.13	55	87
Nonfermented whole milk and milk products	−0.24	180	79	0.07	105	136
Nonfermented medium, low-fat and skimmed milk products	0.04	92	112	−0.08	125	92
Full cream cheese and cheese spreads	−0.01	22	24	0.02	21	23
Low-fat cheese and cheese spreads	−0.03	2.1	1.4	0.09	0.73	2.4
Beef and veal (fresh)	−0.09	49	42	0.30	33	56
Pork (fresh)	0.01	42	43	0.06	42	48
Liver (fresh)	0.15	1.1	3.6	0.13	1.7	3.6
Poultry (fresh)	0.30	6.8	21	0.01	13	14
Processed meat	0.02	21	23	0.02	22	24
Fish (shellfish and crustacean included)	0.29	6.4	21	0.14	11	18
Eggs	0.11	14	19	0.07	16	18
Butter	0.07	7.8	10	−0.01	8.6	9.6
Margarine	−0.30	41	20	0.25	19	37
Low-fat margarine	0.06	5.4	8.3	−0.12	13	5.8
Oil	**0.58**	**0.6**	**5.0**	0.02	2.3	2.6
Dressings and other similar sauces	0.07	3.3	4.9	−0.00	4.2	4.7
Peanuts and other nuts	0.29	4.9	16	−0.11	12	7.8
Savory snacks	0.32	0.4	3.3	−0.22	2.6	0.7
Cakes, sweet breads, cookies, and biscuits	−0.06	27	24	−0.04	27	24
Sweets, candies	0.05	4.9	5.9	−0.01	5.5	4.9
Added sugar	−0.07	33	25	−0.03	33	30
Sweet sandwich spread	−0.11	12	7.8	−0.02	11	8.9
Coffee	−0.02	574	559	0.04	587	589
Tea	−0.06	361	297	0.13	261	361
Fruit juices	0.10	15	30	0.07	19	27
Soft drinks and syrups	0.07	32	49	−0.04	50	40
Beer	0.13	48	127	0.02	95	110
Wine and fortified wine	**0.40**	**7.8**	**87**	−0.06	48	31
Spirits	0.16	14	34	0.09	19	30
% Variance explained	5.6			4.8		

(Continued on the following page)

avoid random and systematic coding errors. Mean daily nutrient intakes were calculated using the computerized Dutch food composition table. After baseline exposure measurement, a subcohort of 2,335 male subjects was randomly sampled from the large cohort. Following the case-cohort approach, this subcohort was followed for migration and vital status to calculate person time at risk. The entire cohort has been followed-up for incidence of cancer. The method of record linkage to obtain information on cancer incidence in the entire cohort has been described previously (19). In short, follow-up for incident cancer has been established by record linkage to the Netherlands Cancer Registry and the Netherlands Pathology Registry. After 9.3 years of follow-up, i.e., from September 1986 to December 1995, 1,583 male lung cancer cases were identified. Subjects with incomplete or inconsistent dietary information were excluded from the analyses according to criteria described elsewhere (18). After excluding subjects who reported prevalent cancer other than skin cancer, and subjects with incident *in situ* lung cancer other than carcinoma (sarcoma, lymphoma, unspecified morphology), or without at least a microscopically confirmed diagnosis, 1,426 male incident primary lung carcinoma cases (International Classification of Diseases for Oncology codes T162.2-T162.9) were available for analysis. From the subcohort, prevalent cancer cases other than nonmelanoma skin cancer were excluded as well, leaving 2,190 men for analysis.

Table 1. Factor loadings and mean daily consumption (grams per day) of food groups, within the highest and the lowest quintile of the dietary pattern scores identified among male subcohort members of the Netherlands Cohort Study (Cont'd)

Pork, processed meat, and potatoes pattern			Sweet foods pattern			Brown/white bread substitution pattern		
Load	Q1	Q5	Load	Q1	Q5	Load	Q1	Q5
0.08	36	44	0.04	41	42	−0.02	42	41
0.12	36	45	−0.05	42	40	0.04	38	42
−0.03	22	22	0.01	23	23	−0.01	23	22
0.00	10	11	0.28	7.3	14	0.14	8.1	13
0.01	31	33	−0.14	40	28	0.08	28	36
−0.06	15	11	0.26	8.4	17	0.23	9.1	17
0.02	22	21	0.00	21	21	0.14	16	25
−0.04	3.7	3.1	−0.11	4.4	3.0	0.06	2.8	3.6
−0.30	85	31	0.13	39	60	0.10	42	64
0.07	12	15	0.25	7.2	21	−0.04	14	10
−0.09	82	60	0.26	40	96	**0.37**	**40**	**115**
−0.05	7.6	6.3	**0.36**	**3.5**	**11**	−0.03	7.5	6.5
0.38	**111**	**199**	0.25	123	180	0.02	152	157
−0.18	24	9.2	0.09	9.4	17	0.02	13	12
0.20	4.8	10	0.06	6.3	7.9	−0.04	6.7	6.3
−0.28	7.7	1.3	0.19	1.6	5.5	0.13	2.1	5.3
0.04	37	41	0.13	29	48	**−0.62**	**108**	**11**
0.33	113	189	0.18	117	164	**0.67**	**54**	**221**
−0.06	21	14	0.28	3.6	35	−0.01	14	13
−0.21	102	39	0.02	63	68	0.33	31	139
0.15	95	151	0.29	66	195	−0.29	191	61
−0.03	108	99	−0.06	112	100	0.19	57	145
0.08	21	25	0.23	15	27	0.22	18	29
−0.02	2.1	1.7	−0.00	1.6	1.9	0.18	0.7	3.8
−0.11	50	37	−0.00	45	44	−0.07	43	41
0.48	**23**	**65**	−0.18	55	37	−0.17	52	36
0.10	1.8	3.4	−0.04	2.9	2.3	−0.10	3.2	1.4
−0.01	14	13	−0.04	15	13	−0.01	13	12
0.53	**11**	**39**	0.07	20	24	−0.08	25	21
−0.06	15	14	−0.11	17	12	−0.11	17	12
0.19	15	21	−0.04	19	17	−0.32	24	13
−0.30	18	3.3	0.14	6.5	12	−0.32	16	3.2
0.25	17	37	0.12	25	34	−0.01	29	29
0.46	**1.4**	**19**	0.08	5.7	8.9	0.28	4.1	17
0.06	2.3	2.6	0.05	2.3	2.5	0.01	2.2	2.4
0.22	2.7	6.3	0.21	2.9	6.1	−0.05	4.7	4.5
0.06	8.2	11	0.20	5.3	13	−0.13	12	7.0
0.06	1.1	1.6	0.13	0.7	2.2	−0.12	1.9	0.8
−0.01	26	25	**0.53**	**12**	**39**	0.02	25	25
−0.04	5.7	4.3	**0.38**	**1.7**	**9.6**	−0.13	6.1	3.7
0.30	20	45	0.25	19	42	−0.34	48	15
0.04	9.0	9.7	**0.45**	**3.2**	**19**	0.01	9.4	11
0.50	**397**	**796**	−0.13	636	525	−0.02	598	555
−0.27	424	246	0.31	199	404	0.01	320	333
−0.02	23	16	0.12	16	34	−0.07	24	17
0.17	25	64	0.17	28	61	−0.21	63	22
0.21	39	164	−0.34	218	35	−0.10	120	51
−0.09	46	23	−0.02	35	33	0.03	26	37
−0.03	26	23	−0.29	47	10	−0.19	34	13
4.3			4.3			4.1		

NOTE: Values in boldface font indicate food groups with factor loadings >0.35 in the respective dietary pattern.

Assessment of Dietary Patterns. As part of the DIETSCAN project, the food items from the food frequency questionnaire were aggregated into 51 food groups based on their role in the diet and possible relevance to cancer etiology (16). To identify dietary patterns, exploratory factor analysis using principal components analysis was conducted in the subcohort, using the food frequency questionnaire-derived food groups (PROC FACTOR in SAS version 8; SAS Institute, Inc., Cary, NC). Factors were rotated by orthogonal Varimax transformation. Eigenvalues >1 and the scree test, traditional criteria in factor analysis, were used to determine the number of factors to be extracted. Labels of the dietary patterns were determined according to the dominant foods (i.e., foods with high factor loadings, representing correlation coefficients between food groups and dietary patterns). Further details regarding the food grouping, dietary pattern assessment, and related sensitivity analyses conducted in DIETSCAN has been described elsewhere (16).

Statistical Analyses. Cox proportional hazards models were constructed to estimate hazard ratios and 95% CI relating the dietary patterns to the incidence of lung cancer (Stata version 8; Stata Corporation, College Station, TX). The proportional hazards assumption was tested using Schoenfeld residuals. Because of the case-cohort design, the 95% CIs were corrected for the additional variance introduced by using a randomly sampled subcohort instead of the complete cohort, by using the robust option.

Prior to etiologic analysis, factor scores for both lung cancer cases and members of the subcohort were determined by summing the standardized intakes from each food group, weighed by the factor loadings (PROC SCORE in SAS). Thus, each participant had a unique score for each factor. High scores represented high intake of foods loading on the corresponding dietary pattern, low scores represented low intake of those foods. Because the factor scores represent standardized variables, each score had a mean of zero and a SD of one. Rate ratio (RR) estimates for the dietary patterns were calculated for quintiles based on the distribution in the subcohort and for continuous variables with an increment of 1 SD. Tests for trend in the RRs were assessed by fitting ordinal exposure categories as continuous variables. The models were adjusted for potential confounding variables determined by previous analyses conducted within the Netherlands Cohort Study on Diet and Cancer (1) and additionally for health-related lifestyle factors that we expected to be associated with one or more dietary patterns: age at baseline (years), total energy intake (kilojoules), current cigarette smoking (yes/no), number of cigarettes smoked per day, number of years of smoking cigarettes, highest attained education (primary school, lower vocational school, junior high school, senior high school, higher vocational school, or university), physical activity outside of the profession (<30, 30-60, 60-90, or >90 minutes/day) and family history of lung cancer in first- or second-degree relatives (yes/no). The models included all factor scores simultaneously because dietary patterns are conditional on each other. Subgroup analyses were done after stratification into never smokers, former smokers, and current smokers, and into Kreyberg group I tumors (squamous cell, large cell, and small cell carcinoma) and Kreyberg group II tumors (adenocarcinoma). To compare the dietary pattern approach with traditional food and nutrient-based analyses, Cox regression models were constructed for food groups or nutrients that corresponded closely to several of the dietary patterns.

Results

Exploratory factor analysis identified five stable dietary patterns that explained 23% of the total variance in the dietary input variables (16). In Table 1, the factor loadings of all food groups and their mean daily consumption within the first and the fifth quintile of dietary pattern scores are presented. The first dietary pattern was one we labeled ''(salad) vegetables''. This pattern was characterized by high factor loadings on several vegetable items, several fruit items, pasta, rice, poultry, fish, and oil, and explained 5.6% of the total variance. A separate ''cooked vegetables'' pattern was also identified, with high loadings on cooked leaf vegetables, cabbages, legumes, and carrots, describing 4.8% of the total variance. A pattern labeled ''pork, processed meat, and potatoes'' was also identified. This pattern also loaded positively on coffee and low-fat margarine, and accounted for 4.3% of the variance in dietary intake. A pattern with high loads on ''sweet foods'', such as cakes and cookies, sweet sandwich spread, sweets and candies, and (straw)berries accounted for 4.3% of total variance. In addition a ''white/brown bread substitution'' pattern was identified that correlated positively with brown/wholemeal bread types and apples and pears; correlating negatively with white types of bread, it explained 4.1% of the total variance.

A description of the 1,426 male lung cancer cases and 2,190 male subcohort members with complete dietary data is presented in Table 2. On average, cases were older, were more likely to smoke cigarettes, and had a lower educational level than members of the subcohort.

The association between potential covariates and dietary patterns was studied in the subcohort by calculating Pearson correlation coefficients (Table 3). Men with a higher score on the salad vegetables pattern tended to be younger and more highly educated. Subjects with higher pork, processed meat, and potatoes pattern scores were younger and less educated, were more likely to smoke and tended to smoke more cigarettes per day and smoked for more years. Men with higher scores on the sweet pattern had lower body mass index and were less likely to be current smokers and tended to smoke fewer cigarettes per day and smoked for fewer years. Consumption of brown/wholemeal types of bread instead of white types of bread was negatively associated with current smoker status, number of cigarettes smoked per day, and duration of smoking. None of the dietary patterns were associated with family history of lung cancer.

RRs of lung cancer and their 95% CIs according to quintiles of scores on the dietary patterns are provided in Table 4. After adjustment for potential confounders and the other identified dietary patterns, the salad vegetables pattern was associated with decreased risk of lung cancer (RR_{Q5}, 0.75; 95% CI, 0.55-1.01; *P* for trend, 0.008). High consumption of the pork, processed meat, and potatoes pattern seemed to be associated with an increased risk of lung cancer. However, after adjustment for smoking, the other potential confounders and the other four dietary patterns, the positive association was less strong and was no longer statistically significant. After adjustment for all potential confounders, the sweet pattern was associated with decreased risk of lung cancer (RR_{Q5}, 0.62; 95% CI, 0.43-0.89; *P* for trend, 0.002). The pattern characterized by brown/white bread seemed to be associated with a decreased risk of lung cancer. However, after adjustment for smoking, the inverse association was less strong and no longer statistically significant. However, the hazard ratio of this pattern did not seem to be constant over time, and neither were the hazard ratios for education,

Table 2. Nondietary baseline characteristics among male lung carcinoma cases and subcohort members with complete dietary data

Characteristics	Subcohort		Cases	
	n	Mean or %	*n*	Mean or %
Age (y)	2,190	61.3	1,426	626
Family history of lung cancer	2,190	10%	1,426	12%
Body mass index (kg/m^2)	2,115	24.9	1,380	24.7
Total energy intake (MJ)	2,190	9.1	1,426	9.0
Smoking habits				
Never smoker	277	13%	52	3.7%
Former smoker	1,128	52%	488	34%
Current smoker	785	36%	886	62%
Number of cigarettes per day	2,065	14.8	1,282	19.1
Duration of smoking cigarettes (y)	2,156	29.4	1,398	39.6
Highest level of education				
Primary school	494	25%	388	32%
Lower vocational	414	21%	305	25%
High school	398	35%	392	32%
Higher vocational/university	356	18%	141	11%
Physical activity outside profession				
<30 min/d	398	18%	319	23%
30-60 min/d	669	31%	408	29%
60-90 min/d	406	19%	259	18%
>90 min/d	691	32%	423	30%

Table 3. Pearson correlation coefficients of potential covariates with dietary patterns among male subcohort member of the Netherlands Cohort Study

Characteristics	*n*	Salad vegetables pattern	Cooked vegetables pattern	Pork, processed meat, and potatoes pattern	Sweet foods pattern	Brown/white bread substitution pattern
Age (y)	2,190	−0.12	0.06	−0.20	−0.00	0.02
Family history of lung cancer	2,190	0.03	0.02	−0.01	−0.02	−0.00
Body mass index (kg/m^2)	2,115	0.02	0.08	0.07	−0.10	−0.03
Total energy intake (MJ)	2,190	0.13	0.21	0.50	0.50	−0.15
Current smoking status	2,190	−0.03	0.02	0.19	−0.14	−0.21
Number of cigarettes per day	2,065	0.07	0.05	0.12	−0.20	−0.12
Years of smoking cigarettes	2,156	−0.07	0.01	0.16	−0.19	−0.14
Higher vocational/university education	1,973	0.21	−0.09	−0.14	0.07	0.02
Physical activity (min/d)	2,164	−0.01	0.08	0.06	0.05	0.03

family history of lung cancer, and physical activity. When this was adjusted for in the model (using time-dependent variables), the hazard ratio for brown/white bread substitution pattern increased from 0.87 after 1.3 years of follow-up to 1.25 after 9.3 years of follow-up. The pattern characterized by consumption of cooked vegetables was not associated with lung cancer risk.

In Table 5, the relative rates for dietary patterns stratified according to smoking status are presented. The salad vegetables pattern seemed to be associated with decreased risk of lung

Table 4. RR and 95% CI of lung cancer according to dietary patterns among men in the Netherlands Cohort Study (9.3 years of follow-up)

Dietary pattern	Quintile of pattern score					$P^{\dagger}$	Continuous, linear (increment, 1 SD)
	Q1*	Q2	Q3	Q4	Q5		
Salad vegetables							
Cases of lung cancer	329	336	300	245	216		1,426
Person-years	3,705	3,736	3,768	3,739	3,817		18,766
RR (age, kJ)	1	1.06	0.95	0.80	0.72	0.000	0.86
95% CI	ref.	0.86-1.31	0.77-1.17	0.65-1.00	0.57-0.90		0.80-0.93
RR (multivariate)‡	1	1.07	1.02	0.75	0.75	0.008	0.87
95% CI	ref.	0.81-1.40	0.77-1.35	0.56-1.01	0.55-1.01		0.78-0.96
Cooked vegetables							
Cases of lung cancer	251	296	327	280	272		1,426
Person-years	3,736	3,763	3,767	3,810	3,688		18,766
RR (age, kJ)	1	1.12	1.28	1.05	1.02	0.88	1.00
95% CI	ref.	0.90-1.40	1.03-1.59	0.84-1.31	0.81-1.28		0.93-1.07
RR (multivariate)‡	1	1.05	1.10	0.90	0.86	0.18	0.95
95% CI	ref.	0.78-1.41	0.83-1.47	0.67-1.22	0.63-1.16		0.86-1.05
Pork, processed meat, and potatoes							
Cases of lung cancer	192	269	316	303	346		1,426
Person-years	3,690	3,736	3,759	3,779	3,802		18,766
RR (age, kJ)	1	1.52	1.90	1.94	2.67	0.000	1.38
95% CI	ref.	1.20-1.92	1.50-2.40	1.53-2.47	2.06-3.47		1.27-1.50
RR (multivariate)‡	1	1.18	1.32	1.24	1.44	0.08	1.10
95% CI	ref.	0.87-1.61	0.96-1.80	0.90-1.71	0.99-2.09		0.98-1.25
Sweet foods							
Cases of lung cancer	437	326	268	209	186		1,426
Person-years	3,683	3,707	3,706	3,786	3,883		18,766
RR (age, kJ)	1	0.68	0.54	0.37	0.31	0.000	0.64
95% CI	ref.	0.55-0.83	0.43-0.66	0.30-0.47	0.24-0.40		0.59-0.70
RR (multivariate)‡	1	0.73	0.70	0.57	0.62	0.002	0.80
95% CI	ref.	0.56-0.95	0.52-0.93	0.41-0.78	0.43-0.89		0.70-0.90
Brown/white bread substitution							
Cases of lung cancer	340	332	265	282	207		1,426
Person-years	3,760	3,681	3,786	3,770	3,768		18,766
RR (age, kJ)	1	0.95	0.76	0.76	0.59	0.000	0.85
95% CI	ref.	0.77-1.17	0.61-0.94	0.61-0.95	0.47-0.74		0.80-0.91
RR (multivariate)‡,§	1	1.14	0.92	0.90	0.89	0.18	0.99
95% CI	ref.	0.87-1.50	0.69-1.23	0.67-1.22	0.65-1.20		0.89-1.09

*Q, quintile.
†All *P* values are from two-sided tests.
‡Adjusted for all other dietary patterns and age at baseline, total energy intake (kJ), current cigarette smoker (yes/no), number of cigarettes smoked per day, years of smoking cigarettes, higher vocational or university education, family history of lung cancer, physical activity (<30, 30-60, 60-90, or >90 min/d).
§The hazard ratio of the brown/white bread substitution pattern and education, history of lung cancer in family and physical activity did not seem to be constant over time. When this was adjusted for in the model (using time-dependent variables), the hazard ratio was $\exp(-0.2027 + 0.0454 \times t)$.

Table 5. RR and 95% CI of lung cancer with an increment in 1 SD in factor score for dietary patterns, stratified by smoking status and histologic type of cancer

Subgroup	Number of cases	Person-time	Salad vegetables pattern	Cooked vegetables pattern	Pork, processed meat, and potatoes pattern	Sweet foods pattern	Brown/white bread substitution pattern
Total*	1,087	15,536	0.89 (0.81-0.97)	0.97 (0.88-1.06)	1.17 (1.05-1.30)	0.81 (0.72-0.90)	1.00 (0.91-1.09)
Smoking status							
Never smoker†	52	2,452	1.21 (0.94-1.56)	0.84 (0.63-1.12)	1.20 (0.85-1.67)	0.51 (0.33-0.77)	0.83 (0.59-1.17)
Former smoker‡	446	9,233	0.77 (0.65-0.90)	1.08 (0.95-1.24)	1.04 (0.88-1.22)	0.88 (0.75-1.04)	0.96 (0.84-1.10)
Current smoker‡	771	5,852	0.90 (0.80-1.01)	0.97 (0.87-1.08)	1.21 (1.06-1.39)	0.79 (0.70-0.91)	1.04 (0.93-1.16)
Histologic type of cancer							
Kreyberg I§	915	17,454	0.85 (0.77-0.94)	0.98 (0.89-1.07)	1.25 (1.12-1.39)	0.77 (0.69-0.86)	0.98 (0.89-1.07)
Kreyberg II§	254	17,381	0.95 (0.83-1.10)	0.97 (0.86-1.10)	0.98 (0.83-1.15)	0.83 (0.71-0.96)	1.00 (0.88-1.15)

*Adjusted for age at baseline, total energy intake (kJ), current cigarette smoker (yes/no), number of cigarettes smoked per day, years of smoking cigarettes, higher vocational or university education, family history of lung cancer, physical activity (not active, moderately active, active, or very active).
†Adjusted for age and energy.
‡Adjusted for age, energy, number of cigarettes smoked, and years of smoking.
§Adjusted for age, energy, current smoking, number of cigarettes, and years of smoking.

cancer in former smokers ($RR_{per\ 1\ SD}$, 0.77; 95% CI, 0.65-0.90; P for interaction, 0.02), and marginally significant in current smokers ($RR_{per\ 1\ SD}$, 0.90; 95% CI, 0.80-1.01; P for interaction, 0.22). The increased risk of lung cancer of the dietary pattern characterized by consumption of pork, processed meat, and potatoes pattern was associated with risk of lung cancer among current smokers ($RR_{per\ 1\ SD}$, 1.21; 95% CI, 1.06-1.39; P for interaction, 0.36). In former smokers, the pork, processed meat, and potatoes pattern was not associated with risk of lung cancer (P for interaction, 0.66), whereas in never smokers, the association was not statistically significant ($RR_{per\ 1\ SD}$, 1.20; 95% CI, 0.85-1.67). The sweet pattern was most strongly associated with decreased risk of lung cancer in never smokers ($RR_{per\ 1\ SD}$, 0.51; 95% CI, 0.33-0.77) and current smokers ($RR_{per\ 1\ SD}$, 0.79; 95% CI, 0.70-0.91; P for interaction, 0.05). In former smokers, the inverse association was not statistically significant (P for interaction, 0.01). The cooked vegetables pattern and the brown/white bread substitution pattern were not associated with risk of lung cancer, in any of the smoking strata (P for interaction ranging from 0.21 to 0.63).

As different histologic groups of carcinomas are suspected to have different etiologies, RRs of dietary patterns are shown for Kreyberg I and II carcinomas separately (Table 5). The salad vegetables pattern was associated with decreased risk of Kreyberg I carcinomas, but not with the less frequent Kreyberg II carcinomas. Similarly, the pork, processed meat, and potatoes pattern increased risk of Kreyberg I carcinomas, but not with Kreyberg II carcinomas. The sweet pattern decreased the risk of both Kreyberg I and II carcinomas. The cooked vegetables or the brown/white bread substitution patterns were not associated with either type of carcinomas.

To compare the salad vegetables pattern and cooked vegetables pattern with traditional food-based analyses, we also calculated the risk estimates for the groups of total vegetables, raw vegetables, and cooked vegetables (Table 6). High consumption of total vegetables was associated with decreased risk of lung cancer (RR_{Q5}, 0.66; 95% CI, 0.50-0.87; test for trend, 0.008). Similarly, raw vegetables (RR_{Q5}, 0.74; 95% CI, 0.57-0.98; test for trend, 0.07) and cooked vegetables (RR_{Q5}, 0.73; 95% CI, 0.55-0.96; test for trend, 0.07) were associated with decreased risk of lung cancer. The risk reduction for the salad vegetables dietary pattern did not seem to be greater than that found when studying the groups of total or raw vegetables. Therefore, we repeated the multivariate analyses, additionally adjusting for either total vegetables or raw vegetables. After adjustment for total vegetables consumption, there was still a suggestion of a negative association for the salad vegetables pattern; however, the point estimate and the test for trend no longer reached statistical significance (RR_{Q5}, 0.91; 95% CI, 0.82-1.01; test for trend, 0.09). Adjustment for raw vegetables yielded similar results.

To compare the sweet foods pattern with the traditional reductionist approach, we also calculated the risk estimates for total fruits consumption and for intake of monosaccharides and disaccharides and alcohol (Pearson correlation coefficients with the sweet pattern of 0.37, 0.58, and −0.40, respectively). High consumption of monosaccharides and disaccharides was not associated with decreased risk of lung cancer (RR_{Q5}, 0.88; 95% CI, 0.63-1.23; test for trend, 0.53). Consumption of alcohol (RR_{Q5}, 1.56; 95% CI, 1.11-2.18; test for trend, 0.03) was associated with increased risk of lung cancer and total fruit consumption (RR_{Q5} 0.69; 95% CI, 0.53-0.91; test for trend, 0.001) was associated with decreased risk. After additional adjustment for total fruits consumption and for total alcohol consumption, the sweet pattern was still inversely associated with risk of lung cancer (respectively, RR_{Q5}, 0.83; 95% CI, 0.74-0.93; test for trend, 0.004; and RR_{Q5}, 0.64; 95% CI, 0.44-0.93; test for trend, 0.006).

Discussion

Five dietary patterns were identified in male participants of the Netherlands Cohort Study. The salad vegetables pattern and the ''sweet'' pattern were inversely associated with lung cancer risk after adjustment for confounders, whereas the pork, processed meat, and potatoes pattern was associated with increased risk of lung cancer. The other dietary patterns were not associated with lung cancer risk.

One of the strengths of our study is the completeness of follow-up of both person-years and cancer cases, ensuring that selection bias due to loss of follow-up is unlikely (20). Another strength is the prospective design in which diet was measured before disease so that information bias due to change in (recall of) diet due to disease was avoided. As preclinical symptoms might have influenced diet, the analyses were repeated excluding cases diagnosed during the first 2 years of follow-up (n = 250). However, this did not affect the results. Even though we used a validated semiquantitative food frequency questionnaire to assess the participants' diet, we cannot exclude the possibility that the dietary pattern approach combined artificial correlations due to grouping of items within the food frequency questionnaire. However, spurious factors due to correlated questionnaire items are unlikely as Hu et al. (21) and Togo et al. (22) showed that qualitative similar eating patterns were observed across food frequency questionnaires and diet records. Furthermore, as the identified dietary patterns were very comparable with dietary patterns identified in other populations (16), and as the patterns correlated as expected with health-related lifestyle variables, it is unlikely that these dietary patterns are merely caused by correlated measurement error.

Another potential limitation that should be considered is whether the food frequency questionnaire at baseline was a reliable estimate of past and future diet. The stability of dietary habits over time was evaluated from five annually repeated questionnaire administrations in independent random samples of the cohort. The mean intakes barely changed and the correlation between two measurements decreased only slightly over an increasing time interval. It was concluded that the single food frequency questionnaire measurement characterizes dietary habits for a period of at least 5 to 10 years (23).

Despite the strengths of factor analysis, the technique has been criticized for its subjective decisions (24). We did extensive sensitivity analyses and our findings showed that the number of extracted factors and other subjective factor analytic decisions did not affect the identified dietary patterns (16). However, as McCann et al. (12) observed that reducing the number of input variables from 168 to 56 to 36 did not affect the derived factors, but did attenuate the odds ratios of endometrial cancer for the healthy pattern, the classification of the food frequency questionnaire items into 51 food groups may have attenuated the RRs.

The inverse association we observed between the salad vegetables pattern and lung cancer risk is in agreement with both case-control and prospective cohort studies, that have consistently shown an inverse association between fruit and vegetable intake and lung cancer risk (2, 25-27). In both previous analyses of the Netherlands Cohort Study (1) and the current reanalyses after 9.3 years of follow-up, we found reduced risks of lung cancer for vegetable and fruit consumption. The risk reduction for the salad vegetables dietary pattern seemed to be not greater than that found when studying total or raw vegetable consumption, and additional adjustment for these food groups diminished the association with the pattern. Therefore, although other foods such as oil and wholemeal products also present in the salad vegetables dietary pattern may act synergistically with vegetables to decrease risk of lung cancer, our results favor a major role for (raw) vegetables. The cooked vegetables pattern was not significantly associated with lung cancer risk, whereas the sum of total cooked vegetables was. Therefore, the other foods associated with the cooked vegetables pattern (such as spirits, beef and veal, margarine, potatoes and potato products, etc.) may counteract the protective effect of the cooked vegetables.

The inverse association between the sweet pattern and lung cancer risk was unexpected and was not in agreement with a case-control study that showed a positive dose-response relation between consumption of particular desserts (cakes and custard/cream pies) and the risk of lung cancer in men and women (28). Although adjustment for smoking behavior changed the RR estimate, it is unlikely that the

Table 6. RR and 95% CI of lung cancer according to quintiles of food groups and nutrients that corresponded closely to several of the dietary patterns (Netherlands Cohort Study, 9.3 years of follow-up)

Food group or nutrient	Quintile of consumption					$P^{\dagger}$
	Q1*	Q2	Q3	Q4	Q5	
Salad vegetables and Cooked vegetables patterns						
Total vegetables						
Median intake (g/d)	102	144	177	218	294	
Cases of lung cancer	339	293	304	254	236	
Person-years	3,684	3,760	3,786	3,812	3,724	
RR (multivariate)‡	1	0.97	0.99	0.92	0.66	
95% CI	ref.	0.74-1.26	0.76-1.30	0.70-1.21	0.50-0.87	0.008
Raw vegetables						
Median intake (g/d)	6	19	31	46	73	
Cases of lung cancer	374	279	290	260	223	
Person-years	3,751	3,706	3,727	3,824	3,758	
RR (multivariate)‡	1	0.82	0.97	0.84	0.74	
95% CI	ref.	0.63-1.07	0.75-1.26	0.64-1.10	0.57-0.98	0.071
Cooked vegetables						
Median intake (g/d)	79	115	143	177	239	
Cases of lung cancer	322	288	280	282	254	
Person-years	3,745	3,700	3,790	3,793	3,737	
RR (multivariate)‡	1	0.89	0.91	0.94	0.73	
95% CI	ref.	0.68-1.16	0.69-1.20	0.72-1.23	0.55-0.96	0.074
Sweet foods pattern						
Total fruits						
Median intake (g/d)	31	88	135	186	290	
Cases of lung cancer	408	305	235	235	243	
Person-years	3,717	3,746	3,737	3,799	3,766	
RR (multivariate)‡	1	0.77	0.67	0.63	0.69	
95% CI	ref.	0.59-0.99	0.52-0.87	0.48-0.83	0.53-0.91	0.001
Alcohol						
Median intake (g/d)	0	2.2	9.3	23	42	
Cases of lung cancer	183	241	337	333	311	
Person-years	2,592	3,880	5,056	4,265	2,973	
RR (multivariate)‡	1	1.11	1.23	1.08	1.56	
95% CI	ref.	0.80-1.54	0.91-1.67	0.80-1.47	1.11-2.18	0.03
Monosaccharides/disaccharides						
Median intake (g/d)	53	78	99	122	163	
Cases of lung cancer	306	302	238	291	289	
Person-years	3,669	3,808	3,736	3,731	3,822	
RR (multivariate)‡	1	0.95	0.80	0.96	0.88	
95% CI	ref.	0.72-1.25	0.59-1.06	0.71-1.30	0.63-1.23	0.53

*Q, quintile.

†All P values are from two-sided tests.

‡Adjusted for age at baseline, total energy intake (kJ), current cigarette smoker (yes/no), number of cigarettes smoked per day, years of smoking cigarettes, higher vocational or university education, family history of lung cancer, physical activity (not active, moderate active, active, or very active), and body mass index (kg/m^2).

observed association is entirely due to residual confounding because the association still existed after stratification according to smoking status and seemed to be strongest in never smokers. As additional adjustment for monosaccharides and disaccharides, total fruit consumption and total alcohol consumption diminished the inverse association only slightly, the higher consumption of monosaccharides and disaccharides and fruits, and the lower consumption of alcohol associated with this dietary pattern, could not account for its full protective effect. Therefore, either the combination of effects of all food groups adds up to the protective effect of this dietary pattern, or other factors associated with this pattern affect lung cancer risk.

The pork, processed meat, and potatoes pattern was associated with increased risk of lung cancer in men, and this association was most evident among current smokers. So far, the evidence that higher meat and fat intake are associated with higher risk of lung cancer is weakly consistent (3, 26). All constituents in the pork, processed meat, and potatoes dietary pattern might act synergistically to increase risk of lung cancer. However, as adjustment for smoking behavior diminished the association, and the association was strongest in current smokers, we cannot exclude the possibility of residual confounding due to smoking.

Dietary patterns represent combinations of foods describing as much variance in the original dietary variables as possible, but not necessarily those combinations of foods that show the strongest association with disease. The dietary pattern approach complements the traditional reductionist approach, by enabling the study of not only the effect of single factors, but also of complex combinations of dietary exposures.

In conclusion, although we cannot entirely exclude the possibility of residual confounding, the dietary patterns which we labeled salad vegetables and sweet foods were associated with a decreased risk of lung cancer, whereas the pork, processed meat, and potatoes pattern might be associated with an increased risk. The protective effect of cooked vegetables seemed to be counteracted by other foods or factors associated with the cooked vegetables pattern. These results show how studying both single factors and dietary patterns gives more insight into the complex, and often seemingly inconsistent, associations between diet and cancer.

Acknowledgments

We thank the participants of this study; the cancer registries (IKA, IKL, IKMN, IKN, IKO, IKR, IKST, IKW, IKZ, and VIKC); the Netherlands nationwide registry of pathology (PALGA); Drs. A. Volovics and A. Kester for statistical advice; L. Schouten, S. van de Crommert, H. Brants, J. Nelissen, C. de Zwart, M. Moll, W. van Dijk, M. Jansen, and A. Pisters for assistance; and H. van Montfort, T. van Moergastel, L. van den Bosch, and R. Schmeitz for programming assistance.

References

1. Voorrips LE, Goldbohm RA, Verhoeven DTH, et al. Vegetable and fruit consumption and lung cancer risk in the Netherlands Cohort Study on diet and cancer. Cancer Causes Control 2000;11:101–5.
2. Smith-Warner SA, Spiegelman D, Yaun SS, et al. Fruits and vegetables and lung cancer: a pooled analysis of cohort studies. Int J Cancer 2003; 107:1001–11.
3. Smith-Warner SA, Ritz J, Hunter DJ, et al. Dietary fat and risk of lung cancer in a pooled analysis of prospective studies. Cancer Epidemiol Biomarkers Prev 2002;11:987–92.
4. Randall E, Marshall JR, Brasure J, Graham S. Dietary patterns and colon cancer in western New York. Nutr Cancer 1992;18:265–76.
5. Slattery ML, Boucher KM, Caan BJ, Potter JD, Ma KN. Eating patterns and risk of colon cancer. Am J Epidemiol 1998;148:4–16.
6. Terry P, Hu FB, Hansen H, Wolk A. Prospective study of major dietary patterns and colorectal cancer risk in women. Am J Epidemiol 2001;154: 1143–9.
7. Fung T, Hu FB, Fuchs C, et al. Major dietary patterns and the risk of colorectal cancer in women. Arch Intern Med 2003;163:309–14.
8. Dixon LB, Balder HF, Virtanen M, et al. Dietary patterns associated with colon and rectal cancers: results from the DIETSCAN project. Am J Clin Nutr 2004;80:1003–11.
9. Palli D, Russo A, Decarli A. Dietary patterns, nutrient intake and gastric cancer in a high-risk area of Italy. Cancer Causes Control 2001;12:163–72.
10. Masaki M, Sugimori H, Nakamura K, Tadera M. Dietary patterns and stomach cancer among middle-aged male workers in Tokyo. Asian Pac J Cancer Prev 2003;4:61–6.
11. Terry P, Suzuki R, Hu FB, Wolk A. A prospective study of major dietary patterns and the risk of breast cancer. Cancer Epidemiol Biomarkers Prev 2001;10:1281–5.
12. McCann SE, Marshall JR, Brasure JR, Graham S, Freudenheim JL. Analysis of patterns of food intake in nutritional epidemiology: food classification in principal components analysis and the subsequent impact on estimates for endometrial cancer. Public Health Nutr 2001;4:989–97.
13. Handa K, Kreiger N. Diet patterns and the risk of renal cell carcinoma. Public Health Nutr 2002;5:757–67.
14. Markaki I, Linos D, Linos A. The influence of dietary patterns on the development of thyroid cancer. Eur J Cancer 2003;39:1912–9.
15. Tsai YY, McGlynn KA, Hu Y, et al. Genetic susceptibility and dietary patterns in lung cancer. Lung Cancer 2003;41:269–81.
16. Balder HF, Virtanen M, Brants HAM, et al. Common and country-specific dietary patterns in four European cohort studies. J Nutr 2003;133: 4246–51.
17. van den Brandt PA, Goldbohm RA, Van't Veer P, Volovics A, Hermus RJ, Sturmans F. A large-scale prospective cohort study on diet and cancer in the Netherlands. J Clin Epidemiol 1990;43:285–95.
18. Goldbohm RA, van den Brandt PA, Brants HA, et al. Validation of a dietary questionnaire used in a large-scale prospective cohort study on diet and cancer. Eur J Clin Nutr 1994;48:253–65.
19. van den Brandt PA, Schouten LJ, Goldbohm RA, Dorant E, Hunen PM. Development of a record linkage protocol for use in the Dutch Cancer Registry for Epidemiological Research. Int J Epidemiol 1990;19:553–8.
20. Goldbohm RA, van den Brandt PA, Dorant E. Estimation of the coverage of Dutch municipalities by cancer registries and PALGA based on hospital discharge data. Tijdschr Soc Gezondheidsz 1994;72:80–4.
21. 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.
22. Togo P, Heitmann BL, Sorensen TIA, Osler M. Consistency of food intake factors by different dietary assessment methods and population groups. Br J Nutr 2003;90:667–78.
23. Goldbohm RA, Van't Veer P, van den Brandt PA, et al. Reproducibility of a food frequency questionnaire and stability of dietary habits determined from five annually repeated measurements. Eur J Clin Nutr 1995; 49:420–9.
24. Martinez ME, Marshall JR, Sechrest L. Invited commentary: Factor analysis and the search for objectivity [comment]. Am J Epidemiol 1998;148:17–9.
25. Ziegler RG, Mayne ST, Swanson CA. Nutrition and lung cancer. Cancer Causes Control 1996;7:157–77.
26. World Cancer Research Fund, Food, nutrition and the prevention of cancer: a global perspective. Washington: World Cancer Research Fund/American Institute for Cancer Research; 1997.
27. Goldbohm RA, Voorrips LE. Epidemiology of nutrition and lung cancer. Nestle Nutr Workshop Ser Clin Perform 2000;4:23–35.
28. Goodman MT, Hankin JH, Wilkens LR, Kolonel LN. High-fat foods and the risk of lung cancer. Epidemiology 1992;3:288–99.