American Journal of Epidemiology
Published by the Johns Hopkins Bloomberg School of Public Health 2008.

Vol. 168, No. 9
DOI: 10.1093/aje/kwn212
Advance Access publication September 12, 2008

Original Contribution

Intakes of Fruit, Vegetables, and Specific Botanical Groups in Relation to Lung Cancer Risk in the NIH-AARP Diet and Health Study

Margaret E. Wright, Yikyung Park, Amy F. Subar, Neal D. Freedman, Demetrius Albanes, Albert Hollenbeck, Michael F. Leitzmann, and Arthur Schatzkin

Received for publication November 16, 2007; accepted for publication June 13, 2008.

Increased fruit and vegetable consumption may protect against lung cancer, although epidemiologic findings are inconclusive. The authors prospectively examined associations between lung cancer risk and intakes of fruit, vegetables, and botanical subgroups in 472,081 participants aged 50–71 years in the National Institutes of Health (NIH)-AARP Diet and Health Study. Diet was assessed at baseline (1995–1996) with a 124-item dietary questionnaire. A total of 6,035 incident lung cancer cases were identified between 1995 and 2003. Total fruit and vegetable intake was unrelated to lung cancer risk in both men and women. Higher consumption of several botanical subgroups, however, was significantly inversely associated with risk, but only in men. For example, the relative risks of lung cancer among men in the highest versus lowest quintiles of intake of rosaceae, convolvulaceae, and umbelliferae were 0.82 (95% confidence interval (CI): 0.73, 0.91), 0.86 (95% CI: 0.75, 0.96), and 0.86 (95% CI: 0.78, 0.96), respectively; corresponding relative risks in women were 0.97 (95% CI: 0.85, 1.12), 0.95 (95% CI: 0.83, 1.09), and 0.92 (95% CI: 0.80, 1.06). These results provide support for a protective role of specific botanical subgroups of fruits and vegetables in lung cancer prevention in men, although the findings could also be due to residual confounding by smoking or chance.

cohort studies; fruit; lung neoplasms; vegetables

Abbreviations: CI, confidence interval; ICD-O-3, *International Classification of Diseases for Oncology*, Third Edition; NIH, National Institutes of Health; RR, relative risk.

Lung cancer is the second most commonly diagnosed malignancy and the leading cause of cancer-related death among men and women in the United States (1). Given the poor overall 5-year survival rate (16%) and the lack of a proven effect of screening on lung cancer mortality (1), primary prevention remains the most effective way to reduce the health burden of this disease.

Cigarette smoking is the predominant cause of lung cancer, accounting for approximately 90% of all cases (2). Although cessation is the most effective preventive strategy among smokers, it is exceedingly difficult to quit. Even if they are successful, former smokers continue to have an increased risk of lung cancer compared with nonsmokers throughout their lifetime (2), and their options for primary prevention are unclear. Relatively few studies have examined risk factors for lung cancer in nonsmokers, even though 10% of lung cancer patients in the United States have never smoked (3).

Cigarette smoke contains reactive oxygen and nitrogen species that cause oxidative DNA damage when inhaled (4). Fruits and vegetables are rich sources of free-radical-scavenging antioxidant nutrients, including carotenoids and vitamin C, and may therefore protect against oxidative insults associated with cigarette smoking. Most prospective cohort studies have shown modest inverse associations between fruit and/or vegetable consumption and lung cancer risk (5–15), and both a pooled analysis of 8 studies (16) and a separate meta-analysis of 11 studies (17) indicated that this was driven primarily by fruit, not vegetable, intake. With the exception of cruciferous vegetables and citrus fruits, most botanical

Correspondence to Dr. Margaret E. Wright, Department of Pathology (MC 847), College of Medicine, University of Illinois at Chicago, 840 South Wood Street, Room 130, Chicago, IL 60612 (e-mail: mewright@uic.edu).

groupings of fruits and vegetables—classified according to their phytochemical content and proposed mechanisms of biologic action—have not been extensively investigated in epidemiologic studies of diet and lung cancer risk.

We prospectively examined whether fruit and vegetable consumption and intakes of specific botanical groups were related to lung cancer risk in the National Institutes of Health (NIH)-AARP Diet and Health Study. With over 6,000 incident lung cancer cases available for analysis, this is the largest prospective study to date to have examined these hypotheses.

MATERIALS AND METHODS

Study population

The NIH-AARP Diet and Health Study was initiated in 1995–1996 when an extensive baseline questionnaire was mailed to 3.5 million AARP members aged 50–71 years residing in one of 6 US states (California, Florida, Pennsylvania, New Jersey, North Carolina, and Louisiana) or 2 US metropolitan areas (Atlanta, Georgia, and Detroit, Michigan) (18). This questionnaire ascertained information on usual dietary intake over the past 12 months, use of individual and multivitamin supplements, smoking history, and other risk factors. A total of 617,119 persons returned the baseline questionnaire, and 567,169 questionnaires were determined to have been satisfactorily completed.

Among the 567,169 persons who returned the baseline questionnaire, we excluded those with duplicate questionnaires ($n = 179$), those who had died or moved out of the study area prior to baseline ($n = 582$), those who withdrew from the study ($n = 6$), those who had questionnaires completed by proxy respondents ($n = 15,760$), those who had been previously diagnosed with cancer (except for nonmelanoma skin cancer; $n = 51,205$), and those with extreme values for total energy intake (beyond twice the interquartile range of Box-Cox log-transformed intake, corresponding to <415 kcal/day and >6,144 kcal/day for men and <317 kcal/day and >4,791 kcal/day for women; total $n = 4,434$). We also excluded subjects who reported extreme intakes of fruits and vegetables (<0.02 and >24 pyramid servings per day for men and <0.07 and >25 pyramid servings per day for women; total $n = 4,394$) and those with missing or inconsistent information on smoking habits ($n = 18,528$). After these exclusions, 281,288 men and 190,793 women were available for analysis.

The NIH-AARP Diet and Health Study was approved by the Special Studies Institutional Review Board of the National Cancer Institute.

Smoking history

Smoking history was determined using responses to 3 specific questions on the baseline questionnaire: 1) "Have you smoked 100 or more cigarettes during your entire life?" (yes, no); 2) "Do you currently smoke cigarettes or have you stopped?" (current smoker, stopped within last year, stopped 1–4 years ago, stopped 5–9 years ago, stopped 10 or more years ago); and 3) "How many cigarettes per day do you or did you usually smoke?" (1–10, 11–20, 21–30, 31–40, 41–60, or ≥61 cigarettes/day). Using this information, we constructed a multilevel categorical variable that integrated smoking status, time since quitting (for former smokers), and smoking dose (for both former and current smokers).

The questionnaire did not ascertain the age at which participants had started to smoke. Therefore, we estimated smoking duration using information obtained from the National Health Interview Survey, which found that, on average, respondents aged 50 years or more who had ever smoked at least 100 cigarettes began smoking at age 18 years, regardless of sex (19). The smoking duration variable was created by assigning a value of 0 to never smokers, subtracting 18 years plus the midpoint value of the category corresponding to time since quitting (i.e., 2.5, 7.5, or 15 years for having stopped 1–4, 5–9, or ≥10 years previously, respectively) from the age at entry for former smokers, and subtracting 18 years from the age at entry for current smokers. Adjustment for estimated smoking duration (continuous) did not materially alter our findings, and we therefore excluded this variable from all multivariate models.

Dietary intake

The dietary component of the baseline questionnaire asked about the frequency of consumption and corresponding portion sizes of 124 food items, including 14 fruit items and 23 vegetable items, during the past 12 months. Participants were queried about their frequency of intake in 10 predefined categories ranging from "never" to "2+ times per day" for solid foods and "never" to "6+ times per day" for beverages. Each line item was accompanied by 3 possible portion size categories. The food items, portion sizes, nutrient database, and pyramid food servings database were constructed using methods developed by Subar et al. (20) with national dietary data from the US Department of Agriculture's 1994–1996 Continuing Survey of Food Intake by Individuals (21). The Pyramid Servings Database utilized a recipe file to disaggregate food mixtures into their component ingredients and assign them to food groups. One pyramid serving of vegetables is equivalent to 1 cup of raw, leafy vegetables, ½ cup of other vegetables, or 6 ounces (178 mL) of juice; one pyramid serving of fruit equals 1 medium-sized piece of fresh fruit, ½ cup of chopped fruit, or 6 ounces of fruit juice (22). We excluded white potatoes from the vegetable group. Fruits and vegetables were also grouped into botanical families, in which foods are classified according to their phytochemical content and proposed mechanisms of biologic action (23). The contributors to fruits, vegetables, and specific botanical groups are listed in the Appendix.

The food frequency questionnaire was validated using 2 24-hour recalls in a subset of the cohort (24). Energy-adjusted correlation coefficients for total fruits and vegetables were 0.72 and 0.61 in men and women, respectively.

Endpoint ascertainment

Incident, first primary lung cancer cases (*International Classification of Diseases for Oncology*, Third Edition

(ICD-O-3) (25), codes C340–C349) were identified through December 31, 2003, via linkage of the NIH-AARP cohort database to the databases of the 8 state cancer registries and the National Death Index Plus. The cancer registries in all 8 states are certified by the North American Association of Central Cancer Registries as meeting the highest standard of quality (90% case ascertainment within 24 months of the close of the diagnosis year). In a validation study, we estimated that 90% of all cancer cases in our cohort were validly identified via linkage to state cancer registries, as compared with self-reports and medical records (26).

A total of 6,035 incident lung cancer cases were identified (3,834 in men and 2,201 in women) during 8 years of follow-up. Adenocarcinoma (ICD-O-3 codes 8140, 8200, 8231, 8250, 8251, 8260, 8290, 8310, 8323, 8430, 8480, 8481, 8490, and 8550) was the most common histologic type, accounting for 47% of cases in men and 54% of cases in women. In men, squamous cell carcinoma (ICD-O-3 codes 8050, 8070, 8071, 8072, 8073, and 8074) was the second most frequent type (27%), followed by small cell (ICD-O-3 codes 8002, 8041, 8042, 8044, and 8045) (18%) and undifferentiated large cell (ICD-O-3 codes 8012, 8020, 8021, 8022, 8031, and 8032) (8%) carcinomas. In women, small cell carcinoma accounted for 21% of lung cancer cases, followed by squamous cell (18%) and large cell (8%) carcinomas.

Statistical analysis

Follow-up time for each participant accrued from the date of return of the baseline questionnaire to the date of lung cancer diagnosis, the date of moving out of the registry ascertainment area, death, or the end of the follow-up period. Cox proportional hazards models with age as the underlying time metric were used to estimate relative risks and 95% confidence intervals for lung cancer according to sex-specific quintiles of intake of fruit, vegetables, and selected botanical groups. Tests for linear trend were conducted using the median value of each exposure category to create a continuous variable. The proportional hazards assumption was tested and upheld in all analyses. In addition to age, all multivariate models were adjusted for smoking status, smoking dose, and time since quitting (described above), body mass index (weight (kg)/height (m)^2; <25, 25–29.9, 30–34.9, $\geq$35), race (White, Black, other), educational attainment (<12 years, 12 years or high school equivalent, some college, college graduate or postgraduate), physical activity level (none, rarely, 1–3 times per month, 1–2 times per week, 3–4 times per week, $\geq$5 times per week), family history of any cancer (yes, no), and alcohol consumption (0, >0–4.9 g/day, 5–14.9 g/day, 15–29.9 g/day, $\geq$30 g/day). Consumption of fruit, vegetables, and individual botanical groups was adjusted for energy intake using the nutrient density method (27) and was expressed as pyramid servings per 1,000 kcal per day.

We examined whether the associations between plant food consumption and lung cancer risk varied according to subgroups defined by smoking status and histologic type of disease. We tested for effect modification by adding the relevant cross-product term to main-effects models.

For all comparisons, P values were 2-sided and $\alpha < 0.05$ indicated statistical significance.

RESULTS

In both men and women, higher fruit and vegetable intake was associated with elements of a healthier lifestyle (Table 1). Persons in higher quintiles were older, leaner, more physically active, and more likely to be a never smoker and to report a higher level of attained education, and they consumed less energy and alcohol than those in the lowest quintile. On average, men reported consuming 3 servings of total fruits and vegetables per 1,000 kcal per day and women reported consuming 4 servings per 1,000 kcal per day.

Higher intake of fruit and vegetables—overall and when analyzed separately—was strongly inversely associated with lung cancer risk in age-adjusted models in both men and women (for highest vs. lowest quintiles of total fruit and vegetable consumption, relative risk (RR) = 0.41, 95% confidence interval (CI): 0.37, 0.46 ($P < 0.0001$) in men and RR = 0.46, 95% CI: 0.41, 0.53 ($P < 0.0001$) in women). Addition of smoking status, smoking dose (in both current and former smokers), and time since quitting (for former smokers) to these models greatly diminished the apparent benefits of fruit and vegetable intake (RR = 0.86, 95% CI: 0.77, 0.95 ($P = 0.003$) in men and RR = 0.92, 95% CI: 0.80, 1.06 ($P = 0.51$) in women), with adjustment for additional covariates further attenuating risk estimates in both sexes (Table 2); there was no evidence of a dose-response trend. Mutual adjustment of total fruits and total vegetables did not alter the observed associations (data not shown).

In men, lung cancer risks tended to be somewhat lower in never and former smokers who consumed more fruits and vegetables, although none of the risk estimates were statistically significant and all P values for interaction were nonsignificant (Table 2). In women, there was little variation in the associations of fruits and vegetables with lung cancer across smoking categories (Table 2). There were no differences when fruit and vegetable intake was examined with respect to the different histologic types of lung cancer in either sex.

Consumption of most botanical groups was unrelated to lung cancer risk (Table 3). Significant inverse associations were noted, however, for intakes of rosaceae (apples, peaches, nectarines, plums, pears, and strawberries), convolvulaceae (sweet potatoes and yams), and umbelliferae (carrots) in men, with borderline inverse relations also apparent for compositae and cruciferae. No such beneficial relations were noted in women. Analyses using deciles of intake—undertaken in order to explore patterns of risk at the extremes of consumption—generally yielded results that were similar to those based on quintiles, although significant inverse associations emerged among men in the top decile of legume consumption (>0.69 servings per 1,000 kcal per day; RR = 0.85, 95% CI: 0.74, 0.98; $P = 0.07$) and among women in the top decile of rutaceae (citrus) fruit intake (>1.51 servings per 1,000 kcal per day; RR = 0.80, 95% CI: 0.66, 0.97; $P = 0.28$).

Smoking status significantly modified the association between intakes of convolvulaceae, umbelliferae, cruciferae, chenopodiaceae, and gramineae foods and lung cancer risk in men (P's for interaction = 0.005, 0.008, 0.05, 0.003, and 0.04, respectively), with the protective effects being limited

Table 1. Baseline Characteristics (Means and Proportions) of the Study Population According to Quintile of Total Fruit and Vegetable Consumption, NIH-AARP Diet and Health Study, 1995–1996

Characteristic	Quintile of Total Fruit and Vegetable Consumption				
	1	2	3	4	5
Men					
No. of participants	56,257	56,258	56,258	56,258	56,257
Age, years	61.4	62.0	62.2	62.5	62.7
Body mass index[a]	27.4	27.4	27.3	27.2	27.0
Physical activity[b], no. of times per week	2.8	3.1	3.3	3.4	3.6
Race, %					
White	95.1	94.9	94.5	93.5	91.4
Black	2.3	2.3	2.3	2.6	3.5
Other[c]	2.6	2.8	3.2	3.9	5.1
Smoking status, %					
Never smoker	21.2	28.3	31.6	34.0	36.2
Former smoker	53.2	57.3	58.4	58.5	58.2
Quit ≥10 years ago	73.6	78.4	80.7	81.9	83.1
Quit 5–9 years ago	16.1	13.7	12.8	12.1	11.2
Quit 1–4 years ago	10.3	7.9	6.5	6.0	5.7
Current smoker	25.6	14.4	10.1	7.6	5.6
≤20 cigarettes/day	48.8	58.2	61.8	67.1	71.9
21–40 cigarettes/day	43.6	37.4	33.8	29.8	25.4
>40 cigarettes/day	7.6	4.4	4.4	3.1	2.7
Education ≥college degree, %	36.1	45.5	49.3	52.0	53.2
Family history of any cancer, % yes	47.5	48.1	47.9	47.2	46.1
Total energy intake, kcal/day	2,310	2,091	1,992	1,887	1,725
Median alcohol intake, g/day	5.3	4.7	4.0	3.3	2.0
Women					
No. of participants	38,158	38,159	38,159	38,159	38,158
Age, years	61.2	61.7	62.0	62.1	62.2
Body mass index	27.3	27.1	26.9	26.7	26.2
Physical activity, no. of times per week	2.4	2.8	3.0	3.2	3.6
Race, %					
White	93.0	92.9	92.0	90.6	87.9
Black	4.4	4.4	5.0	5.7	7.3
Other	2.6	2.8	3.1	3.7	4.8
Smoking status, %					
Never smoker	38.1	44.7	47.2	48.4	49.2
Former smoker	32.3	37.2	38.8	39.8	40.7
Quit ≥10 years ago	63.8	69.0	70.8	72.6	73.2
Quit 5–9 years ago	20.9	19.0	17.9	17.2	17.3
Quit 1–4 years ago	15.4	12.0	11.3	10.2	9.5
Current smoker	29.6	18.2	14.0	11.8	10.1
≤20 cigarettes/day	64.2	73.5	77.0	79.0	84.0
21–40 cigarettes/day	32.8	24.7	21.5	19.8	15.4
>40 cigarettes/day	3.0	1.8	1.5	1.2	0.6
Education ≥college degree, %	24.8	31.2	34.6	36.0	38.1
Family history of any cancer, % yes	51.7	52.3	52.0	51.5	50.5
Total energy intake, kcal/day	1,697	1,623	1,565	1,511	1,392
Median alcohol intake, g/day	0.85	1.1	1.1	0.87	0.65

Abbreviation: NIH, National Institutes of Health.

[a] Weight (kg)/height (m)^2.

[b] Defined as engaging in physical activity for at least 20 minutes that caused increases in breathing or heart rate or worked up a sweat.

[c] Includes Hispanics, Asians, Pacific Islanders, and American Indians/Alaska Natives.

Table 2. Relative Risk of Lung Cancer According to Quintile of Fruit and Vegetable Consumption, Overall and by Smoking Status, NIH-AARP Diet and Health Study, 1995–2003

Sex and Intake, servings/1,000 kcal/day	All Participants			Never Smokers			Former Smokers			Current Smokers		
	No. of Cases	RR[a]	95% CI	No. of Cases	RR[b]	95% CI	No. of Cases	RR[c]	95% CI	No. of Cases	RR[d]	95% CI
Men												
Total fruit and vegetables												
<1.82	1,212	1.00		23	1.00		501	1.00		688	1.00	
1.82–2.53	809	0.95	0.87, 1.04	25	0.77	0.43, 1.36	421	0.90	0.79, 1.03	363	1.05	0.89, 1.24
2.54–3.25	681	0.95	0.86, 1.04	28	0.76	0.44, 1.33	424	0.94	0.83, 1.08	229	1.02	0.84, 1.23
3.26–4.29	600	0.93	0.84. 1.04	32	0.79	0.46, 1.38	384	0.88	0.77, 1.01	184	1.12	0.91, 1.38
>4.29	532	0.93	0.83, 1.04	33	0.77	0.44, 1.35	380	0.91	0.79, 1.05	119	1.00	0.77, 1.29
P trend			0.17			0.56			0.22			0.69
Total fruit												
<0.65	1,195	1.00		21	1.00		497	1.00		677	1.00	
0.65–1.08	837	0.98	0.89, 1.07	27	0.89	0.50, 1.58	441	0.94	0.82, 1.07	369	1.01	0.89, 1.15
1.09–1.56	660	0.91	0.83, 1.01	33	0.92	0.53, 1.60	384	0.86	0.75, 0.99	243	0.96	0.83, 1.12
1.57–2.27	619	0.95	0.86, 1.06	24	0.60	0.33, 1.09	412	0.95	0.83, 1.09	183	0.97	0.82, 1.14
>2.27	523	0.91	0.82, 1.02	36	0.81	0.46, 1.41	376	0.91	0.79, 1.05	111	0.84	0.69, 1.04
P trend			0.10			0.35			0.36			0.12
Total vegetables												
<0.87	1,058	1.00		27	1.00		495	1.00		536	1.00	
0.87–1.22	826	0.98	0.89, 1.07	27	0.85	0.50, 1.45	445	0.97	0.85, 1.10	354	0.98	0.86, 1.12
1.23–1.61	725	0.96	0.87, 1.05	27	0.84	0.49, 1.43	418	0.92	0.80, 1.05	280	0.99	0.86, 1.15
1.62–2.20	633	0.91	0.82, 1.01	28	0.85	0.50, 1.45	360	0.81	0.70, 0.93	245	1.07	0.92, 1.25
>2.20	592	0.93	0.83, 1.03	32	0.94	0.56, 1.59	392	0.88	0.77, 1.01	168	0.97	0.81, 1.16
P trend			0.08			0.99			0.01			0.90
Women												
Total fruit and vegetables												
<2.39	658	1.00		25	1.00		162	1.00		471	1.00	
2.39–3.25	439	0.94	0.83, 1.07	30	0.97	0.57, 1.65	155	0.95	0.76, 1.18	254	0.96	0.82, 1.12
3.26–4.14	384	0.95	0.83, 1.08	36	1.06	0.63, 1.77	172	1.04	0.84, 1.30	176	0.88	0.73, 1.05
4.15–5.37	390	1.06	0.93, 1.21	43	1.18	0.71, 1.97	174	1.06	0.85, 1.33	173	1.06	0.89, 1.27
>5.37	330	0.98	0.85, 1.13	36	0.99	0.58, 1.69	172	1.03	0.82, 1.29	122	0.93	0.76, 1.15
P trend			0.56			0.86			0.55			0.73
Total fruit												
<0.89	674	1.00		23	1.00		179	1.00		472	1.00	
0.89–1.41	454	0.96	0.85, 1.09	30	1.00	0.58, 1.72	168	0.94	0.76, 1.16	256	0.98	0.84, 1.14
1.42–1.97	394	0.97	0.86, 1.11	43	1.28	0.77, 2.13	160	0.92	0.74, 1.14	191	0.98	0.82, 1.16
1.98–2.76	358	0.98	0.86, 1.12	32	0.84	0.49, 1.46	176	1.03	0.83, 1.28	150	0.95	0.79, 1.15
>2.76	321	0.97	0.84, 1.11	42	1.08	0.64, 1.84	152	0.94	0.75, 1.17	127	0.95	0.78, 1.17
P trend			0.70			0.99			0.85			0.58
Total vegetables												
<1.11	564	1.00		32	1.00		135	1.00		397	1.00	
1.11–1.56	426	0.94	0.83, 1.07	34	0.96	0.59, 1.55	150	1.09	0.86, 1.38	242	0.89	0.76, 1.05
1.57–2.08	410	0.99	0.87, 1.13	37	0.99	0.61, 1.60	183	1.29	1.03, 1.62	190	0.84	0.70, 1.00
2.09–2.86	410	1.03	0.91, 1.18	41	1.09	0.68, 1.75	171	1.18	0.94, 1.48	198	0.97	0.82, 1.16
>2.86	391	1.05	0.92, 1.21	26	0.72	0.42, 1.22	196	1.26	1.01, 1.58	169	1.01	0.84, 1.22
P trend			0.23			0.27			0.07			0.75

Abbreviations: CI, confidence interval; NIH, National Institutes of Health; RR, relative risk.

[a] Adjusted for age, energy intake, race, education, body mass index, smoking status, smoking dose, time since quitting smoking, alcohol intake, physical activity, and family history of any cancer.

[b] Adjusted for age, energy intake, race, education, body mass index, alcohol intake, physical activity, and family history of any cancer.

[c] Adjusted for age, energy intake, race, education, body mass index, time since quitting smoking, past smoking dose, alcohol intake, physical activity, and family history of any cancer.

[d] Adjusted for age, energy intake, race, education, body mass index, current smoking dose, alcohol intake, physical activity, and family history of any cancer.

Table 3. Multivariate[a] Relative Risk of Lung Cancer According to Quintile of Intake of Selected Botanical Groups, NIH-AARP Diet and Health Study, 1995–2003

Botanical Group and Sex	Quintile of intake																		P for Trend
	1[b]		2				3				4				5				
	No. of Cases	Intake[c]	No. of Cases	Intake	RR	95% CI	No. of Cases	Intake	RR	95% CI	No. of Cases	Intake	RR	95% CI	No. of Cases	Intake	RR	95% CI	
Cucurbitaceae																			
Men	932	0.003	803	0.01	0.97	0.88, 1.07	717	0.03	0.98	0.88, 1.08	728	0.07	1.04	0.94, 1.14	654	0.20	0.95	0.86, 1.05	0.50
Women	564	0.006	439	0.02	0.92	0.81, 1.04	401	0.05	0.92	0.81, 1.05	378	0.13	0.92	0.81, 1.05	419	0.33	1.05	0.92, 1.19	0.14
Musaceae																			
Men	1,003	0.01	849	0.05	1.00	0.91, 1.09	722	0.14	0.97	0.88, 1.07	645	0.29	0.94	0.85, 1.04	615	0.53	0.97	0.87, 1.08	0.42
Women	604	0.01	480	0.07	0.98	0.86, 1.10	343	0.18	0.80	0.70, 0.91	378	0.35	0.94	0.83, 1.08	396	0.63	1.01	0.88, 1.15	0.73
Rosaceae																			
Men	1,154	0.03	824	0.09	0.94	0.85, 1.02	699	0.18	0.92	0.83, 1.01	654	0.35	0.97	0.88, 1.07	503	0.72	0.82	0.73, 0.91	0.002
Women	670	0.05	481	0.14	1.00	0.89, 1.12	403	0.27	1.01	0.89, 1.14	334	0.49	0.93	0.81, 1.07	313	0.94	0.97	0.85, 1.12	0.52
Rutaceae																			
Men	1,068	0.04	808	0.19	0.99	0.91, 1.09	715	0.43	0.97	0.88, 1.07	635	0.73	0.95	0.86, 1.05	608	1.35	0.99	0.89, 1.10	0.68
Women	633	0.05	432	0.21	0.90	0.79, 1.01	369	0.50	0.86	0.75, 0.98	414	0.84	1.00	0.88, 1.13	353	1.51	0.91	0.79, 1.04	0.55
Vitaceae																			
Men	1,045	0.002	808	0.01	0.95	0.87, 1.05	689	0.03	0.93	0.84, 1.03	672	0.06	0.97	0.87, 1.07	620	0.18	0.94	0.85, 1.04	0.47
Women	567	0	434	0.02	0.93	0.82, 1.05	408	0.04	0.95	0.84, 1.08	408	0.08	1.03	0.91, 1.17	384	0.29	0.99	0.86, 1.13	0.70
Chenopodiaceae																			
Men	957	0	675	0.006	0.91	0.83, 1.01	824	0.01	1.03	0.94, 1.14	738	0.04	0.97	0.88, 1.07	640	0.13	0.94	0.85, 1.04	0.33
Women	474	0	455	0.01	1.02	0.89, 1.16	440	0.03	1.00	0.88, 1.14	432	0.06	1.02	0.89, 1.16	400	0.21	1.00	0.87, 1.15	0.93
Compositae																			
Men	1,026	0.01	824	0.06	0.98	0.90, 1.08	720	0.14	0.96	0.87, 1.05	657	0.26	0.95	0.86, 1.05	607	0.63	0.92	0.83, 1.02	0.09
Women	543	0.02	424	0.09	0.89	0.78, 1.01	413	0.21	0.94	0.83, 1.07	395	0.39	0.93	0.81, 1.06	426	0.91	1.02	0.89, 1.16	0.29
Convolvulaceae																			
Men	897	0	808	0.005	1.00	0.91, 1.10	799	0.01	0.98	0.89, 1.08	695	0.03	0.91	0.82, 1.00	635	0.07	0.86	0.75, 0.96	0.0008
Women	529	0	432	0.009	0.88	0.77, 1.00	416	0.02	0.88	0.77, 1.00	421	0.04	0.93	0.82, 1.06	403	0.11	0.95	0.83, 1.09	0.77
Cruciferae																			
Men	1,024	0.03	839	0.08	0.97	0.89, 1.07	719	0.15	0.91	0.83, 1.00	641	0.25	0.90	0.81, 0.99	611	0.50	0.92	0.83, 1.02	0.09
Women	522	0.06	439	0.14	0.96	0.84, 1.09	417	0.24	0.99	0.87, 1.13	427	0.39	1.06	0.93, 1.20	396	0.77	1.00	0.87, 1.14	0.65
Gramineae																			
Men	924	0.009	781	0.03	0.93	0.85, 1.03	777	0.05	0.98	0.89, 1.08	661	0.08	0.86	0.78, 0.95	691	0.16	0.96	0.87, 1.06	0.30
Women	525	0.008	460	0.03	0.97	0.85, 1.09	443	0.05	0.99	0.87, 1.12	407	0.09	0.95	0.84, 1.09	366	0.17	0.91	0.79, 1.04	0.16
Legumes																			
Men	918	0.08	808	0.17	0.98	0.89, 1.07	716	0.27	0.90	0.81, 0.99	733	0.40	0.98	0.89, 1.08	659	0.69	0.92	0.83, 1.02	0.16
Women	512	0.09	483	0.19	1.07	0.94, 1.21	414	0.30	0.95	0.83, 1.08	380	0.45	0.93	0.81, 1.06	412	0.81	1.07	0.94, 1.22	0.65
Solanaceae																			
Men	911	0.07	766	0.15	0.97	0.88, 1.07	734	0.23	0.98	0.89, 1.08	674	0.34	0.94	0.85, 1.03	749	0.60	1.02	0.93, 1.13	0.60
Women	472	0.08	428	0.17	0.98	0.86, 1.11	411	0.27	1.00	0.87, 1.14	423	0.40	1.04	0.91, 1.19	467	0.71	1.11	0.97, 1.27	0.04
Umbelliferae																			
Men	1,070	0.005	848	0.02	0.97	0.89, 1.07	703	0.03	0.89	0.80, 0.98	655	0.07	0.91	0.83, 1.01	558	0.21	0.86	0.78, 0.96	0.01
Women	593	0.008	475	0.03	0.96	0.85, 1.08	416	0.06	0.92	0.81, 1.04	388	0.13	0.93	0.82, 1.06	329	0.36	0.92	0.80, 1.06	0.38

Abbreviations: CI, confidence interval; NIH, National Institutes of Health; RR, relative risk.

[a] Adjusted for age, energy intake, race, education, body mass index, smoking status, smoking dose, time since quitting smoking, alcohol intake, physical activity, and family history of any cancer.

[b] Reference category (RR = 1).

[c] Median number of servings/1,000 kcal/day.

to former smokers in each instance (Table 4). With the exception of compositae intake ($P = 0.02$), associations between specific botanical groups and lung cancer risk did not vary according to smoking status in women (Table 4; all P's $>$ 0.05). In analyses stratified by histologic type, among men, higher consumption of the compositae group was most protective against squamous cell carcinoma (for highest quintile vs. lowest, RR = 0.74, 95% CI: 0.58, 0.93; $P = 0.02$), whereas intake of cruciferae was most strongly protective against small cell carcinoma (RR = 0.74, 95% CI: 0.55, 0.99; $P = 0.10$). Among women, notable inverse associations were observed between squamous cell carcinoma and intakes of musaceae (RR = 0.66, 95% CI: 0.45, 0.96; $P =$ 0.02) and rutaceae (RR = 0.65, 95% CI: 0.45, 0.96; $P =$ 0.04) foods, whereas higher consumption of umbelliferae foods was related to a lower risk of undifferentiated large cell lung cancer (RR = 0.52, 95% CI: 0.28, 0.94; $P = 0.03$).

Exclusion of lung cancer cases diagnosed during the first year of follow-up ($n = 459$ in men and $n = 264$ in women) resulted in slightly stronger inverse associations between fruit and vegetable intake and lung cancer risk in men, with virtually no changes in risk estimates in women (data not shown).

DISCUSSION

In this large prospective study, we observed no relations between total intake of fruit and vegetables combined or total fruit or total vegetable consumption considered separately and lung cancer risk. Higher consumption of several botanical groups, however, including rosaceae (apples, peaches, nectarines, plums, pears, and strawberries), convolvulaceae (sweet potatoes and yams), and umbelliferae (carrots), was significantly inversely associated with lung cancer risk in men. Smoking status modified these relations, with the most pronounced inverse associations being apparent in former smokers; no such effect modification was noted in women.

More than 10 prospective studies have examined whether higher fruit and vegetable consumption is associated with a decreased risk of lung cancer, with most (5–15), though not all (28, 29), showing some degree of protection. Both a pooled analysis (16) and a meta-analysis (17) of these reports indicated that fruit consumption was primarily responsible for the observed beneficial associations. In several studies, investigators further analyzed the relation of individual plant foods and/or plant food groups with lung cancer risk; the most consistent findings were for apples and pears (5, 6, 9), cruciferous vegetables (5, 11, 14), and citrus fruits (5, 14). In the present study, we found an inverse association between consumption of rosaceae foods (apples, peaches, nectarines, plums, pears, and strawberries) and lung cancer risk in men but observed no such relation for rutaceae (citrus) intake. While increased consumption of cruciferous vegetables was associated with modest reductions in risk among men in our study, these findings were of borderline statistical significance. Similarly to Knekt et al. (8), we demonstrated a strong reduction in lung cancer risk among men who consumed higher amounts of carrots.

A relatively large number of lung tumors (311 total: 141 in men, 170 in women) were diagnosed among never smokers in our cohort, which enabled us to examine the relation between plant food intake and lung cancer risk in a subgroup that was free from possible residual confounding by smoking habits. Several previous studies showed that the protective effects of fruit and/or vegetable consumption were strongest in current (9, 14, 15) or former (13) smokers, whereas 2 reports identified a significant beneficial association between total fruit intake and lung cancer risk in never smokers (5, 10). We did not observe significant inverse relations between overall fruit and vegetable intake and lung cancer risk in any smoking stratum, although analyses of individual botanical groups revealed that, in men, the most pronounced inverse associations occurred in former smokers.

It is unclear why beneficial associations between intakes of several botanical groups and overall lung cancer risk were restricted to males in our cohort. Men and women appear to have similar risks of lung cancer in response to a given level of tobacco exposure (30), although there is evidence that the biology of the disease differs between the sexes. Hormonal (i.e., estrogen), genetic, and metabolic factors may contribute to these differences (30). The observed gender differences may also be related to differential reporting errors, with women being more likely to overreport consumption of foods perceived as healthy (31, 32). Any ensuing exposure misclassification could have attenuated modest associations between fruit and vegetable intake and lung cancer risk.

It is not known which bioactive compounds in specific botanical classes of fruits and vegetables might be responsible for a protective effect of these foods against lung cancer. Plant foods are rich sources of a myriad of potentially anticarcinogenic substances, including antioxidant nutrients, folic acid, and fiber (33). In our analysis, we observed that increased consumption of umbelliferae, convolvulaceae, and rosaceae foods was associated with a decreased risk of lung cancer in men. Carrots are a major contributor to the umbelliferae botanical group and sweet potatoes are part of the convolvulaceae family; both of these root vegetables are rich sources of pro-vitamin A carotenoids, including beta-carotene, and vitamin C. Both micronutrients are powerful free-radical-scavenging antioxidants, with beta-carotene additionally serving as a precursor of vitamin A—an essential regulator of epithelial cell division, growth, differentiation, and proliferation (34, 35). However, 2 randomized controlled trials demonstrated that supplemental beta-carotene increased the risk of lung cancer in smokers (36, 37), which contradicted the large body of observational data that consistently showed higher dietary intake of this micronutrient to be linked with a lower risk of lung cancer (38). These discrepancies probably arose because supplements contain only large quantities of a single nutrient, whereas dietary antioxidants are likely to exert their protective effects through interactions with other vitamins and phytochemicals found in the same food sources. Furthermore, beta-carotene appears to exhibit distinct biologic functions at different concentrations: It is an effective antioxidant at low doses but adversely alters retinoid signaling through an antioxidant-independent mechanism at higher

Table 4. Relative Risk of Lung Cancer in the Highest Versus the Lowest, Quintile of Intake of Selected Botanical Groups, by Smoking Status, NIH-AARP Diet and Health Study, 1995–2003

	Never Smokers				Former Smokers				Current Smokers			
	No. of Cases (High/Low)[a]	RR[b]	95% CI	*P* for Trend[c]	No. of Cases (High/Low)	RR[d]	95% CI	*P* for Trend	No. of Cases (High/Low)	RR[e]	95% CI	*P* for Trend
Cucurbitaceae												
Men	32/22	1.21	0.70, 2.09	0.87	419/444	0.93	0.81, 1.06	0.49	203/466	0.94	0.80, 1.11	0.66
Women	38/28	1.21	0.74, 1.97	0.71	202/161	1.10	0.89, 1.36	0.06	179/375	0.99	0.83, 1.19	0.94
Musaceae												
Men	34/14	1.57	0.83, 2.96	0.34	420/443	0.95	0.83, 1.09	0.25	161/546	0.98	0.82, 1.18	0.67
Women	43/32	0.97	0.61, 1.54	0.48	192/185	1.03	0.84, 1.27	0.62	161/387	0.96	0.80, 1.16	0.65
Rosaceae												
Men	37/27	0.75	0.45, 1.26	0.98	359/512	0.81	0.70, 0.93	0.007	107/615	0.76	0.62, 0.94	0.07
Women	42/20	1.31	0.76, 2.26	0.50	153/186	0.91	0.73, 1.13	0.17	118/464	1.01	0.82, 1.24	0.95
Rutaceae												
Men	26/24	0.63	0.36, 1.10	0.09	407/485	1.00	0.88, 1.15	0.92	175/559	0.99	0.84, 1.18	0.68
Women	36/28	0.82	0.50, 1.36	0.81	167/188	1.00	0.81, 1.24	0.57	150/417	0.84	0.70, 1.02	0.18
Vitaceae												
Men	24/29	0.57	0.33, 0.99	0.07	394/497	0.94	0.82, 1.07	0.53	202/519	0.98	0.83, 1.16	0.89
Women	32/31	0.88	0.54, 1.45	0.43	167/177	0.97	0.78, 1.20	0.99	185/359	1.03	0.86, 1.23	0.42
Chenopodiaceae												
Men	39/28	1.56	0.95, 2.56	0.02	358/515	0.88	0.77, 1.01	0.08	243/414	0.98	0.83, 1.15	0.84
Women	30/39	0.72	0.44, 1.18	0.32	168/162	1.01	0.81, 1.26	0.93	202/273	1.03	0.86, 1.25	0.92
Compositae												
Men	23/27	0.78	0.44, 1.37	0.17	387/497	0.87	0.76, 1.00	0.06	197/502	0.99	0.84, 1.18	0.91
Women	35/32	1.16	0.71, 1.89	0.45	212/155	1.15	0.93, 1.42	0.05	179/356	0.91	0.76, 1.10	0.44
Convolvulaceae												
Men	34/33	0.71	0.43, 1.15	0.46	361/466	0.79	0.69, 0.91	<0.0001	240/398	1.00	0.85, 1.18	0.99
Women	37/43	0.64	0.41, 1.00	0.45	190/169	1.08	0.88, 1.34	0.09	176/317	0.89	0.74, 1.07	0.34
Cruciferae												
Men	32/24	1.10	0.64, 1.87	0.61	385/514	0.85	0.74, 0.97	0.03	194/486	0.99	0.84, 1.17	0.83
Women	25/34	0.66	0.39, 1.12	0.06	179/139	1.13	0.90, 1.42	0.53	192/349	1.01	0.84, 1.20	0.46
Gramineae												
Men	34/26	1.19	0.71, 1.99	0.62	399/514	0.85	0.74, 0.97	0.02	258/384	1.10	0.94, 1.29	0.40
Women	38/26	1.30	0.79, 2.15	0.40	139/194	0.91	0.73, 1.13	0.41	189/305	0.86	0.72, 1.03	0.10
Legumes												
Men	32/32	0.87	0.53, 1.43	0.85	390/466	0.86	0.75, 0.99	0.03	237/420	1.00	0.85, 1.17	0.79
Women	34/33	0.83	0.51, 1.35	0.59	178/172	1.14	0.93, 1.41	0.26	200/307	1.05	0.88, 1.26	0.87
Solanaceae												
Men	34/24	1.59	0.94, 2.70	0.10	443/429	1.00	0.87, 1.14	0.50	272/458	1.04	0.90, 1.22	0.26
Women	31/39	0.82	0.51, 1.33	0.74	205/153	1.08	0.87, 1.34	0.26	231/280	1.17	0.98, 1.40	0.06
Umbelliferae												
Men	32/29	0.76	0.46, 1.27	0.78	340/524	0.77	0.67, 0.89	0.0009	186/517	1.02	0.86, 1.21	0.81
Women	30/28	0.76	0.45, 1.28	0.13	155/180	0.90	0.72, 1.12	0.64	144/385	0.99	0.81, 1.21	0.80

Abbreviations: CI, confidence interval; NIH, National Institutes of Health; RR, relative risk.

[a] Numbers of cases in the highest and lowest quintiles, respectively.

[b] Adjusted for age, energy intake, race, education, body mass index, alcohol intake, physical activity, and family history of any cancer.

[c] *P* value for trend across quintiles of intake.

[d] Adjusted for age, energy intake, race, education, body mass index, time since quitting smoking, past smoking dose, alcohol intake, physical activity, and family history of any cancer.

[e] Adjusted for age, energy intake, race, education, body mass index, current smoking dose, alcohol intake, physical activity, and family history of any cancer.

levels (38). The rosaceae botanical family includes foods that are rich in flavonoids—a group of phytochemicals with demonstrated antioxidant activity (39). Several prospective studies have shown an inverse association between flavonoid intake (particularly quercetin, which is concentrated in apples) and lung cancer risk (40). Analyses of the relation of lung cancer to individual antioxidant nutrients, including carotenoids, flavonoids, tocopherols, vitamin C, and selenium, in the NIH-AARP Diet and Health Study are under way.

Strengths of this study include its prospective design, the large number of incident lung cancer cases, including a substantial number of cases in never smokers, a wide range of self-reported fruit and vegetable intakes, and our ability to control for a large number of potentially important confounders. Additional strengths include investigation of botanical subgroups and our ability to evaluate whether relations varied by histologic type of disease. A limitation is the lack of information on smoking duration, although we were able to estimate it using data on age at smoking initiation from the National Health Interview Survey—a nationally representative sample of the civilian noninstitutionalized population of the United States (19). Addition of this variable to multivariate models did not alter any of the risk estimates. Nevertheless, it is still possible that our significant subgroup findings could be explained by inadequate control for smoking characteristics, as it is well-known that cigarette smoking is highly associated with patterns of nutrient intake, including fruit and vegetable consumption (41, 42). Furthermore, we had no information on passive smoke exposure, which could have confounded the observed relations. Another limitation is that we relied on 1 measure of fruit and vegetable consumption from a food frequency questionnaire, which is an instrument that is subject to measurement error (43). Furthermore, we ascertained intake in middle age, which may not be the etiologically relevant exposure period. Finally, we performed a large number of statistical tests, and some of our significant associations could have been due to chance.

In summary, we found that increased consumption of several botanical groups, but not overall fruit and vegetable intake, was associated with modest reductions in lung cancer risk, with the former results being evident only in men. Although provocative, our findings should be interpreted with caution since they could be due to chance. Because smoking remains the predominant risk factor for lung cancer, public health efforts should continue to focus primarily on smoking prevention and cessation as a means of reducing the incidence of and mortality from this lethal disease.

ACKNOWLEDGMENTS

Author affiliations: Department of Pathology, College of Medicine, University of Illinois at Chicago, Chicago, Illinois (Margaret E. Wright); Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland (Margaret E. Wright, Yikyung Park, Neal D. Freedman, Demetrius Albanes, Michael F. Leitzmann, Arthur Schatzkin); Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, Maryland (Amy F. Subar); Cancer Prevention Fellowship Program, Division of Cancer Prevention, National Cancer Institute, Bethesda, Maryland (Neal D. Freedman); and AARP, Washington, DC (Albert Hollenbeck).

The authors thank Drs. Jay Lubin and Neil Caporaso for statistical advice.

Cancer incidence data from the Atlanta, Georgia, metropolitan area were collected by the Georgia Center for Cancer Statistics, Department of Epidemiology, Rollins School of Public Health, Emory University. Cancer incidence data from California were collected by the California Department of Health Services, Cancer Surveillance Section. Cancer incidence data from the Detroit, Michigan, metropolitan area were collected by the Michigan Cancer Surveillance Program, Community Health Administration, State of Michigan. The Florida cancer incidence data used in this report were collected by the Florida Cancer Data System under a contract with the Department of Health. (The views expressed herein are solely those of the authors and do not necessarily reflect those of the contractor or the Department of Health.) Cancer incidence data from Louisiana were collected by the Louisiana Tumor Registry, Louisiana State University Medical Center in New Orleans. Cancer incidence data from New Jersey were collected by the New Jersey State Cancer Registry, Cancer Epidemiology Services, New Jersey State Department of Health and Senior Services. Cancer incidence data from North Carolina were collected by the North Carolina Central Cancer Registry. Cancer incidence data from Pennsylvania were supplied by the Division of Health Statistics and Research, Pennsylvania Department of Health. (The Pennsylvania Department of Health specifically disclaims responsibility for any analyses, interpretations, or conclusions.)

Conflict of interest: none declared.

REFERENCES

1. American Cancer Society. *Cancer Facts & Figures 2007*. Washington, DC: American Cancer Society; 2007.
2. Alberg AJ, Samet JM. Epidemiology of lung cancer. *Chest*. 2003;123(1 suppl):21S–49S.
3. Subramanian J, Govindan R. Lung cancer in never smokers: a review. *J Clin Oncol.* 2007;25(5):561–570.
4. Hecht SS. Tobacco smoke carcinogens and lung cancer. *J Natl Cancer Inst.* 1999;91(14):1194–1210.
5. Feskanich D, Ziegler RG, Michaud DS, et al. Prospective study of fruit and vegetable consumption and risk of lung cancer among men and women. *J Natl Cancer Inst.* 2000; 92(22):1812–1823.
6. Fraser GE, Beeson WL, Phillips RL. Diet and lung cancer in California Seventh-day Adventists. *Am J Epidemiol*. 1991; 133(7):683–693.
7. Holick CN, Michaud DS, Stolzenberg-Solomon R, et al. Dietary carotenoids, serum beta-carotene, and retinol and risk of lung cancer in the Alpha-Tocopherol, Beta-Carotene cohort study. *Am J Epidemiol*. 2002;156(6):536–547.
8. Knekt P, Jarvinen R, Teppo L, et al. Role of various carotenoids in lung cancer prevention. *J Natl Cancer Inst.* 1999; 91(2):182–184.

9. Linseisen J, Rohrmann S, Miller AB, et al. Fruit and vegetable consumption and lung cancer risk: updated information from the European Prospective Investigation into Cancer and Nutrition (EPIC). *Int J Cancer*. 2007;121(5):1103–1114.
10. Miller AB, Altenburg HP, Bueno-de-Mesquita B, et al. Fruits and vegetables and lung cancer: findings from the European Prospective Investigation into Cancer and Nutrition. *Int J Cancer*. 2004;108(2):269–276.
11. Neuhouser ML, Patterson RE, Thornquist MD, et al. Fruits and vegetables are associated with lower lung cancer risk only in the placebo arm of the Beta-Carotene and Retinol Efficacy Trial (CARET). *Cancer Epidemiol Biomarkers Prev*. 2003; 12(4):350–358.
12. Ocke MC, Bueno-de-Mesquita HB, Feskens EJ, et al. Repeated measurements of vegetables, fruits, beta-carotene, and vitamins C and E in relation to lung cancer. The Zutphen Study. *Am J Epidemiol*. 1997;145(4):358–365.
13. Steinmetz KA, Potter JD, Folsom AR. Vegetables, fruit and lung cancer in the Iowa Women's Health Study. *Cancer Res*. 1993;53(3):536–543.
14. Voorrips LE, Goldbohm RA, Verhoeven DT, et al. Vegetable and fruit consumption and lung cancer risk in the Netherlands Cohort Study on diet and cancer. *Cancer Causes Control*. 2000;11(2):101–115.
15. Yong LC, Brown CC, Schatzkin A, et al. Intake of vitamins E, C, and A and risk of lung cancer. The NHANES I Epidemiologic Followup Study. *Am J Epidemiol*. 1997;146(3):231–243.
16. Smith-Warner SA, Spiegelman D, Yaun SS, et al. Fruits, vegetables and lung cancer: a pooled analysis of cohort studies. *Int J Cancer*. 2003;107(6):1001–1011.
17. Riboli E, Norat T. Epidemiologic evidence of the protective effect of fruit and vegetables on cancer risk. *Am J Clin Nutr*. 2003;78(3 suppl):559S–569S.
18. Schatzkin A, Subar AF, Thompson FE, et al. Design and serendipity in establishing a large cohort with wide dietary intake distributions: the National Institutes of Health-American Association of Retired Persons Diet and Health Study. *Am J Epidemiol*. 2001;154(12):1119–1125.
19. Schoenborn CA, Adams PF, Schiller JS. *National Health Interview Survey, 2000: Summary Health Statistics for the U.S. Population*. Darby, PA: Diane Publishing Company; 2004.
20. Subar AF, Midthune D, Kulldorff M, et al. Evaluation of alternative approaches to assign nutrient values to food groups in food frequency questionnaires. *Am J Epidemiol*. 2000;152(3): 279–286.
21. Tippett KS, Cypel YS, eds. *Design and Operation: The Continuing Survey of Food Intakes by Individuals and Diet and Health Knowledge Survey, 1994–96*. (NFS report no. 96-1). Beltsville, MD: US Department of Agriculture; 1997.
22. US Department of Agriculture. *The Food Guide Pyramid*. (Home and Garden Bulletin no. 252). Beltsville, MD: US Department of Agriculture; 1992.
23. Smith SA, Campbell DR, Elmer PJ, et al. The University of Minnesota Cancer Prevention Research Unit vegetable and fruit classification scheme (United States). *Cancer Causes Control*. 1995;6(4):292–302.
24. Thompson FE, Kipnis V, Midthune D, et al. Performance of a food-frequency questionnaire in the US NIH-AARP (National Institutes of Health-American Association of Retired Persons) Diet and Health Study. *Public Health Nutr*. 2008;11(2):183–195.
25. Fritz AG. *International Classification of Diseases for Oncology: ICD-O*. Third Edition. Geneva, Switzerland: World Health Organization; 2000.
26. Michaud DS, Midthune D, Hermansen S, et al. Comparison of cancer registry case ascertainment with SEER estimates and self-reporting in a subset of the NIH-AARP Diet and Health Study. *J Regist Manage*. 2005;32(1):70–75.
27. Willett WC. *Nutritional Epidemiology*. 2nd ed. New York, NY: Oxford University Press; 1998.
28. Liu Y, Sobue T, Otani T, et al. Vegetables, fruit consumption and risk of lung cancer among middle-aged Japanese men and women: JPHC Study. *Cancer Causes Control*. 2004;15(4): 349–357.
29. Shibata A, Paganini-Hill A, Ross RK, et al. Intake of vegetables, fruits, beta-carotene, vitamin C and vitamin supplements and cancer incidence among the elderly: a prospective study. *Br J Cancer*. 1992;66(4):673–679.
30. Patel JD. Lung cancer in women. *J Clin Oncol*. 2005;23(14): 3212–3218.
31. Bogers RP, Brug J, van Assema P, et al. Explaining fruit and vegetable consumption: the theory of planned behaviour and misconception of personal intake levels. *Appetite*. 2004;42(2): 157–166.
32. Subar AF, Kipnis V, Troiano RP, et al. Using intake biomarkers to evaluate the extent of dietary misreporting in a large sample of adults: The OPEN Study. *Am J Epidemiol*. 2003;158(1): 1–13.
33. Steinmetz KA, Potter JD. Vegetables, fruit, and cancer prevention: a review. *J Am Diet Assoc*. 1996;96(10):1027–1039.
34. Rock CL. Carotenoids: biology and treatment. *Pharmacol Ther*. 1997;75(3):185–197.
35. Lee KW, Lee HJ, Surh YJ, et al. Vitamin C and cancer chemoprevention: reappraisal. *Am J Clin Nutr*. 2003;78(6): 1074–1078.
36. The Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study Group. The effect of vitamin E and beta carotene on the incidence of lung cancer and other cancers in male smokers. *N Engl J Med*. 1994;330(15):1029–1035.
37. Omenn GS, Goodman GE, Thornquist MD, et al. Effects of a combination of beta carotene and vitamin A on lung cancer and cardiovascular disease. *N Engl J Med*. 1996;334(18): 1150–1155.
38. Krinsky NI, Mayne ST, Sies H. *Carotenoids in Health and Disease*. New York, NY: Marcel Dekker; 2004.
39. Le Marchand L. Cancer preventive effects of flavonoids—a review. *Biomed Pharmacother*. 2002;56(6):296–301.
40. Arts IC, Hollman PC. Polyphenols and disease risk in epidemiologic studies. *Am J Clin Nutr*. 2005;81(1 suppl):317S–325S.
41. Dallongeville J, Marecaux N, Fruchart JC, et al. Cigarette smoking is associated with unhealthy patterns of nutrient intake: a meta-analysis. *J Nutr*. 1998;128(9):1450–1457.
42. Stram DO, Huberman M, Wu AH. Is residual confounding a reasonable explanation for the apparent protective effects of beta-carotene found in epidemiologic studies of lung cancer in smokers? *Am J Epidemiol*. 2002;155(7):622–628.
43. Schatzkin A, Kipnis V. Could exposure assessment problems give us wrong answers to nutrition and cancer questions? *J Natl Cancer Inst*. 2004;96(21):1564–1565.

APPENDIX

Food Group Constituents

Total fruits: apples, apple sauce, pears, bananas, dried fruit excluding apricots, peaches, nectarines, plums, cantaloupe, other melons, strawberries, oranges, tangerines, tangelos, grapefruit, grapes, orange and grapefruit juice, and other fruit juices and drinks

Total vegetables (no white potatoes): spinach, turnips, collard greens, mustard, kale, coleslaw, cabbage, sauerkraut, carrots, string beans, green beans, peas, corn, broccoli, cauliflower, brussels sprouts, mixed vegetables, tomatoes, sweet peppers, lettuce salad, sweet potatoes, yams, tomato juice, tomato sauce, chili, and salsa

Musaceae: bananas

Cucurbitaceae: cantaloupe, watermelon, and honeydew melon

Vitaceae: grapes

Rosaceae: apples, peaches, nectarines, plums, pears, and strawberries

Rutaceae: oranges, tangerines, tangelos, and grapefruit

Chenopodiaceae: raw spinach and cooked spinach

Gramineae: corn

Cruciferae: broccoli, cauliflower, brussels sprouts, turnips, cabbage, coleslaw, collard greens, mustard greens, and kale

Legumes: dried beans, string beans, and peas

Compositae: lettuce

Solanaceae: tomatoes and peppers

Convolvulaceae: sweet potatoes and yams

Umbelliferae: carrots