ARTICLE

Check for updates

Diet quality and lung cancer incidence in a low-income population in the United States

Heather M. Munro[1], Danxia Yu[2], Wei Zheng[2], William J. Blot[2,3], Qiuyin Cai[2] and Martha J. Shrubsole[1,2]✉

BACKGROUND: Although tobacco smoking is the leading cause of lung cancer, interest in the relationship of diet quality on risk has been growing.
METHODS: We examined the association between Healthy Eating Index-2010 (HEI-10) at enrollment and lung cancer incidence among 70,802 participants in a predominantly African American and low-income prospective cohort in the southern United States. Outcomes were ascertained through linkages with state cancer registries and the National Death Index (NDI). Hazard ratios by HEI-10 quartiles were assessed using Cox proportional hazard models adjusted for potential confounders.
RESULTS: During ≤16 years of follow-up, 1454 incident lung cancers were identified. The lowest HEI-10 quartile compared to the highest was adversely associated with lung cancer risk (HR: 1.89, 95% CI 1.16–3.07) among male former smokers and female never smokers (HR: 2.58, 95% CI 1.06–6.28).
CONCLUSIONS: Low-quality diet was associated with increased lung cancer risk among male former smokers and female never smokers but cautious interpretation of the findings should be taken due to the small number of lung cancers among never smokers and the possibility of residual confounding by smoking in ever smokers.

British Journal of Cancer (2023) 129:626–635; https://doi.org/10.1038/s41416-023-02342-7

INTRODUCTION

Lung cancer accounts for about 23% of cancer deaths in the United States and ranks second among the most common types of cancer diagnosed in males and females [1, 2]. While smoking shows one of the strongest associations with lung cancer, there is some evidence that diet may also contribute to lung cancer risk [3–6]. There is limited evidence suggesting that red meat, processed meat, and alcohol may increase the risk of lung cancer and limited evidence suggesting fruits, vegetables and foods containing beta-carotene, carotenoids, vitamin C and isoflavones might decrease the risk of lung cancer [6]. In addition, there is growing recognition that multiple components of diet may contribute simultaneously to disease risk and prevention. Thus, several dietary pattern scores have been created to assess overall diet quality using foods, food groups, and nutrient intake information [7, 8], such as the Healthy Eating Index (HEI) [9–11].

Although there is substantial interest in dietary patterns and their role in cancer, there are very limited data within African American individuals or individuals with low socioeconomic status (SES), who are at elevated risk of lung cancer [12] but often underrepresented and underserved and whose diets may substantially vary from the general US population. The Southern Community Cohort Study (SCCS) is a large predominantly African American prospective cohort of mostly low SES participants living in the southeastern United States. Low SES, residing in the US South and being a member of a Black racial group have been associated with lower diet quality and higher cancer incidence [13–16]. We previously found that HEI was inversely associated with all-cause mortality, cardiovascular disease mortality, and overall cancer mortality in the SCCS [17]. In the analyses herein, we examine associations between HEI-2010 and lung cancer incidence, as well as the relationship of the HEI-10 with biomarkers related to cancer risk that we hypothesized were related to dietary quality.

METHODS

The SCCS is a prospective cohort study designed to investigate determinants for health disparities between African Americans and Whites in underserved populations. A total of 84,508 adults between the ages of 40 and 79 were enrolled during 2002–2009 from the following 12 Southeastern states: Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, Virginia, and West Virginia. Participants were enrolled either in person at one of 71 community health centers (CHC, 85%) or by mailings to stratified random samples of the general populations (GP) of these states [18, 19]. Structured in-person interviews were conducted by trained staff at CHCs to obtain detailed information on the participants' demographic, socioeconomic, and anthropometric characteristics, personal and family medical history, and lifestyle choices such as alcohol use and diet. GP participants provided the same information by completing a self-administered questionnaire. Periodic direct contact with participants and linkages with the National Death Index (NDI) and the Social Security Administration's Service for Epidemiologic Research were used to determine vital status. Cancer

[1]International Epidemiology Field Station, Vanderbilt University Medical Center, Nashville, TN, USA. [2]Department of Medicine, Division of Epidemiology, Vanderbilt Epidemiology Center, Vanderbilt University, Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center, Nashville, TN, USA. [3]International Epidemiology Institute, Rockville, MD, USA. ✉email: martha.shrubsole@vanderbilt.edu

Received: 12 January 2022 Revised: 15 June 2023 Accepted: 20 June 2023
Published online: 3 July 2023

incidence was ascertained through linkages with cancer registries from the 12 Southeastern states. The study was approved by the institutional review boards of Vanderbilt University and the Meharry Medical College, and participants gave written informed consent at study enrollment.

Participants completed a food frequency questionnaire (FFQ) containing 89 food and 5 alcohol questions, which asked about the frequency of consumption of certain types of foods and alcohol over the last year [20, 21]. Data for participants ages 30–84 living in the South census region from the National Health and Nutrition Examination Survey (NHANES) and the USDA Continuing Survey of Food Intakes by Individuals (CSFII) were used to estimate portion size for each food item after a pilot study determined that collecting individual portion size data was not necessary [22]. The FFQ was validated by comparing the amounts of select nutrients measured from biosamples among a subset of the cohort to those nutrient amounts estimated from the scored FFQ [23].

The HEI-10 was calculated for participants who completed at least 79 of the 89 FFQ questions and had energy intake between 600 and 8000 kcal/day ($N = 77{,}822$) [11, 17]. The HEI-10 was calculated by linking the FFQ data with the MyPyramid Equivalents Database (MPED, version 2.0) [24] to determine equivalent intake in cups or ounces per 1000 kcal for food groups in the USDA *Dietary Guidelines for Americans* (DGA) [25]. The 12 components of the HEI-10 were calculated using reference SAS code provided by the USDA [26] and HEI-2010 standards, which recommended increasing consumption of total fruit, whole fruit, total vegetables, greens and beans, whole grains, dairy, total protein foods, seafood and plant proteins, and the ratio of unsaturated to saturated fatty acids and decreasing consumption of refined grains, sodium, and calories from solid fats, excessive alcohol, and added sugars. The HEI-10 total score ranges from 0 to 100.

After excluding those who reported having had cancer other than non-melanoma skin cancer at baseline ($N = 6223$), those who died less than a month after enrollment ($N = 16$), those who were missing information on smoking status or intensity ($N = 781$), and those who were diagnosed with lung cancer less than a month after enrollment ($N = 7$), 70,795 participants were available for the analyses. The HEI-10 score was analyzed using quartiles derived from the cohort with males and females combined, similar to previous studies [27–29], and linear trends in the scores were assessed by treating the quartiles as a linear variable using the median value for each quartile.

To inform the possible mechanisms by which the HEI-10 may affect lung cancer risk, we evaluated the relation of the HEI-10 with several biomarkers measured from samples collected at enrollment in a subset of the cohort. Biomarkers included c-reactive protein (CRP), urinary prostaglandin E2 metabolite (PGE-M), interleukins 1β and 6 (IL-1β, IL-6), tumor necrosis factor alpha (TNF-α), and hemoglobin A1c (HbA1c). We calculated partial spearman correlation coefficients adjusted for age, sex, BMI (kg/m^2) and smoking status among cohort participants who had at least one measure of biomarkers related to cancer risk that we hypothesized to be potentially related to diet quality ($n = 6090$). Plasma high sensitivity CRP levels were measured by using a Millipore Human CRP enzyme-linked immunosorbent assay kit as described previously [30, 31]. Urinary PGE-M (11 alpha-hydroxy-9, 15-dioxo-2, 3, 4, 5-tetranor-prostane-1,20-dioic acid) level was measured using a liquid chromatography/tandem mass spectrometric method described previously [32]. Plasma levels of IL-1β, IL-6, and TNFα were measured by using MILLIPLEX MAP Human High Sensitivity T Cell Panel Premixed 21-plex - Immunology Multiplex Assay kit on Luminex 200TM Analyzer following the manufacturer's protocols. HbA1c level was measured using the Bio-Rad Variant II Hemoglobin Testing System (HPLC) with the 270–2101 NU kit in the Clinical Chemistry laboratories at the Vanderbilt University Medical Center following the manufacturer's protocol.

Chi-squared tests and ANOVA were used to test the univariate associations between HEI-10 and baseline characteristics. Cox proportional hazards models were used to test the associations between HEI-10 and incident lung cancer using a cohort study design following participants from age (in months) at enrollment through age (in months) at lung cancer diagnosis, death, loss to follow-up ($n = 41$), or date through which cancer registry data is complete (12/31/2016). Models were adjusted for sex (male, female), self-reported race (White, African American, other race), enrollment site (CHC or GP), education (less than high school, high school graduate, some college or technical training, and college graduate or higher education), annual household income (<\$15,000, \$15,000–\$24,999, \$25,000–\$49,999, and \$50,000 or more), marital status (married, not married), health insurance coverage (yes, no), BMI (<18.5, 18.5–24.9, 25–29.9, 30+ kg/m^2), smoking status and intensity (current smoker of 20 or more cigarettes per day, current smoker of 10–19 cigarettes per day, current smoker of less than 10 cigarettes per day, former smoker of 20 or more cigarettes per day, former smoker of 10–19 cigarettes per day, former smoker of less than 10 cigarettes per day, never smoker), history of diabetes (yes, no), heart attack (yes, no), stroke (yes, no), hypertension (yes, no), hypercholesterolemia (yes, no), COPD (yes, no) and HIV/AIDS (yes, no), total physical activity MET-hours per day (none, tertiles among those reporting activity), total hours spent sitting, total energy intake (kcal/day), and menopause status (postmenopausal, premenopausal) and having ever used hormone replacement therapy (yes, no) among females. Sex-stratified analyses were conducted to account for differences in dietary patterns between males and females. Additional analyses were conducted stratified by sex and race (African Americans and Whites; there were too few in the other race category for separate analysis) and by sex and smoking status (current, former, and never smokers and never smokers combined with former smokers who quit more than 15 years before enrollment; there were only 27 incident lung cancers in male never smokers resulting in unstable estimates). Sensitivity analyses excluding the first 2 years of follow-up were conducted to assess the potential for reverse causation. Analyses examining the association of each individual component of the HEI-10 with lung cancer incidence were also conducted by sex. Missing covariate data were imputed using multiple imputation with five imputations. Sensitivity analyses using participants without missing data for the covariates included in the models were performed producing very similar results. Likelihood ratio tests comparing a model with a linear term for the HEI-10 and a model with splines with four knots for the HEI-10 were used to test for non-linearity. There was no evidence for non-linearity of the HEI-10 in the overall or stratified models. Likelihood ratio tests comparing models with and without interaction terms were used to test for interactions between the HEI-10 and sex, race, and smoking status and intensity. The proportional hazards assumption was tested by including an interaction between the HEI-10 quartiles and age. The assumption held for all models among males, former smokers and never smokers. Additional models among female current smokers were stratified by age at enrollment to account for the change in HEI-10 estimates by age. All analyses use two-sided tests and were performed using SAS software, version 9.4 (SAS Institute, Inc, Cary, NC).

RESULTS

Participants 40–79 years old (mean: 51.9, SD: 8.6) at enrollment were followed on average for 10.6 years (range: 0.1–14.8). There were 1454 incident lung cancers ($N = 757$ males and $N = 697$ females). Average HEI-10 scores were lower for participants who developed lung cancer (mean: 55.1, SD: 11.5) compared to participants who did not develop lung cancer (mean: 57.7, SD: 12.0). In univariate analyses, the HEI-10 distribution was significantly different by sex and all covariates considered were associated with HEI-10 quartiles among both males and females except for total hours spent sitting and stroke which were not associated with HEI-10 in males (Table 1). Higher educational attainment, greater household income, being married, having health insurance, having a higher BMI, more time spent participating in physical activity, being postmenopausal, ever having hormone replacement therapy, never or former smoking, and having diabetes, a heart attack, hypertension, or high cholesterol were associated with higher quality diets while being enrolled at a CHC, having a higher caloric intake, or having COPD or HIV/AIDS were associated with lower-quality diets.

The lowest HEI-10 quartile compared to the highest was adversely related to lung cancer incidence among men but not women (Table 2), although the HEI-10 by sex interaction was not significant. Among males, there was a 35% increase in lung cancer risk (95% CI: 1.05–1.75, $P = 0.02$) for those with the lowest quality diet compared with the highest quality diet, whereas the corresponding HR among females was 1.06 (0.85–1.33). Associations between HEI-10 and lung cancer were observed for the whole cohort and for men and women separately in models without adjustment for smoking (Supplemental Table 1) but the

Table 1. Baseline characteristics by sex and quartiles of HEI-10 scores in the Southern Community Cohort Study.

Characteristic	Males ($N = 28,951$)				Females ($N = 41,844$)			
	Quartile 1 (low; $N = 9133$)	Quartile 2 ($N = 8073$)	Quartile 3 ($N = 6789$)	Quartile 4 (high; $N = 4956$)	Quartile 1 (low; $N = 8565$)	Quartile 2 ($N = 9627$)	Quartile 3 ($N = 10,909$)	Quartile 4 (high, $N = 12,743$)
HEI-10 score								
Median	43.6	53.2	61.2	71.4	43.9	53.5	61.4	72.5
Range	13.7–49.0	49.0–57.3	57.3–65.9	65.9–95.2	16.9–49.0	49.0–57.3	57.3–65.9	65.9–96.7
Age at enrollment (years), mean (SD)	50.0 (7.6)	50.7 (7.9)	52.1 (8.3)	54.7 (9.0)	49.3 (7.6)	51.0 (8.3)	52.3 (8.7)	55.0 (9.2)
Race, *N* (%)								
Non-Hispanic White	2760 (30.2)	2059 (25.5)	1701 (25.1)	1531 (30.9)	2911 (34.0)	2726 (28.3)	2935 (26.9)	3621 (28.4)
Non-Hispanic African American	6090 (66.7)	5709 (70.7)	4800 (70.7)	3123 (63.0)	5342 (62.4)	6531 (67.8)	7495 (68.7)	8456 (66.4)
All other	283 (3.1)	305 (3.8)	288 (4.2)	302 (6.1)	312 (3.6)	370 (3.8)	479 (4.4)	666 (5.2)
Education, *N* (%)								
<High school	3183 (35.0)	2572 (32.0)	1846 (27.3)	1041 (21.2)	2861 (33.5)	3012 (31.4)	2975 (27.3)	2641 (20.8)
High school	3420 (37.6)	2912 (36.2)	2226 (32.9)	1228 (25.0)	3163 (37.0)	3260 (33.9)	3527 (32.4)	3666 (28.9)
Some college or training	1874 (20.6)	1778 (22.1)	1723 (25.5)	1327 (27.0)	1960 (22.9)	2426 (25.3)	2997 (27.5)	3879 (30.5)
College graduate or more	629 (6.9)	776 (9.7)	962 (14.2)	1315 (26.8)	562 (6.6)	905 (9.4)	1382 (12.7)	2512 (19.8)
Missing, *N*	27	35	32	45	19	24	28	45
Income, *N* (%)								
<$15,000	5627 (62.1)	4623 (57.8)	3535 (52.7)	1880 (38.5)	5338 (63.0)	5686 (59.8)	5998 (55.7)	5924 (47.2)
$15,000–$24,999	1846 (20.4)	1718 (21.5)	1326 (19.8)	925 (19.0)	1882 (22.2)	2101 (22.1)	2383 (22.1)	2810 (22.4)
$25,000–$49,999	1071 (11.8)	1016 (12.7)	1012 (15.1)	906 (18.6)	872 (10.3)	1180 (12.4)	1560 (14.5)	2227 (17.8)
$50,000 or more	515 (5.7)	636 (8.0)	829 (12.4)	1166 (23.9)	376 (4.4)	548 (5.8)	822 (7.6)	1578 (12.6)
Missing, *N*	74	80	87	79	97	112	146	204
Currently married, *N* (%)	2977 (32.7)	2774 (34.5)	2694 (40.0)	2444 (50.1)	2755 (32.3)	3079 (32.1)	3631 (33.5)	4353 (34.4)
Missing, *N*	33	44	52	73	31	35	54	80
Health insurance, *N* (%)	4479 (49.1)	4182 (51.8)	4049 (59.7)	3438 (69.5)	4762 (55.7)	5726 (59.6)	6844 (62.8)	8732 (68.6)
Missing, *N*	11	6	8	9	9	13	16	13
CHC enrollment, *N* (%)	8351 (91.4)	7224 (89.5)	5736 (84.5)	3674 (74.1)	7896 (92.2)	8773 (91.1)	9802 (89.9)	10,916 (85.7)
Body mass index (BMI, kg/m^2), *N* (%)								
<18.5	153 (1.7)	114 (1.4)	45 (0.7)	25 (0.5)	151 (1.8)	148 (1.6)	133 (1.2)	80 (0.6)
18.5–24.9	3280 (36.1)	2768 (34.5)	2036 (30.2)	1212 (24.7)	1783 (21.1)	1790 (18.9)	1860 (17.3)	2183 (17.4)
25.0–29.9	3118 (34.4)	2792 (34.8)	2421 (35.9)	1931 (39.3)	2072 (24.5)	2394 (25.3)	2785 (25.9)	3375 (26.8)
30 or more	2524 (27.8)	2346 (29.3)	2239 (33.2)	1747 (35.5)	4437 (52.6)	5128 (54.2)	5982 (55.6)	6943 (55.2)
Missing, *N*	58	53	48	41	122	167	149	162
Physical activity (MET-hours/day), *N* (%)								
None	615 (6.8)	439 (5.5)	376 (5.6)	181 (3.7)	179 (2.1)	196 (2.1)	173 (1.6)	178 (1.4)
Tertile 1 (>0–12.6)	2915 (32.3)	2502 (31.4)	2146 (32.2)	1582 (32.7)	2959 (35.0)	3170 (33.4)	3574 (33.3)	4158 (33.2)
Tertile 2 (12.7–26.4)	2047 (22.7)	1958 (24.6)	1664 (24.9)	1477 (30.5)	2929 (34.6)	3380 (35.6)	3965 (36.9)	4850 (38.7)
Tertile 3 (> 26.4)	3449 (38.2)	3061 (38.5)	2486 (37.3)	1605 (33.1)	2387 (28.2)	2750 (29.0)	3032 (28.2)	3342 (26.7)
Missing, *N*	107	113	117	111	111	131	165	215

Table 1. continued

Characteristic	Males ($N = 28,951$)				Females ($N = 41,844$)			
	Quartile 1 (low; $N = 9133$)	Quartile 2 ($N = 8073$)	Quartile 3 ($N = 6789$)	Quartile 4 (high; $N = 4956$)	Quartile 1 (low; $N = 8565$)	Quartile 2 ($N = 9627$)	Quartile 3 ($N = 10,909$)	Quartile 4 (high, $N = 12,743$)
Total sitting time (hours/day), mean (SD)	9.2 (5.3)	9.2 (5.1)	9.3 (5.1)	9.2 (4.9)	9.7 (5.2)	9.5 (5.1)	9.4 (5.0)	8.9 (4.7)
Missing, *N*	79	43	39	30	80	92	122	116
Smoking status, cigarettes per day, *N* (%)								
Never	1721 (18.8)	1630 (20.2)	1760 (25.9)	1701 (34.3)	3184 (37.2)	3947 (41.0)	5109 (46.8)	6808 (53.4)
Former, <10	332 (3.6)	357 (4.4)	427 (6.3)	470 (9.5)	405 (4.7)	577 (6.0)	793 (7.3)	1317 (10.3)
Former, 10–19	360 (3.9)	411 (5.1)	424 (6.2)	454 (9.2)	280 (3.3)	452 (4.7)	555 (5.1)	835 (6.6)
Former, 20+	935 (10.2)	970 (12.0)	929 (13.7)	903 (18.2)	578 (6.7)	784 (8.1)	1028 (9.4)	1248 (9.8)
Current, <10	1413 (15.5)	1373 (17.0)	1171 (17.2)	591 (11.9)	1048 (12.2)	1255 (13.0)	1315 (12.1)	1078 (8.5)
Current, 10–19	1875 (20.5)	1625 (20.1)	1159 (17.1)	429 (8.7)	1323 (15.4)	1271 (13.2)	1133 (10.4)	844 (6.6)
Current, 20+	2497 (27.3)	1707 (21.1)	919 (13.5)	408 (8.2)	1747 (20.4)	1341 (13.9)	976 (8.9)	613 (4.8)
Energy intake (Kcal/day), mean (SD)	3197 (1555)	3278 (1627)	3077 (1614)	2588 (1340)	2411 (1344)	2400 (1353)	2273 (1309)	1915 (976)
Health history, *N* (%)								
Diabetes	1166 (12.8)	1339 (16.6)	1469 (21.7)	1338 (27.0)	1351 (15.8)	1932 (20.1)	2683 (24.6)	3736 (29.3)
Missing, *N*	4	3	5	4	3	3	2	2
Heart attack	691 (7.6)	638 (7.9)	609 (9.0)	529 (10.7)	391 (4.6)	524 (5.4)	619 (5.7)	772 (6.1)
Missing, *N*	9	3	5	4	7	6	7	9
Stroke	532 (5.8)	461 (5.7)	419 (6.2)	318 (6.4)	471 (5.5)	593 (6.2)	731 (6.7)	892 (7.0)
Missing, *N*	10	9	7	4	8	9	6	9
Hypertension	4220 (46.2)	3837 (47.5)	3562 (52.5)	2710 (54.7)	4449 (52.0)	5538 (57.5)	6463 (59.3)	7795 (61.2)
Missing, *N*	5	3	2	2	1	3	4	4
Hypercholesterolemia	2112 (23.2)	2088 (25.9)	2160 (31.9)	2024 (40.9)	2509 (29.4)	3139 (32.7)	3924 (36.0)	5430 (42.7)
Missing, *N*	36	24	9	13	17	21	23	16
COPD	712 (7.8)	535 (6.6)	402 (5.9)	238 (4.8)	987 (11.5)	1094 (11.4)	1057 (9.7)	1047 (8.2)
Missing, *N*	6	8	8	3	9	7	9	11
HIV/AIDS	231 (2.5)	217 (2.7)	152 (2.2)	90 (1.8)	106 (1.2)	126 (1.3)	95 (0.9)	69 (0.5)
Missing, *N*	4	4	0	1	3	2	1	4
Postmenopause, *N* (%)					5035 (58.8)	6076 (63.2)	7382 (67.8)	9631 (75.7)
Missing, *N*					9	13	20	18
Hormone therapy, *N* (%)					1917 (22.4)	2347 (24.5)	3074 (28.3)	4503 (35.4)
Missing, *N*					18	33	31	40

P values were computed using Chi-square tests for categorical variables and ANOVA F-tests for continuous variables. All *P* values were <0.0001 except for total sitting time ($P = 0.15$), stroke ($P = 0.32$) and HIV/AIDS ($P = 0.009$) among males and currently married ($P = 0.0008$) among females.

Table 2. Associations between HEI-10 score and lung cancer incidence by sex, the Southern Community Cohort Study.

	Overall ($N = 70{,}795$)		**Male ($N = 28{,}951$)**		**Female ($N = 41{,}844$)**	
	Cases, *N*	**HR (95% CI)**	**Cases, *N***	**HR (95% CI)**	**Cases, *N***	**HR (95% CI)**
Diet quality						
Highest quality diet (Q4)	257	1.0	84	1.0	173	1.0
Higher quality diet	341	1.09 (0.93–1.29)	158	1.16 (0.89–1.53)	183	1.07 (0.86–1.32)
Lower-quality diet	390	1.06 (0.90–1.25)	226	1.23 (0.95–1.60)	164	0.95 (0.76–1.18)
Lowest quality diet (Q1)	466	1.18 (1.00–1.39)	289	1.35 (1.05–1.75)	177	1.06 (0.85–1.33)
P trend		0.07		0.02		0.83
P interaction for sex[a]	0.55–0.57					

[a]Range of interaction *P* values for the HEI-10 and sex using likelihood ratio tests from overall models with and without the interaction terms across the five imputations.

HEI-10 scores were split into quartiles: lowest quality diet (13.72–49.04), lower-quality diet (49.04–57.32), higher quality diet (57.32–65.91), highest quality diet (reference, 65.91–96.72).

Models were adjusted for sex (male vs female in analyses containing both sexes), race (Black, other races, vs White), enrollment source (GP vs CHC), education (less than high school, high school, some college or training, vs college graduate or higher), income (<\$15,000, \$15,000–\$24,999, \$25,000–\$49,999, vs \$50,000 or more), marital status (married vs not married), health insurance coverage (yes vs no), BMI (<18.5, 25–29.9, 30+ vs 18.5–24.9 kg/m^2), smoking status and intensity (current smoker of 20 or more cigarettes per day, current smoker of 10–19 cigarettes per day, current smoker of less than 10 cigarettes per day, former smoker of 20 or more cigarettes per day, former smoker of 10–19 cigarettes per day, former smoker of less than 10 cigarettes per day, vs never smoker), diabetes (yes vs no), history of heart attack (yes vs no), history of stroke (yes vs no), hypertension (yes vs no), hypercholesterolemia (yes vs no), COPD (yes vs no), HIV/AIDS (yes vs no), total physical activity MET-hours per day (none, lowest tertile among those reporting some physical activity, middle tertile, vs highest tertile), total hours spent sitting per day, energy intake (kcal/day), and menopausal status (postmenopausal vs premenopausal) and ever use of hormone replacement therapy (yes vs no) among females.

strength of the associations was attenuated when confounding by smoking status and intensity was adequately controlled. In Table 3, the patterns were generally similar between African Americans and Whites (though somewhat stronger for Whites). Among males the association appeared stronger for former than current smokers (Table 4). Among females alone and males and females combined the association was only seen among never smokers, however the number of lung cancers among never smokers was small and the number of male never smokers with lung cancer was too small to produce reliable risk estimates. When combining never smokers with former smokers who had quit smoking more than 15 years before enrollment (Supplemental Table 2), significant adverse trends remained for both sexes combined and females alone. Hazard ratios remained higher than 1.0 for the 3 lower-quality diet quartiles among males, however, the number of lung cancers remained small and the associations were not statistically significant.

Sensitivity analyses excluding the first 2 years of follow-up showed similar patterns to analyses using the entire follow-up time suggesting reverse causation is not an issue. A significant adverse trend between lower HEI-10 and lung cancer incidence was observed among all never smokers but not among all current or all former smokers (Supplemental Table 3). Among male former smokers there was a more significant trend and a stronger association for the lowest quality diet quartile compared to the highest (HR = 2.47, 95% CI 1.41–4.33, $P = 0.002$, P trend$=0.002$). Analyses among female current smokers stratified by age at enrollment (<50 vs 50+ and <55 vs 55 +) showed an adverse association between low diet quality and lung cancer risk among participants who were younger at enrollment, however, the trend reversed for participants who enrolled at age 55 or later (Supplemental Table 4).

No clearer patterns emerged when lung cancer histological types (squamous cell carcinoma, adenocarcinoma, large cell carcinoma, or small cell lung cancer) were examined separately, with the statistical significance of the HEI-10 association among men disappearing in these sub-type analyses (data not shown). Analyses were also conducted to evaluate individual components of the HEI-10 among males and females (Supplemental Tables 5 and 6, respectively). In general, there were no associations between individual components and lung cancer risk.

Overall, there were no strong correlations between biomarker levels and HEI-10 (Table 5). The HEI-10 was negatively correlated with CRP overall and among males, females, current smokers, former smokers, and the participants who did not develop lung cancer, however all correlations were weak (<10%).

DISCUSSION

In this large study of low-income US adults, we observed some evidence of an increased risk of lung cancer incidence associated with a lower-quality diet among male former smokers and female never smokers and long-term former smokers. The strength of the associations with about an 89% difference in risk between highest and lowest quartile diet quality in male former smokers and an over twofold greater risk in never smokers with lower-quality diets, was much less than associated with smoking (where differences in risk of lung cancer mortality exceed 2000% in this population with hazard ratios exceeding 20.0 among current moderate and heavy smokers compared to never smokers [33]) and may be an over estimate of the true difference in risk due to the inability to completely eliminate all confounding by smoking, but nevertheless warrants further study because lung cancer is still the most common cause of cancer death among Americans [12]. The finding of an adverse association with lowest diet quality among male former smokers, female never smokers combined with long-term former smokers, and younger female current smokers suggests that diet quality may have a small effect on risk of lung cancer, however, the reverse in the association for older female current smokers raises caution about a causal interpretation. Aside from menopausal status and hormone replacement therapy, models among males were adjusted for the same covariates as models among females so that confounding factors should not have made a major contribution to sex differences, although sex differences in lung cancer have in the past led to speculation about hormonal and other exposure and/or response differences between men and women [34]. Interactions between sex and the other covariates included in the models were assessed and the only significant

Table 3. Association between HEI-10 score and lung cancer incidence by sex and race.

	Males				Females			
	African American ($N = 19,722$)		White ($N = 8051$)		African American ($N = 27,824$)		White ($N = 12,193$)	
	Cases, *N*	HR (95% CI)	Cases, *N*	HR (95% CI)	Cases, *N*	HR (95% CI)	Cases, *N*	HR (95% CI)
Diet quality								
Highest quality diet	57	1.0	22	1.0	101	1.0	64	1.0
Higher quality diet	114	1.11 (0.80–1.53)	38	1.26 (0.74–2.16)	107	1.02 (0.77–1.34)	68	1.10 (0.78–1.56)
Lower-quality diet	162	1.19 (0.87–1.62)	57	1.36 (0.81–2.28)	96	0.93 (0.69–1.25)	59	0.86 (0.59–1.25)
Lowest quality diet	195	1.29 (0.95–1.75)	89	1.60 (0.96–2.65)	98	1.10 (0.82–1.49)	72	0.91 (0.63–1.32)
P trend		0.07		0.05		0.66		0.42
P interaction for race[a]	0.87							

[a]Interaction *P* values were the same across the five imputations for the HEI-10 and race using likelihood ratio tests from overall models with and without the interaction terms.
HEI-10 scores were split into quartiles: lowest quality diet (13.72–49.04), lower-quality diet (49.04–57.32), higher quality diet (57.32–65.91), highest quality diet (reference, 65.91–96.72).
Models were adjusted for enrollment source (GP vs CHC), education (less than high school, high school, some college or training, vs college graduate or higher), income (<$15,000, $15,000–$24,999, $25,000–$49,999, vs $50,000 or more), marital status (married vs not married), health insurance coverage (yes vs no), BMI (< 18.5, 25–29.9, 30+ vs 18.5–24.9 kg/m^2), smoking status and intensity (current smoker of 20 or more cigarettes per day, current smoker of 10–19 cigarettes per day, current smoker of less than 10 cigarettes per day, former smoker of 20 or more cigarettes per day, former smoker of 10–19 cigarettes per day, former smoker of less than 10 cigarettes per day, vs never smoker), diabetes (yes vs no), history of heart attack (yes vs no), history of stroke (yes vs no), hypertension (yes vs no), hypercholesterolemia (yes vs no), COPD (yes vs no), HIV/AIDS (yes vs no), total physical activity MET-hours per day (none, lowest tertile among those reporting some physical activity, middle tertile, vs highest tertile), total hours spent sitting per day, energy intake (kcal/day), and menopausal status (postmenopausal vs premenopausal) and ever use of hormone replacement therapy (yes vs no) for models among females.

interaction was for sex by BMI, however, adding this interaction to the model did not change the HEI-10 quartile estimates. Although risk was examined across the same ranges of diet exposure, the distribution of the HEI-10 scores was significantly different between males and females with 30% of females consuming the highest quality diet while only 17% of males had HEI-10 scores in the highest quartile which may have contributed to the sex differences. However, when we calculated sex-specific quartiles and re-ran the analyses, the results were similar and the conclusions remained the same.

The HEI has been associated with lung cancer in some recent studies, although results are not consistent. In a large American cohort that was predominantly White and of higher socio-economic status, an inverse association was found among both former and current smokers [27]. The US Multiethnic Cohort Study (MEC) and Women's Health Initiative (WHI) also have both reported that a higher quality diet, as measured by HEIs, was inversely associated with risk of squamous cell cancers [28, 29]. Our findings were evident for lung cancer overall and not specific for squamous cell carcinoma, possibly due to sample size. One important factor that may explain some of the differences between the studies is that the HEI-10 levels in the MEC and WHI are higher on average than in the SCCS. However, similar to our study, the MEC reported an inverse association with lung cancer incidence and the WHI reported an inverse association with lung cancer mortality. Sex differences were not reported in the other studies.

High diet quality has been hypothesized to be related to cancer risk through individual and combined intakes of potentially beneficial nutrients (e.g., vitamins or fiber) and non-nutrients (e.g., polyphenols or probiotics) as well as the relatively lower intake of potentially adverse food exposures (e.g., nitrates or total and saturated fats) [5, 35–40]. These, in turn, are hypothesized to impact multiple physiological mechanisms, including inflammation and insulin resistance. In this study, we did not observe a strong relationship between the HEI-10 and multiple biomarkers previously hypothesized or observed to be associated with cancer, including lung cancer risk [41–48], which is consistent with a previous study [49]. This suggests that, for SCCS participants, the observed relationship of the HEI-10 with risks of lung cancer incidence cannot be primarily explained via these pathways. Future studies should continue to evaluate these and other relationships to better understand potential mechanisms.

The SCCS is a large cohort study, predominantly comprised of African Americans, females, and people of low socioeconomic status, funded to investigate the higher burden of disease in these groups who are often underrepresented in research, have a higher burden of disease, and have less access to healthy foods and medical treatment. Results from this study may not be generalizable to other populations. The SCCS data is primarily self-reported; therefore, misclassification error is possible. However, the FFQ and adjustment covariates such as physical activity, have been validated [18, 20, 50]. The large sample size allowed us to adjust for numerous factors related to lung cancer risk and healthy eating, however residual confounding by factors that were not measured or not included in the models cannot be ruled out.

Since smoking is a very strong risk factor for lung cancer, we attempted to control for the effects of smoking by adjusting for smoking status and intensity. However, the decreases in strength of the HEI-10 estimates observed between models without adjustment for smoking (Supplemental Table 1) and models adjusted for smoking status and intensity (Table 2) suggest some residual confounding by smoking is still likely. Effect estimates for HEI-10, a modest risk factor, could be biased if adjustment for smoking in the models was inadequate. In sensitivity analyses, we attempted to control for the residual confounding of smoking in a

Table 4. Association between HEI-10 score and lung cancer incidence by sex and smoking status.

	All current smokers (*N* = 29,111)		Male current smokers (*N* = 15,167)		Female current smokers (*N* = 13,944)	
	Cases, *N*	**HR (95% CI)**	**Cases, *N***	**HR (95% CI)**	**Cases, *N***	**HR (95% CI)**
Diet quality						
Highest quality diet	158	1.0	53	1.0	105	1.0
Higher quality diet	238	0.94 (0.77–1.15)	109	0.92 (0.66–1.29)	129	0.98 (0.75–1.27)
Lower-quality diet	295	0.89 (0.73–1.09)	177	1.00 (0.73–1.37)	118	0.80 (0.61–1.04)
Lowest quality diet	393	1.03 (0.85–1.25)	235	1.08 (0.79–1.47)	158	1.03 (0.79–1.33)
P trend		0.54		0.29		0.96
	All former smokers (*N* = 15,824)		**Male former smokers (*N* = 6972)**		**Female former smokers (*N* = 8852)**	
	Cases, *N*	**HR (95% CI)**	**Cases, *N***	**HR (95% CI)**	**Cases, *N***	**HR (95% CI)**
Diet quality						
Highest quality diet	87	1.0	30	1.0	57	1.0
Higher quality diet	74	1.09 (0.80–1.50)	40	1.39 (0.86–2.24)	34	0.95 (0.62–1.47)
Lower-quality diet	64	1.09 (0.78–1.52)	39	1.39 (0.85–2.26)	25	0.92 (0.57–1.50)
Lowest quality diet	56	1.19 (0.83–1.71)	47	1.89 (1.16–3.07)	9	0.52 (0.25–1.08)
P trend		0.36		0.01		0.15
	All never smokers (*N* = 25,860)		**Male never smokers (*N* = 6812)**		**Female never smokers (*N* = 19,048)**	
	Cases, *N*	**HR (95% CI)**	**Cases, *N***	**HR (95% CI)**	**Cases, *N***	**HR (95% CI)**
Diet quality						
Highest quality diet	12	1.0	1	1.0	11	1.0
Higher quality diet	29	3.15 (1.59–6.24)	9	–	20	2.77 (1.32–5.85)
Lower-quality diet	31	4.14 (2.08–8.21)	10	–	21	3.86 (1.83–8.18)
Lowest quality diet	17	2.70 (1.25–5.83)	7	–	10	2.58 (1.06–6.28)
P trend		0.004		–		0.006
P interaction HEI-10 x smoking status		<0.001		0.03		<0.001

P values for the interaction between the HEI-10 and smoking status were computed using likelihood ratio tests comparing models with and without interaction terms.
HEI-10 scores were split into quartiles: lowest quality diet (13.72–49.04), lower-quality diet (49.04–57.32), higher quality diet (57.32–65.91), highest quality diet (reference, 65.91–96.72).
Models were adjusted for sex (male vs female in analyses including both sexes), race (Black, other races, vs White), enrollment source (GP vs CHC), education (less than high school, high school, some college or training, vs college graduate or higher), income (<\$15,000, \$15,000–\$24,999, \$25,000–\$49,999, vs \$50,000 or more), marital status (married vs not married), health insurance coverage (yes vs no), BMI (<18.5, 25–29.9, 30+ vs 18.5–24.9 kg/m^2), smoking intensity for models among current or former smokers (20 or more cigarettes per day, 10–19 cigarettes per day, vs less than 10 cigarettes per day), diabetes (yes vs no), history of heart attack (yes vs no), history of stroke (yes vs no), hypertension (yes vs no), hypercholesterolemia (yes vs no), COPD (yes vs no), HIV/AIDS (yes vs no), total physical activity MET-hours per day (none, lowest tertile among those reporting some physical activity, middle tertile, vs highest tertile), total hours spent sitting per day, energy intake (kcal/day), and menopausal status (postmenopausal vs premenopausal) and ever use of hormone replacement therapy (yes vs no) for models among females.

similar manner to previous studies among other cohorts [27–29] by additionally adjusting for packyears in analyses containing current or former smokers and the number of years since quitting smoking in analyses restricted to former smokers. After adding adjustment for packyears and for years of cessation among former smokers, differences in the results were minimal. We also analyzed never smokers alone and never smokers combined with former smokers who had quit smoking more than 15 years before enrollment. We saw significant adverse trends between low diet quality and lung cancer risk for all never smokers and all never and long-term former smokers combined and for female never smokers separately and female never and long-term former smokers, however the number of never smokers who developed lung cancer were small so the results would need to be replicated in studies with larger numbers of never smokers who developed lung cancer to determine whether HEI-10 has an effect on lung cancer risk.

In a predominantly African American cohort with low socioeconomic status, decreased risk of lung cancer incidence was observed among male former smokers and female never and long-term former smokers with higher quality diets. It is possible that non-smokers may be able to reduce their lung cancer risk by following a healthy diet, however future studies are needed to investigate the lack of consistent results by sex and smoking status and potential biological pathways underlying diet quality and risk of lung cancer. Studies with larger numbers of never smokers who develop lung cancer are needed to adequately assess the association between diet and lung cancer and remove the potential for residual confounding by smoking status. The components of the HEI-10 did not strongly influence the results, suggesting that the quality of the diet overall played a more important role than the individual food components and that future studies should focus on overall diet quality.

Table 5. Cross-sectional partial correlations between HEI-10 score and baseline markers of inflammation, immune response, and insulin resistance among 6090 participants with at least one biomarker measured.

	C-reactive protein (CRP)	Urinary prostaglandin E2 metabolite (PGE-M)	Tumor necrosis factor alpha (TNF-α)	Interleukin 1β (IL-1β)	Interleukin 6 (IL-6)	Hemoglobin A1c (HbA1c)
All participants						
N	5922	296	1272	1272	1272	2129
Spearman correlation	−0.05	0.02	0.01	0.003	−0.03	0.03
P value	<0.0001	0.68	0.63	0.92	0.26	0.22
Males						
N	2282	168	699	699	699	348
Spearman correlation	−0.04	0.12	0.04	0.01	−0.03	0.03
P value	0.04	0.12	0.27	0.82	0.46	0.61
Females						
N	3640	128	573	573	573	1781
Spearman correlation	−0.05	−0.08	−0.03	−0.005	−0.03	0.03
P value	0.001	0.35	0.48	0.91	0.48	0.20
Current smokers						
N	2590	163	662	662	662	832
Spearman correlation	−0.05	0.10	0.01	0.03	−0.02	0.03
P value	0.01	0.21	0.76	0.51	0.62	0.44
Former smokers						
N	1388	72	296	296	296	484
Spearman correlation	−0.08	−0.06	−0.03	−0.08	−0.04	−0.02
P value	0.003	0.60	0.57	0.19	0.49	0.69
Never smokers						
N	1944	61	314	314	314	813
Spearman correlation	−0.03	−0.11	0.06	0.05	−0.02	0.05
P value	0.13	0.42	0.27	0.35	0.76	0.12
Participants who did not develop lung cancer						
N	5328	195	898	898	898	2072
Spearman correlation	−0.05	0.01	0.02	0.003	−0.05	0.02
P value	0.0002	0.87	0.51	0.93	0.15	0.27
Participants who developed lung cancer						
N	565	101	367	367	367	49
Spearman correlation	−0.01	0.05	−0.002	0.03	0.03	0.06
P value	0.81	0.61	0.96	0.53	0.63	0.70

P value for partial Spearman's rank correlations controlling for enrollment age, BMI (kg/m^2), sex (in analyses including both sexes), and smoking status (current, former, or never in analyses including all smoking statuses).

DATA AVAILABILITY

Requests for data are to be made at www.southerncommunitystudy.org.

REFERENCES

1. Centers for Disease Control and Prevention. An Update on Cancer Deaths in the United States. Atlanta, GA: US Department of Health and Human Services, Centers for Disease Control and Prevention, Division of Cancer Prevention and Control; 2021 [cited 2021 Oct 29]. Available from: https://www.cdc.gov/cancer/dcpc/research/update-on-cancer-deaths/index.htm.
2. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2021. CA Cancer J Clin. 2021;71:7–33.
3. Kushi LH, Doyle C, McCullough M, Rock CL, Demark-Wahnefried W, Bandera EV, et al. American Cancer Society guidelines on nutrition and physical activity for cancer prevention: reducing the risk of cancer with healthy food choices and physical activity. CA Cancer J Clin. 2012;62:30–67.
4. World Cancer Research Fund/American Institute for Cancer Research. Continuous Update Project: Diet, Nutrition, Physical Activity and the Prevention of Cancer. Summary of Strong Evidence [cited 2021 Oct 29]. Available from: https://www.wcrf.org/cupmatrix.

5. Wang C, Yang T, Guo X-F, Li D. The associations of fruit and vegetable intake with lung cancer risk in participants with different smoking status: a meta-analysis of prospective cohort studies. Nutrients. 2019;11:1791.
6. World Cancer Research Fund/American Institute for Cancer Research. Diet, Nutrition, Physical Activity and Cancer: a Global Perspective. Continuous Update Project Expert Report 2018. Available at https://www.dietandcancerreport.org.
7. Wirt A, Collins CE. Diet quality—what is it and does it matter? Public Health Nutr. 2009;12:2473–92.
8. Kant AK. Dietary patterns and health outcomes. J Am Diet Assoc. 2004;104:615–35.
9. Kennedy ET, Ohls J, Carlson S, Fleming K. The healthy eating index: design and applications. J Am Diet Assoc. 1995;95:1103–8.
10. Guenther PM, Reedy J, Krebs-Smith SM. Development of the healthy eating index-2005. J Am Diet Assoc. 2005;108:1896–901.
11. Guenther PM, Casavale KO, Reedy J, Kirkpatrick SI, Hiza HAB, Kuczynski KJ, et al. Update of the healthy eating index: HEI-2010. J Acad Nutr Diet. 2013;113:569–80.
12. Cancer of the lung and Bronchus - Cancer stat facts [Internet]. Cancer.gov. [cited 2021 Oct 29]. Available from: https://seer.cancer.gov/statfacts/html/lungb.html.
13. Centers for Disease Control and Prevention. Racial/ethnic disparities and geographic differences in lung cancer incidence—38 states and the District of Columbia, 1998–2006. Morbid Mortal Weekly Rep. 2010;59:1434–8.
14. U.S. Cancer Statistics Working Group. U.S. Cancer Statistics Data Visualizations Tool, based on 2020 submission data (1999-2018): U.S. Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute [released in June 2021, cited 2021 Oct 29]. www.cdc.gov/cancer/dataviz.
15. O'Connor JM, Sedghi T, Dhodapkar M, Kane MJ, Gross CP. Factors associated with cancer disparities among low-, medium-, and high-income US counties. JAMA Netw Open. 2018;1:e183146.
16. Vadiveloo M, Perraud E, Parker HW, Juul F, Parekh N. Geographic differences in the dietary quality of food purchases among participants in the nationally representative Food Acquisition and Purchase Survey (FoodAPS). Nutrients. 2019;11:1233.
17. Yu D, Sonderman J, Buchowski MS, McLaughlin JK, Shu X-O, Steinwandel M, et al. Healthy eating and risks of total and cause-specific death among low-income populations of African-Americans and other adults in the southeastern United States: a prospective cohort study. PLoS Med. 2015;12:e1001830.
18. Signorello LB, Hargreaves MK, Steinwandel MD, Zheng W, Cai Q, Schlundt DG, et al. Southern community cohort study: establishing a cohort to investigate health disparities. J Natl Med Assoc. 2005;97:972–9.
19. Signorello LB, Hargreaves MK, Blot WJ. The Southern Community Cohort Study: investigating health disparities. J Health Care Poor Underserved. 2010;21:26–37.
20. Buchowski MS, Schlundt DG, Hargreaves MK, Hankin JH, Signorello LB, Blot WJ. Development of a culturally sensitive food frequency questionnaire for use in the Southern Community Cohort Study. Cell Mol Biol (Noisy-le-grand). 2003;49:1295–304.
21. Signorello LB, Munro HM, Buchowski MS, Schlundt DG, Cohen SS, Hargreaves MK, et al. Estimating nutrient intake from a food frequency questionnaire: incorporating the elements of race and geographic region. Am J Epidemiol. 2009;170:104–11.
22. Schlundt DG, Buchowski MS, Hargreaves MK, Hankin JH, Signorello LB, Blot WJ. Separate estimates of portion size were not essential for energy and nutrient estimation: results from the Southern Community Cohort food-frequency questionnaire pilot study. Public Health Nutr. 2007;10:245–51.
23. Signorello LB, Buchowski MS, Cai Q, Munro HM, Hargreaves MK, Blot WJ. Biochemical validation of food frequency questionnaire-estimated carotenoid, alpha-tocopherol, and folate intakes among African Americans and non-Hispanic Whites in the Southern Community Cohort Study. Am J Epidemiol. 2010;171:488–97. https://doi.org/10.1093/aje/kwp402.
24. Bowman SA, Friday JE, Moshfegh AJ. MyPyramid Equivalents Database, 2.0 for USDA Survey Foods, 2003-2004 [Online]. Beltsville, MD: U.S. Department of Agriculture, ARS [2008, cited 2021 Oct 29]. www.ars.usda.gov/ba/bhnrc/fsrg.
25. U.S. Department of Agriculture and U.S. Department of Health and Human Services. Dietary Guidelines for Americans, 2010. 7th Edition, Washington, DC: U.S. Government Printing Office, December 2010 [cited 2021 Oct 29]. https://health.gov/our-work/nutrition-physical-activity/dietary-guidelines/previous-dietary-guidelines/2010.
26. Overview & background of Healthy Eating Index (HEI) [Internet]. Cancer.gov. [cited 2021 Oct 29]. https://epi.grants.cancer.gov/hei/.
27. Anic GM, Park Y, Subar AF, Schap TE, Reedy J. Index-based dietary patterns and risk of lung cancer in the NIH-AARP diet and health study. Eur J Clin Nutr. 2016;70:123–9.
28. Park S-Y, Boushey CJ, Shvetsov YB, Wirth MD, Shivappa N, Hébert JR, et al. Diet quality and risk of lung cancer in the Multiethnic Cohort Study. Nutrients. 2021;13:1614.
29. Myneni AA, Giovino GA, Millen AE, LaMonte MJ, Wactawski-Wende J, Neuhouser ML, et al. Indices of diet quality and risk of lung cancer in the Women's Health Initiative Observational Study. J Nutr. 2021;151:1618–27.
30. Wu J, Cai Q, Li H, Cai H, Gao J, Yang G, et al. Circulating C-reactive protein and colorectal cancer risk: a report from the Shanghai Men's Health Study. Carcinogenesis. 2013;34:2799–803.
31. Zhang X, Shu X-O, Signorello LB, Hargreaves MK, Cai Q, Linton MF, et al. Correlates of high serum C-reactive protein levels in a socioeconomically disadvantaged population. Dis Markers. 2008;24:351–9.
32. Murphey LJ, Williams MK, Sanchez SC, Byrne LM, Csiki I, Oates JA, et al. Quantification of the major urinary metabolite of PGE2 by a liquid chromatographic/mass spectrometric assay: determination of cyclooxygenase-specific PGE2 synthesis in healthy humans and those with lung cancer. Anal Biochem. 2004;334:266–75.
33. Munro HM, Tarone RE, Wang TJ, Blot WJ. Menthol and nonmenthol cigarette smoking: all-cause deaths, cardiovascular disease deaths, and other causes of death among blacks and whites: all-cause deaths, cardiovascular disease deaths, and other causes of death among blacks and whites. Circulation. 2016;133:1861–6.
34. Blot WJ, McLaughlin JK. Are women more susceptible to lung cancer? J Natl Cancer Inst. 2004;96:812–3.
35. Rock CL, Thomson C, Gansler T, Gapstur SM, McCullough ML, Patel AV, et al. American Cancer Society guideline for diet and physical activity for cancer prevention. CA Cancer J Clin. 2020;70:245–71.
36. Sun Y, Li Z, Li J, Li Z, Han J. A healthy dietary pattern reduces lung cancer risk: a systematic review and meta-analysis. Nutrients. 2016;8:134.
37. Hodge AM, Bassett JK, Shivappa N, Hébert JR, English DR, Giles GG, et al. Dietary inflammatory index, Mediterranean diet score, and lung cancer: a prospective study. Cancer Causes Control. 2016;27:907–17.
38. Yang JJ, Yu D, Takata Y, Smith-Warner SA, Blot W, White E, et al. Dietary fat intake and lung cancer risk: a pooled analysis. J Clin Oncol. 2017;35:3055–64.
39. Yang JJ, Yu D, Xiang Y-B, Blot W, White E, Robien K, et al. Association of dietary fiber and yogurt consumption with lung cancer risk: a pooled analysis. JAMA Oncol. 2020;6:e194107.
40. Willemsen RF, McNeil J, Heer E, Johnson ST, Friedenreich CM, Brenner DR. Dietary patterns with combined and site-specific cancer incidence in Alberta's Tomorrow Project cohort. Eur J Clin Nutr. 2021. https://doi.org/10.1038/s41430-021-00958-7.
41. Il'yasova D, Colbert LH, Harris TB, Newman AB, Bauer DC, Satterfield S, et al. Circulating levels of inflammatory markers and cancer risk in the health aging and body composition cohort. Cancer Epidemiol Biomarkers Prev. 2005;14:2413–8.
42. Chaturvedi AK, Caporaso NE, Katki HA, Wong H-L, Chatterjee N, Pine SR, et al. C-reactive protein and risk of lung cancer. J Clin Oncol. 2010;28:2719–26.
43. Shiels MS, Pfeiffer RM, Hildesheim A, Engels EA, Kemp TJ, Park J-H, et al. Circulating inflammation markers and prospective risk for lung cancer. J Natl Cancer Inst. 2013;105:1871–80.
44. Shiels MS, Katki HA, Hildesheim A, Pfeiffer RM, Engels EA, Williams M, et al. Circulating inflammation markers, risk of lung cancer, and utility for risk stratification. J Natl Cancer Inst. 2015;107. https://doi.org/10.1093/jnci/djv199.
45. Travier N, Jeffreys M, Brewer N, Wright CS, Cunningham CW, Hornell J, et al. Association between glycosylated hemoglobin and cancer risk: a New Zealand linkage study. Ann Oncol. 2007;18:1414–9.
46. Dibaba DT, Judd SE, Gilchrist SC, Cushman M, Pisu M, Safford M, et al. Association between obesity and biomarkers of inflammation and metabolism with cancer mortality in a prospective cohort study. Metabolism. 2019;94:69–76.
47. Goto A, Noda M, Sawada N, Kato M, Hidaka A, Mizoue T, et al. High hemoglobin A1c levels within the non-diabetic range are associated with the risk of all cancers: hemoglobin A1c and cancer risk. Int J Cancer. 2016;138:1741–53.
48. Rentsch CT, Farmer RE, Eastwood SV, Mathur R, Garfield V, Farmaki A-E, et al. Risk of 16 cancers across the full glycemic spectrum: a population-based cohort study using the UK Biobank. BMJ Open Diabetes Res Care. 2020;8:e001600.
49. Crowe-White KM, Ellis AC, Mehta T, Locher JL, Ard JD. Dietary quality assessed by the HEI-2010 and biomarkers of cardiometabolic disease: an exploratory analysis. J Am Coll Nutr. 2019;38:640–7.
50. Buchowski MS, Matthews CE, Cohen SS, Signorello LB, Fowke JH, Hargreaves MK, et al. Evaluation of a questionnaire to assess sedentary and active behaviors in the Southern Community Cohort Study. J Phys Act Health. 2012;9:765–75.

ACKNOWLEDGEMENTS

The authors wish to thank the Southern Community Cohort Study (SCCS) team members who have contributed to the SCCS.

AUTHOR CONTRIBUTIONS

HMM and MJS conceived the work that led to the submission. DY, WZ, WJ, and QC acquired data. All authors played an important role in interpreting the results, drafted or revised the manuscript, approved the final version and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

FUNDING

Research reported in this publication was supported by the National Cancer Institute of the National Institutes of Health under Award Numbers U01CA202979 and R01CA092447. Biomarker assays were partially supported by NIH grant P50CA090949. Biospecimens and data collection and use as well as cytokine assays were conducted by the Survey and Biospecimen Shared Resource, which is supported in part by the Vanderbilt-Ingram Cancer Center (P30 CA68485).

COMPETING INTERESTS

WZ and MJS receive research support from Pfizer, Inc which is unrelated to the contents of this manuscript. HMM, DY, WJB, and QC declare no competing interests.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

All study participants provided written informed consent. The study was approved by the Institutional Review Board at Vanderbilt University and Meharry Medical College. The study was performed in accordance with the Declaration of Helsinki.

CONSENT FOR PUBLICATION

The manuscript does not contain any individual person's data.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41416-023-02342-7.

Correspondence and requests for materials should be addressed to Martha J. Shrubsole.