



---

Other Health-related Behaviours and Cancer

## Coffee consumption and incidence of lung cancer in the NIH-AARP Diet and Health Study

Kristin A Guertin,<sup>1,\*</sup> Neal D Freedman,<sup>1</sup> Erikka Loftfield,<sup>1</sup>  
Barry I Graubard,<sup>2</sup> Neil E Caporaso<sup>3</sup> and Rashmi Sinha<sup>1</sup>

<sup>1</sup>Nutritional Epidemiology Branch, <sup>2</sup>Biostatistics Branch and <sup>3</sup>Genetic Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Department of Health and Human Services, Bethesda, MD, USA

\*Corresponding author. National Cancer Institute, NIH, 9609 Medical Center Dr. RM 6E326, MSC 9760, Bethesda, MD 20892, USA. E-mail: kristin.guertin@nih.gov

Accepted 19 May 2015

### Abstract

**Background:** Coffee drinkers had a higher risk of lung cancer in some previous studies, but as heavy coffee drinkers tend to also be cigarette smokers, such findings could be confounded. Therefore, we examined this association in the nearly half a million participants of the US NIH-AARP Diet and Health Study.

**Methods:** Typical coffee intake and smoking history were queried at baseline. During 4 155 256 person-years of follow-up, more than 9000 incident lung cancer cases occurred. We used Cox proportional hazards regression to estimate hazard ratios (HRs) and 95% confidence intervals for coffee intake and subsequent incidence of lung cancer. We also comprehensively adjusted for tobacco smoking and examined associations by detailed strata of tobacco use.

**Results:** Coffee drinkers were far more likely to smoke than non-drinkers. Although coffee drinking was associated with lung cancer in age- and sex- adjusted models (HR for  $\geq 6$  cups/day compared with none: 4.56, 4.08–5.10), this association was substantially attenuated after adjusting for smoking (HR: 1.27, 1.14–1.42). Similar findings were observed for each different histological type of lung cancer, and for participants drinking predominantly caffeinated or decaffeinated coffee. Little evidence for an association was observed in our stratified analyses, either within never smokers or in most categories of tobacco use.

**Conclusions:** Coffee drinking was positively associated with lung cancer in our study, although the association was substantially attenuated after adjustment for tobacco smoking. As our adjustment for lifetime tobacco use was imperfect, it is likely that the remaining association is due to residual confounding by smoking, although other explanations are possible.

**Key words:** Coffee, caffeine, lung cancer, smoking, tobacco

---

### Key Messages

- Data from a number of epidemiological studies have suggested that coffee drinkers may have an increased risk of lung cancer. However, nearly all of these studies were unable to comprehensively adjust for tobacco smoking, a behaviour highly correlated with coffee consumption and the main risk factor for lung cancer.
- Using data from the US NIH-AARP Diet and Health Study, we investigated the association between self-reported coffee intake (total, caffeinated and decaffeinated) and incident lung cancer with careful adjustment for tobacco use (current smoking status, number of cigarettes smoked per day, time since smoking cessation among former smokers, and whether a participant ever smoked pipes/cigars).
- Coffee drinkers were much more likely to smoke in our cohort than non-drinkers.
- We observed a positive association between coffee drinking and lung cancer that was substantially attenuated after adjustment for tobacco smoking. We observed little evidence for an association among never smokers or in finely detailed strata of cigarette smoking.
- As our assessment of lifetime smoking use was imperfect, residual confounding by tobacco use is a likely explanation for our findings, although we cannot specifically exclude the possibility of a positive association between coffee drinking and lung cancer.

## Introduction

Lung cancer is the second most common incident cancer and the leading cause of cancer death in the USA, with an estimated 221 200 new cases and 158 040 deaths in 2015.<sup>1</sup> Coffee is a commonly consumed beverage and is a significant dietary source of antioxidants and other bioactive compounds that may play a role in carcinogenesis.<sup>2</sup> Coffee consumption has recently been linked to lower overall risk of mortality and a number of different cancer types,<sup>3–7</sup> but a recent meta-analysis<sup>8</sup> indicated increased lung cancer risk with heavy coffee drinking.

However, residual confounding by cigarette smoking may be a non-causal explanation for these results.<sup>8</sup> Cigarette smoking is the main risk factor for lung cancer,<sup>9</sup> and smoking and coffee consumption are positively correlated behaviours.<sup>10–12</sup> Nevertheless, prior studies of coffee intake and lung cancer risk have lacked the very large size and case numbers necessary for examining the confounding from smoking in their analyses.<sup>10,13–28</sup>

Other questions also remain. For example, there is epidemiological evidence, although scarce, that decaffeinated coffee may be inversely related to lung cancer,<sup>14,22</sup> in contrast to results for caffeinated coffee. In addition, associations may also differ by histological subtype. Although smoking is associated with increased risk of all lung cancers, squamous cell and small-cell carcinomas account for the majority of lung cancer diagnoses among smokers.<sup>29</sup> Never smokers, in contrast, are predominantly diagnosed with adenocarcinoma.<sup>30</sup> Despite the potential aetiological differences by subtype, only lung adenocarcinoma has been examined separately to date.<sup>8</sup>

We therefore examined the association between coffee consumption and incidence of lung cancer within the large,

prospective NIH-AARP cohort; previous studies within this cohort have identified evidence of lower risk of several site-specific cancers<sup>5,6,31,32</sup> and all-cause mortality<sup>3</sup> with coffee drinking. With more than 9000 incident lung cancers, this cohort includes more cases than all previous studies combined.<sup>8</sup> The large number of cases and detailed information on smoking allowed us to both carefully adjust for tobacco use and to evaluate associations among finer strata of tobacco use. In addition, we evaluated potential differences by histological subtype and caffeinated or decaffeinated coffee.

## Methods

### Study population

The NIH-AARP Diet and Health Study [<http://dietandhealth.cancer.gov/>] has been previously described.<sup>33</sup> Briefly, 3.5 million AARP members aged 50–71 years and residing in six US states (California, Florida, Louisiana, New Jersey, North Carolina and Pennsylvania) and two metropolitan areas (Atlanta, GA, and Detroit, MI) were mailed questionnaires in 1995–96, which queried demographics, health-related behaviours and diet. AARP, formerly the American Association of Retired Persons, is a nonprofit, nonpartisan organization that provides services to its members; individuals are eligible for AARP membership if they reside in the USA and are at least 50 years of age. The study cohort included 566 398 participants who satisfactorily completed the baseline questionnaire and provided informed consent. We excluded: proxy respondents ( $n = 15\,760$ ); individuals with prevalent cancer except non-melanoma skin cancer ( $n = 49\,318$ ); with missing

information on coffee intake ( $n=2699$ ); with missing information on cigarette smoking ( $n=30\ 200$ ); with implausible or missing total energy ( $n=4193$ ), fruit ( $n=801$ ) or vegetable ( $n=1761$ ) intake; and individuals who died on or before the day their questionnaire was received ( $n=41$ ); 457 366 individuals were thus included in our analysis. This study was approved by the Special Studies Institutional Review Board at the National Cancer Institute.

### Case ascertainment

Incident lung cancer cases were identified by probabilistic linkage to state cancer registries from the original eight states and three additional states (Arizona, Texas and Nevada) to which participants were most likely to move during follow-up; further details have been published previously.<sup>34</sup> Lung cancer cases were defined according to the *International Classification of Diseases for Oncology* (ICD), third edition<sup>35</sup> codes as previously described<sup>36</sup> and included carcinomas of the bronchus and lung (ICD 34.0–34.9). Examined histological subtypes included adenocarcinoma, squamous cell carcinoma, undifferentiated carcinoma or large-cell carcinoma, small-cell carcinoma and lung cancers not otherwise specified.

### Coffee assessment

At baseline, participants completed a 124-item Food Frequency Questionnaire (FFQ)<sup>37</sup> that queried typical diet, including coffee intake, over the past year. Participants reported the number of cups of coffee they consumed over the past year using 10 categories, ranging from none to  $\geq 6$  cups/day, and whether they drank caffeinated or decaffeinated coffee more than half of the time. We used these data to categorize participants into six categories of usual total coffee consumption over the past year: none (never drank coffee during the past year),  $< 1$ , 1, 2–3, 4–5, or  $\geq 6$  cups/day. The same categories were created for caffeinated and decaffeinated coffee intake, and an indicator variable was created for individuals missing information on caffeine type. In stratified analyses, we further collapsed coffee intake into four categories: none,  $\leq 1$ , 2–3 or  $\geq 4$  cups/day. Using a previously reported method,<sup>31</sup> we performed a calibration analysis where we used data from a subset of participants with two 24-h dietary recalls ( $n=1686$ ) to correct daily coffee intake estimates in the entire cohort for measurement error ([Supplemental Methods](#), available as [Supplementary data](#) at *IJE* online). Lung cancer hazard ratio (HR) estimates for total coffee intake using measurement-error corrected intakes were similar in magnitude to those found with uncorrected estimates ([Supplemental](#)

[Table 1](#), available as [Supplementary data](#) at *IJE* online); thus, we focus on the uncorrected estimates.

### Covariate assessment

All covariates were assessed at baseline; the main covariates were age, sex and smoking; details regarding the distribution of covariates in our study and the statistical adjustments for these covariates are provided in [Table 1](#) and [Table 2](#), respectively. Briefly, our assessment of tobacco smoking included current cigarette smoking status, number of cigarettes smoked per day among current and former smokers, time since smoking cessation among former smokers, and whether a participant ever smoked pipes/cigars. Individuals who reported having quit smoking within the past year were categorized as current smokers.

### Statistical analysis

We used Cox proportional hazards regression to estimate HRs and 95% confidence intervals (CIs) for coffee intake with total lung cancer, with non-drinkers as the referent group and person-years as the underlying time metric; non-drinkers were defined as participants who reported drinking no coffee over the past year. Person-years were calculated beginning on the date of questionnaire return until cancer diagnosis, movement out of the registry area, loss to follow-up, death or the end of follow-up (31 December 2006), whichever came first. Histological subtypes of lung cancer (adenocarcinoma, squamous cell carcinoma, undifferentiated carcinoma and small-cell carcinoma) were considered in separate models. The midpoint of each coffee intake category was used to calculate *P*-values for trends. We compared models that were adjusted for age and sex with those additionally adjusted for tobacco smoking and other covariates (see [Table 2](#) for details); primary analyses focused on models adjusted for age, sex and tobacco smoking. Analyses were conducted using SAS 9.3 (SAS Institute, Cary, NC). All tests of statistical significance were two-sided.

To better understand how residual confounding may affect observed hazard estimates, we conducted analyses stratified by detailed smoking categories, including current cigarette smoking status, number of cigarettes smoked per day among current and former smokers, and time since smoking cessation among former smokers.

In secondary analyses, we calculated the HRs for each type of coffee intake, namely caffeinated and decaffeinated coffee, within one model; participants who reported drinking coffee but who were missing caffeine type were included in the model using an indicator variable. Additionally, we conducted analyses stratified by gender.

**Table 1.** Baseline characteristics of the NIH-AARP Diet and Health Study by coffee intake ( $N = 457\ 366$ )<sup>a</sup>

Characteristic	Coffee intake						Total
	None	<1 cup/day	1 cup/day	2–3 cups/day	4–5 cups/day	≥ 6 cups/day	
No. of subjects	46 369	74 796	75 383	188 205	55 503	17 110	457 366
No. of lung cancers (%)	510 (1)	987 (1)	1122 (1)	4022 (2)	1746 (3)	809 (5)	9196 (2)
Caffeinated coffee, % <sup>b</sup>	0	42	58	69	75	79	57
Decaffeinated coffee, % <sup>b</sup>	0	51	38	28	22	18	30
Age (years), mean (SD)	61.2 (5.5)	61.9 (5.4)	62.7 (5.3)	62.2 (5.3)	61.4 (5.3)	60.9 (5.4)	62.0 (5.4)
Men, %	53	57	56	62	67	68	60
Non-Hispanic White, %	90	86	89	94	96	96	92
Family history of cancer, %	49	48	48	49	50	50	49
Currently married, %	65	65	68	71	72	69	69
Education, %							
≤ High school	33	33	37	35	36	41	35
≥ College graduate	42	41	38	38	38	32	41
Body mass index (kg/m <sup>2</sup> ), mean (SD)	27.3 (5.7)	27.3 (5.4)	27.1 (5.2)	27.0 (4.8)	27.0 (4.8)	27.0 (5.2)	27.1 (5.1)
Emphysema, %	2	2	2	3	3	5	3
Diabetes, %	10	10	10	8	8	8	9
Cigarette smoking status, % <sup>c</sup>							
Never <sup>d</sup>	55	42	37	27	18	11	32
Former	35	47	50	55	52	44	50
Current	7	8	9	15	26	42	14
Cigarettes smoked per day, % ( $n = 294\ 574$ ) <sup>e</sup>							
≤ 10	31	33	32	25	17	12	25
11–20	31	32	33	34	32	30	33
21–40	29	27	28	33	38	42	32
>40	9	9	8	9	12	16	10
Years since quitting smoking, % ( $n = 228\ 875$ ) <sup>f</sup>							
≥ 10	81	80	79	76	72	66	77
5–9	12	13	14	15	17	19	15
≥ 1–4	7	7	7	9	11	15	9
Ever smokers of pipes/cigars, %	10	14	15	19	23	25	18
Alcohol, > 3 drinks/day, %	4	6	7	9	10	9	8
Vigorous physical activity ≥ 5/week, %	21	19	19	19	19	19	19
Poor/fair reported health, %	13	14	13	11	11	13	12
Total energy (kcal/day), mean (SD)	1793.5 (805.9)	1750.2 (785.3)	1767.4 (770.3)	1841.3 (781.5)	1986.8 (845.4)	2176.0 (960.7)	1839.6 (804.0)

NIH-AARP, National Institutes of Health-AARP; SD, standard deviation.

<sup>a</sup>275 328 men and 182 038 women; all values shown are percentages unless otherwise noted; some percentages do not sum to 100 due to missing data.

<sup>b</sup>Type of coffee consumed (caffeinated or decaffeinated) among coffee drinkers was based on which type of coffee the participant reported drinking more than half the time. Due to missing data on caffeine type among coffee drinkers, caffeinated and decaffeinated coffee consumption does not sum to 100 for some categories of coffee consumption.

<sup>c</sup>Individuals who smoked cigars but not cigarettes contribute to denominator but not numerator.

<sup>d</sup>Never smokers of any tobacco products (cigarettes, pipes, cigars).

<sup>e</sup>Current and former smokers.

<sup>f</sup>Former smokers.

Lastly, we assessed the proportional hazard assumption by testing for an interaction between coffee intake and person-years for total lung cancer.

## Results

Over the course of 4 155 256 person-years of follow-up, 9196 incident lung cancer cases were identified.

Descriptively, the majority of lung cancers were diagnosed in current (46%) and former (49%) smokers (data not shown). Of the 6507 cases with specified histology, approximately half (46%,  $n = 3022$ ) were adenocarcinomas.

The mean age at baseline was 62 years and the median follow-up time was 10.5 years. Most participants reported drinking coffee; 75% of participants drank at least 1 cup/day (Table 2) and the median coffee intake among

**Table 2.** Hazard ratios (95% confidence intervals) for lung cancer according to coffee intake in the NIH-AARP Diet and Health Study ( $N = 457\ 366$ )

Model adjustments	Coffee intake						P-trend
	None (ref.)	< 1 cup/day	1 cup/day	2–3 cups/day	4–5 cups/day	≥ 6 cups/day	
<b>Lung cancer (all)</b>							
No. of cases	510	987	1122	4022	1746	809	
Age, sex	1.00	1.15 (1.03-1.28)	1.22 (1.10-1.36)	1.82 (1.66-2.00)	2.85 (2.58-3.15)	4.56 (4.08-5.10)	<0.0001
Age, sex, smoking	1.00	1.00 (0.90-1.11)	0.97 (0.88-1.08)	1.06 (0.97-1.17)	1.14 (1.03-1.26)	1.27 (1.14-1.42)	<0.0001
Multivariate-adjusted	1.00	1.01 (0.91-1.12)	0.99 (0.89-1.10)	1.10 (1.00-1.20)	1.18 (1.07-1.31)	1.29 (1.15-1.45)	<0.0001
<b>Adenocarcinoma</b>							
No. of cases	166	361	391	1327	548	229	
Age, sex	1.00	1.30 (1.09-1.57)	1.33 (1.11-1.59)	1.88 (1.60-2.21)	2.78 (2.33-3.30)	3.99 (3.27-4.88)	<0.0001
Age, sex, smoking	1.00	1.14 (0.95-1.37)	1.08 (0.90-1.29)	1.16 (0.99-1.37)	1.26 (1.05-1.50)	1.33 (1.09-1.64)	0.0013
Multivariate-adjusted	1.00	1.14 (0.94-1.37)	1.07 (0.89-1.29)	1.17 (1.00-1.40)	1.28 (1.07-1.52)	1.35 (1.10-1.65)	0.0005
<b>Squamous cell carcinoma</b>							
No. of cases	98	149	196	732	323	175	
Age, sex	1.00	0.88 (0.68-1.14)	1.08 (0.85-1.38)	1.65 (1.34-2.04)	2.60 (2.07-3.26)	4.88 (3.81-6.25)	<0.0001
Age, sex, smoking	1.00	0.73 (0.57-0.94)	0.80 (0.63-1.02)	0.85 (0.69-1.05)	0.88 (0.70-1.11)	1.11 (0.86-1.42)	0.0070
Multivariate-adjusted	1.00	0.74 (0.57-0.95)	0.82 (0.64-1.04)	0.90 (0.73-1.11)	0.94 (0.74-1.18)	1.14 (0.88-1.47)	0.0024
<b>Undifferentiated</b>							
No. of cases	28	47	55	206	94	50	
Age, sex	1.00	0.99 (0.62-1.58)	1.08 (0.69-1.71)	1.68 (1.13-2.50)	2.75 (1.80-4.20)	5.04 (3.17-8.01)	<0.0001
Age, sex, smoking	1.00	0.84 (0.53-1.34)	0.84 (0.53-1.32)	0.95 (0.64-1.41)	1.07 (0.70-1.64)	1.39 (0.87-2.23)	0.0124
Multivariate-adjusted	1.00	0.88 (0.55-1.40)	0.88 (0.56-1.40)	1.01 (0.68-1.50)	1.12 (0.73-1.73)	1.44 (0.89-2.31)	0.0130
<b>Small cell carcinoma</b>							
No. of cases	64	122	136	586	293	131	
Age, sex	1.00	1.15 (0.85-1.55)	1.20 (0.89-1.62)	2.17 (1.67-2.80)	3.89 (2.96-5.10)	5.98 (4.43-8.06)	<0.0001
Age, sex, smoking	1.00	1.01 (0.75-1.37)	0.96 (0.71-1.29)	1.14 (0.88-1.48)	1.27 (0.97-1.67)	1.26 (0.93-1.71)	0.0020
Multivariate-adjusted	1.00	1.02 (0.75-1.38)	0.97 (0.72-1.31)	1.18 (0.91-1.53)	1.32 (1.00-1.73)	1.27 (0.93-1.72)	0.0016

Referent group was non-drinkers of coffee. All models were adjusted for age at study baseline (continuous) and sex. Smoking adjustment included current cigarette smoking status (current, former, never), number of cigarettes smoked per day (1–10, 11–20, 21–30, 31–40, 41–60, ≥ 60), time of smoking cessation among former smokers (< 1 year, 1 to < 5 years, 5 to < 10 years, or ≥ 10 years before baseline) and whether a participant ever smoked pipe/cigars (yes/no). Additional covariates in the multivariate model included race/ethnicity (non-Hispanic White, non-Hispanic Black, other), body mass index (BMI; as < 18.5, 18.5 to < 25, 25 to < 30, and ≥ 30 kg/m<sup>2</sup>), level of education (≤ 11 years, high school graduate, some college, college graduate), alcohol consumption (0, ≤ 1, 2–3 or > 3 drinks/day), health status (good/excellent, good, poor/fair), total energy intake (kcal, continuous), nutrient density-adjusted fruit intake (continuous), nutrient density-adjusted vegetable intake (continuous), supplement use (yes/no), current marital status (married/not married), physical activity (never/rarely, 1–3 x month, 1–2 x week, 3–4 x week or ≥ 5 x week), history of cardiovascular disease (yes/no), diabetes (yes/no) and family history of cancer (yes/no).

consumers was 2.5 cups/day. The majority of coffee drinkers (63%) drank predominantly caffeinated coffee (data not shown). Coffee drinkers were more likely to be men and to have a lower level of education.

Over half of the participants reported a history of cigarette smoking (50% former, 14% current smokers), and current and former smokers were more likely to drink coffee; 33% ( $n = 21\ 750$ ) of current smokers drank ≥ 4 cups of coffee per day, compared with 8% ( $n = 11\ 790$ ) of never smokers (data not shown). Ever use of pipes or cigars was also more common among participants with higher coffee intakes. Individuals who reported greater smoking intensities (number of cigarettes smoked per day) also reported greater consumption of coffee (Supplemental Table 2, available as Supplementary data at *IJE* online). For example, consumption of at least 6 cups of coffee per day

was reported by 24% of current smokers with a smoking intensity of > 40 cigarettes per day and, in contrast, only 5% of current smokers with a smoking intensity of 1–10 cigarettes per day. This association was similar, but weaker, among former smokers. In linear regression models, individuals who smoked a greater number of cigarettes per day also reported higher coffee consumption ( $P < 0.0001$ ; data not shown); this association persisted after adjustment for age and sex.

In models adjusted for age and sex, coffee intake was positively associated with lung cancer; as compared with non-drinkers, those who consumed ≥ 6 cups/day had the highest hazard (HR = 4.56; 95% CI = 4.08-5.10) (Table 2). After further adjustment for smoking, however, the association was substantially attenuated (comparing ≥ 6 cups/day with 0, HR = 1.27; 95% CI = 1.14-1.42).

**Table 3.** Coffee type (caffeinated, decaffeinated): hazard ratios and 95% confidence intervals for lung cancer according to coffee intake in the NIH-AARP Diet and Health Study ( $N = 457\,366$ )

		Caffeinated coffee intake				P-trend
		None (ref.)	≤ 1 cup/day	2–3 cups/day	≥ 4 cups/day	
Lung cancer (all)	No. of cases	510	1059	2839	2004	–
	HR (95% CI)	1.00	0.93 (0.83-1.03)	1.06 (0.96-1.17)	1.18 (1.07-1.31)	<0.0001
Adenocarcinoma	No. of cases	166	376	928	612	–
	HR (95% CI)	1.00	1.05 (0.88-1.26)	1.16 (0.98-1.37)	1.30 (1.09-1.55)	0.0004
Squamous cell carcinoma	No. of cases	98	185	509	393	–
	HR (95% CI)	1.00	0.76 (0.59-0.97)	0.94 (0.67-1.04)	0.96 (0.76-1.20)	0.0130
Undifferentiated carcinoma	No. of cases	28	48	140	122	–
	HR (95% CI)	1.00	0.74 (0.46-1.17)	0.91 (0.60-1.37)	1.26 (0.83-1.91)	0.0165
Small-cell carcinoma	No. of cases	64	125	423	323	–
	HR (95% CI)	1.00	0.87 (0.64-1.17)	1.15 (0.88-1.49)	1.21 (0.92-1.59)	0.0050
		Decaffeinated coffee intake				
		None (ref.)	≤ 1 cup/day	2–3 cups/day	≥ 4 cups/day	P-trend
Lung cancer (all)	No. of cases	510	903	1027	454	–
	HR (95% CI)	1.00	1.02 (0.92-1.14)	1.04 (0.94-1.16)	1.13 (1.00-1.29)	0.0003
Adenocarcinoma	No. of cases	166	324	351	140	–
	HR (95% CI)	1.00	1.14 (0.94-1.37)	1.16 (0.96-1.40)	1.20 (0.96-1.50)	0.0583
Squamous cell carcinoma	No. of cases	98	127	197	86	–
	HR (95% CI)	1.00	0.71 (0.54-0.92)	0.90 (0.70-1.14)	0.90 (0.67-1.20)	0.0866
Undifferentiated carcinoma	No. of cases	28	48	57	14	–
	HR (95% CI)	1.00	0.97 (0.60-1.54)	1.01 (0.64-1.54)	0.61 (0.32-1.15)	0.9539
Small-cell carcinoma	No. of cases	64	116	145	86	–
	HR (95% CI)	1.00	1.11 (0.82-1.51)	1.15 (0.86-1.55)	1.52 (1.10-2.10)	0.0004

Caffeinated (cups/day), decaffeinated (cups/day), and missing coffee intake (yes/no) were in one model adjusted for age, sex and smoking. The referent group was non-drinkers of coffee (reported no coffee intake of any kind over the past year). Type of coffee consumed (caffeinated or decaffeinated) among coffee drinkers was based on which type of coffee the participant reported drinking more than half the time; an indicator variable represented coffee drinkers who were missing data on caffeine type.

Smoking adjustments included current cigarette smoking status (current, former, never), number of cigarettes smoked per day (1–10, 11–20, 21–30, 31–40, 41–60, ≥ 60), time of smoking cessation among former smokers (< 1 year, 1 to < 5 years, 5 to < 10 years, or ≥ 10 years before baseline) and whether a participant ever smoked pipe/cigars (yes/no).

Further adjustment for other covariates did not substantially alter these estimates. There was an interaction between coffee intake and person-years ( $P = 0.0043$ ). Therefore, we subdivided the data into three subsets ( $\leq 3.7$ ,  $> 3.7$  to  $7.2$  and  $> 7.2$  person-years), each with approximately one-third of cases, to estimate how HRs varied over the follow-up time period. The HR comparing  $\geq 6$  cups/day with none was greater than 1 in each subset although slightly higher in the later follow-up categories; HRs (95% CI) were 1.07 (0.88-1.31), 1.39 (1.14-1.68) and 1.37 (1.13-1.67) for the early, middle and later person-year categories, respectively. Since the direction of the association was similar in each stratum of follow-up time, we present the overall HRs for our main analysis; these HRs represent an average of the HRs over the entire follow-up period.

We also found similar associations for each histological type of lung cancer, and for participants drinking predominantly caffeinated or decaffeinated coffee (Table 3). The

association between high intake of caffeinated coffee and lung cancer hazard was largely driven by adenocarcinoma, but hazard rates were elevated in each category of caffeinated coffee intake compared with non-drinkers. For decaffeinated coffee, hazard rates were highest for small-cell carcinoma with coffee drinkers having a higher hazard compared with non-drinkers.

The highest level of coffee intake ( $\geq 4$  cups/day) was associated with a higher hazard rate of lung cancer in men (HR = 1.25; 95% CI = 1.09-1.43) but this was less prominent in women (HR = 1.10; 95% CI = 0.95-1.26) (Supplemental Table 3, available as Supplementary data at *IJE* online). No clear sex differences were observed for the histological subtypes.

We examined interactions between coffee intake and person-years in all aforementioned stratified analyses to test the proportional hazards assumption; this interaction term was significant among squamous cell lung cancer ( $P = 0.02$ ) and total lung cancer among women ( $P = 0.02$ ),

**Table 4.** Hazard ratios (95% confidence intervals) for lung cancer according to coffee intake by smoking status, time since smoking cessation and number of cigarettes smoked per day in the NIH-AARP Diet and Health Study ( $N = 442\ 280$ )

Smoking subgroup, cigarettes per day ( $n$ )	Cancer	Coffee intake				$P$ -trend
		None (ref.)	$\leq 1$ cup/day	2–3 cups/day	$\geq 4$ cups/day	
Never smokers ( $n = 147\ 706$ )	No. of cases	66	188	144	33	0.5699
	HR (95% CI)	1.00	1.13 (0.86-1.47)	0.97 (0.74-1.29)	1.08 (0.74-1.58)	
Former smokers ( $n = 228\ 875$ )						
	Quit $\geq 10$ years ago					
$\leq 10$ (49 721)	No. of cases	30	96	90	28	0.6550
	HR (95% CI)	1.00	0.68 (0.45-1.03)	0.60 (0.40-0.91)	0.81 (0.48-1.36)	
11–20 (53 718)	No. of cases	48	218	295	76	0.4408
	HR (95% CI)	1.00	0.89 (0.65-1.21)	0.88 (0.65-1.20)	0.82 (0.57-1.18)	
21–40 (53 447)	No. of cases	49	285	516	210	0.0002
	HR (95% CI)	1.00	1.26 (0.93-1.71)	1.42 (1.06-1.91)	1.67 (1.23-2.28)	
$>40$ (18 877)	No. of cases	27	134	249	112	0.2949
	HR (95% CI)	1.00	1.15 (0.76-1.75)	1.32 (0.89-1.97)	1.26 (0.82-1.91)	
Quit 5–9 years ago						
	$\leq 10$ (5461)	No. of cases	6	24	37	9
	HR (95% CI)	1.00	0.72 (0.30-1.77)	0.89 (0.37-2.11)	0.93 (0.33-2.61)	
11–20 (10 003)	No. of cases	15	78	136	41	0.5215
	HR (95% CI)	1.00	0.86 (0.49-1.50)	1.00 (0.59-1.71)	0.96 (0.59-1.71)	
21–40 (13 487)	No. of cases	25	135	254	116	0.3208
	HR (95% CI)	1.00	0.90 (0.59-1.38)	0.96 (0.63-1.45)	1.03 (0.67-1.60)	
$>40$ (4678)	No. of cases	11	49	131	72	0.3323
	HR (95% CI)	1.00	0.82 (0.43-1.58)	1.22 (0.66-2.25)	1.03 (0.55-1.96)	
Quit 1–4 years ago						
	$\leq 10$ (3365) <sup>a</sup>	No. of cases	1	24	26	17
	HR (95% CI)	–	–	–	–	
11–20 (6308)	No. of cases	17	55	126	61	0.0446
	HR (95% CI)	1.00	0.61 (0.36-1.06)	0.83 (0.50-1.38)	0.97 (0.56-1.66)	
21–40 (7659)	No. of cases	18	77	189	113	0.0685
	HR (95% CI)	1.00	0.95 (0.57-1.59)	1.11 (0.68-1.80)	1.24 (0.75-2.05)	
$>40$ (2151)	No. of cases	8	25	69	53	0.3415
	HR (95% CI)	1.00	0.64 (0.29-1.42)	0.93 (0.45-1.94)	0.90 (0.43-1.90)	
Current smokers ( $n = 65\ 699$ )						
	$\leq 10$ (16 382)	No. of cases	27	178	296	152
	HR (95% CI)	1.00	1.17 (0.78-1.76)	1.21 (0.81-1.79)	1.36 (0.90-2.04)	
11–20 (26 081)	No. of cases	74	266	722	556	0.0022
	HR (95% CI)	1.00	0.90 (0.69-1.16)	1.03 (0.81-1.31)	1.14 (0.89-1.45)	
21–40 (20 769)	No. of cases	72	217	637	750	0.0012
	HR (95% CI)	1.00	0.98 (0.75-1.28)	1.00 (0.78-1.28)	1.18 (0.93-1.51)	
$>40$ (2457)	No. of cases	11	37	85	145	0.2852
	HR (95% CI)	1.00	1.28 (0.65-2.50)	1.20 (0.64-2.25)	1.38 (0.75-2.55)	

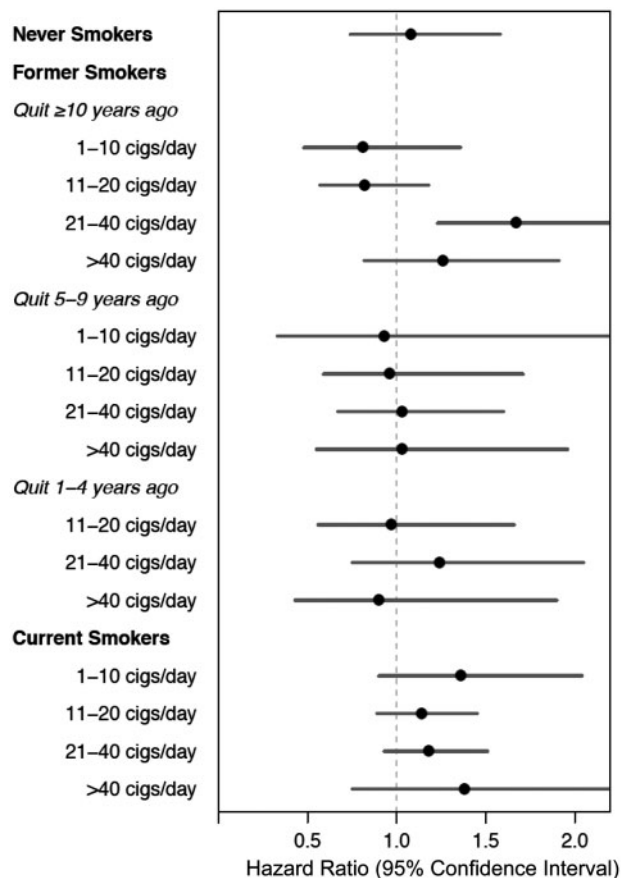
Participants who smoked pipes/cigars but not cigarettes ( $n = 15\ 086$ ) were excluded from model. Referent group was non-drinkers of coffee. Hazard ratios were adjusted for age at study baseline (continuous) and sex, current cigarette smoking status (current, former, never), number of cigarettes smoked per day (1–10, 11–20, 21–30, 31–40, 41–60,  $\geq 60$ ), time of smoking cessation among former smokers ( $< 1$  year, 1 to  $< 5$  years, 5 to  $< 10$  years, or  $\geq 10$  years before baseline), whether a participant ever smoked pipe/cigars (yes/no).

<sup>a</sup>Hazard estimates not provided for former smokers who quit 1–4 years ago and had smoked  $\leq 10$  cigarettes/day, due to small number of cases ( $n = 1$ ) in referent group.

and for caffeinated and decaffeinated coffee ( $P = 0.01$  and  $P < 0.0001$ , respectively). Using the methodology detailed above for total lung cancer, we examined the HRs by follow-up time period; the pattern of the HRs across the categories of coffee consumption did not differ qualitatively by the three time periods in these sub-analyses, so we present the overall HRs herein.

To better understand the impact of potential residual confounding by smoking on the observed associations, we

estimated HRs within more detailed smoking strata including cigarettes per day among current and former smokers, and additionally by years since quitting smoking among former smokers (Table 4 and Figure 1). Whereas the association looked qualitatively different by smoking strata (Figure 1), there was no statistical interaction between coffee and smoking ( $P$ -value for interaction = 0.427). Among never smokers ( $n = 431$  cases), no association was observed for coffee drinking and lung cancer overall



**Figure 1.** Hazard ratios for lung cancer in the NIH-AARP Diet and Health Study ( $N=442\ 280$ ), comparing coffee consumption of  $\geq 4$  cups/day with none, by smoking status, time since quitting smoking and number of cigarettes smoked per day. Participants who smoked pipes/cigars but not cigarettes ( $n=15\ 086$ ) were excluded from model. Cox proportional hazards regression used non-drinkers of coffee (reported never drinking coffee during the past year) as the referent group. Hazard ratios were adjusted for age at study baseline (continuous) and sex, current cigarette smoking status (current, former, never), number of cigarettes smoked per day (1-10, 11-20, 21-30, 31-40, 41-60,  $\geq 60$ ), time of smoking cessation among former smokers ( $< 1$  year, 1 to  $< 5$  years, 5 to  $< 10$  years, or  $\geq 10$  years before baseline) and whether a participant ever smoked pipe/cigars (yes/no). Hazard estimates not provided for former smokers who quit 1-4 years ago and had smoked  $\leq 10$  cigarettes/day, due to small number of cases ( $n=1$ ) in referent group.

(comparing  $\geq 4$  cups/day with 0, HR = 1.08; 95% CI = 0.74-1.58) or within each examined histological subtype stratum (Supplemental Table 4, available as Supplementary data at *IJE* online). Additionally, we found no consistent association between coffee drinking and lung cancer among strata of smoking intensity. Although positive associations were seen among several categories of cigarettes per day in current smokers (current smoking participants who smoked 11-20 or 21-39 cigarettes per day), we observed little evidence for an association in nearly every other examined stratum, including participants who quit 1-4 years before study baseline.

## Discussion

In our study, we observed a positive association between coffee drinking and subsequent lung cancer, with participants drinking  $\geq 6$  cups/day at about 30% greater hazard than non-drinkers. However, we observed little evidence for an association among examined subgroups of smoking use. For example, we saw no association among former and never smokers, which are the largest subgroups in our cohort. Residual confounding is also of substantial concern as coffee drinkers in our cohort were substantially more likely to be cigarette smokers, and our adjustment for cigarette smoking was likely imperfect. Together, these observations suggest that the association we saw overall was most likely not causal and would have been further attenuated with perfect adjustment for lifetime tobacco use. However, we cannot completely exclude the possibility of an association, particularly as we had only modest case numbers in many of our subgroups of cigarette use. In this study, adenocarcinoma was the lung cancer subtype most strongly associated with coffee intake, although squamous cell carcinoma is most closely linked to smoking. Additionally, caffeinated coffee was most strongly associated with adenocarcinoma whereas decaffeinated coffee was most strongly associated with small-cell carcinoma. Further study is needed to determine whether these findings are due to chance or reflect potential biological differences in disease aetiology between subtypes of lung cancer.

A recent meta-analysis of the topic reported a positive association but advocated a cautious interpretation, as many of the included studies did not adjust for tobacco use.<sup>8,38</sup> Among the previous studies which detected increased risk among coffee drinkers,<sup>13-19</sup> many were case-control studies;<sup>14-16</sup> these study designs may be subject to greater recall and selection biases and thus further evidence from prospective studies was needed. Previous studies were not as big as ours and used various methods to address residual confounding by smoking. Many of the positive associations between coffee drinking and lung cancer incidence or mortality were detected among smokers, either in studies limited to smokers<sup>14</sup> or in strata of smokers in stratified analyses within larger studies.<sup>20</sup>

As coffee drinkers were more likely to smoke than non-drinkers and smoking is the most important cause of lung cancer, we investigated potential residual confounding using a multifaceted approach. First, the higher hazard observed with coffee intake in the age- and sex-adjusted models was substantially attenuated once we carefully adjusted for smoking. Secondly, hazard estimates in never smokers were close to 1. Although we observed elevated associations among ever smokers, in stratified analyses these elevations persisted only in some categories of



smoking, particularly current smokers; we would expect residual confounding by smoking to exert a maximum effect among this group. Although we assessed a number of different aspects of smoking use in our study, we lacked data that would allow us to more comprehensively adjust for smoking, including depth of inhalation, environmental smoke exposure, age at smoking initiation, years smoked during one's lifetime, nicotine dependence and detailed information on pipe and cigar use. Based on the substantial attenuation we observed after adjustment for age at smoking cessation and typical cigarettes per day, it is plausible that additional adjustment for tobacco use would have further attenuated our observed associations. Supporting this idea, associations for coffee and lung cancer appeared to be somewhat stronger in men whereas tobacco use appeared to be somewhat lower in women. However as we had only modest case numbers in many of our stratified analyses, we cannot exclude the presence of a modest positive or inverse association.

In our study, greater smoking intensity was strongly correlated with heavier coffee consumption. A potential contributing factor for this correlation is the shared CYP1A2 metabolic pathway. In genome-wide association (GWAS) studies reported by our group and confirmed by others, the CYP1A2 gene is related to self-reported habitual coffee consumption.<sup>39,40–42</sup> Moreover, CYP1A2 performs the first step in the metabolic processing of caffeine. Both caffeine (a naturally occurring compound in coffee) and compounds in tobacco smoke including polycyclic aromatic hydrocarbons (PAHs) and nicotine, upregulate the CYP1A2 pathway.<sup>43–45</sup> Consequently, heavier smokers may need to drink more coffee to receive the same effect from caffeine as non-smoking counterparts. Also, individuals with high coffee intake may need to smoke more cigarettes to receive the same effect from nicotine.

The prospective study design is a major strength of our study. Our study also benefited from the large size of the underlying cohort, such that the number of incident lung cancers was higher than that of all previous studies combined, allowing greater statistical power to conduct stratified analyses.

However, our study also had limitations. Coffee consumption was self-reported, and participants were only queried about typical coffee consumption in the past year; we lack data on cumulative exposure, but coffee consumption is considered to be relatively stable over time. The onset of exposure (here, coffee) among coffee drinkers likely occurred decades before the administration of the food frequency questionnaire, and thus there may be a selection bias in this sample of individuals who are still alive and cancer free at study baseline; this type of bias is

common in cohort studies.<sup>46–49</sup> We categorized participants' coffee consumption as decaffeinated or caffeinated based on which type they drank more than half the time, and we do not have quantitative information on caffeine. The majority of coffee consumed in the USA is filtered coffee, and thus our findings largely reflect the association between filtered coffee and lung cancer; it is plausible that associations may differ for other types of coffee (unfiltered, percolated, espresso) due to varying amounts of caffeine and other constituents, and future studies should investigate these potential differences.

In addition, the effect of measurement error associated with data derived from food frequency questionnaires<sup>50</sup> is unpredictable due to potential residual confounding from other confounders measured with error. To minimize this measurement error, we conducted sensitivity analyses in which self-reported coffee intake was calibrated using a subgroup of participants who completed two 24-h dietary recalls. In these analyses, we found that the magnitude and direction of the associations for total coffee intake with overall lung cancer and each lung cancer subtype were similar to our main findings. Some potential confounding influences may remain in our data, as we are missing data on past occupational exposures, including carcinogens, which may impact on lung cancer hazard; however, this cohort largely consists of retired individuals with a high level of education, and thus it is unlikely that such exposures were common in this group. Additionally, although movement out of the area linked to study data was minimal,<sup>51</sup> incident cases could not be identified in cohort members who moved outside the area covered by study-linked registries. Lastly, examination of associations between coffee consumption and risk of subtype-specific lung cancers is a strength of our study, but approximately one-third of lung cancer cases had unspecified histology. Cases diagnosed by fine-needle aspiration or cytology, such as advanced stage lung cancer cases that are not candidates for curative resection, often have insufficient tissue to establish a precise histological subtype. However, we did not observe marked differences in coffee-related risk among lung cancer cases with unspecified compared with well-characterized histology, and the confidence intervals for all the histology groups substantially overlap.

In summary, we observed a positive association between coffee drinking and lung cancer in our study, with participants who drank  $\geq 6$  cups a day at about 30% higher hazard compared with non-drinkers. However, the association was substantially attenuated after adjustment for tobacco smoking. As our adjustment for lifetime tobacco use is imperfect, it is likely that the observed association between coffee drinking and lung cancer in our

cohort is due to residual confounding by tobacco smoking. Nevertheless, we cannot rule out a positive association.

## Supplementary Data

Supplementary data are available at *IJE* online.

## Acknowledgments

K.A.G. and R.S. had primary responsibility for the final content of the manuscript.

## Funding

This work was supported by the Intramural Research Program of the National Institutes of Health, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department of Health and Human Services.

**Conflict of interest:** None declared.

## References

- American Cancer Society. *Cancer Facts and Figures 2015*. Atlanta, GA: American Cancer Society, 2015.
- Farah A. Coffee constituents. In: Chu Y-F (ed). *Coffee: Emerging Health Effects and Disease Prevention*. Hoboken, NJ: Wiley-Blackwell;2012.
- Freedman ND, Park Y, Abnet CC, Hollenbeck AR, Sinha R. Association of coffee drinking with total and cause-specific mortality. *N Engl J Med* 2012;**366**:1891–904.
- Lai GY, Weinstein SJ, Albanes D *et al*. The association of coffee intake with liver cancer incidence and chronic liver disease mortality in male smokers. *Br J Cancer* 2013;**109**:1344–51.
- Sinha R, Cross AJ, Daniel CR *et al*. Caffeinated and decaffeinated coffee and tea intakes and risk of colorectal cancer in a large prospective study. *Am J Clin Nutr* 2012;**96**:374–81.
- Gunter MJ, Schaub JA, Xue X *et al*. A prospective investigation of coffee drinking and endometrial cancer incidence. *Int J Cancer* 2012;**131**:E530–36.
- Crippa A, Discacciati A, Larsson SC, Wolk A, Orsini N. Coffee consumption and mortality from all causes, cardiovascular disease, and cancer: a dose-response meta-analysis. *Am J Epidemiol* 2014;**180**:763–75.
- Tang N, Wu Y, Ma J, Wang B, Yu R. Coffee consumption and risk of lung cancer: A meta-analysis. *Lung Cancer* 2010;**67**:17–22.
- Lee PN, Forey BA, Coombs KJ. Systematic review with meta-analysis of the epidemiological evidence in the 1900s relating smoking to lung cancer. *BMC Cancer* 2012;**12**:385.
- Nomura A, Heilbrun LK, Stemmermann GN. Prospective study of coffee consumption and the risk of cancer. *J Natl Cancer Inst* 1986;**76**:587–90.
- Nilsson L, Johansson I, Lenner P, Lindahl B, Van Guelpen B. Consumption of filtered and boiled coffee and the risk of incident cancer: a prospective cohort study. *Cancer Causes Control* 2010;**21**:1533–44.
- Ren JS, Freedman ND, Kamangar F *et al*. Tea, coffee, carbonated soft drinks and upper gastrointestinal tract cancer risk in a large United States prospective cohort study. *Eur J Cancer* 2010;**46**:1873–81.
- Stensvold I, Jacobsen BK. Coffee and cancer: a prospective study of 43,000 Norwegian men and women. *Cancer Causes Control* 1994;**5**:401–08.
- Baker JA, McCann SE, Reid ME, Nowell S, Beehler GP, Moysich KB. Associations between black tea and coffee consumption and risk of lung cancer among current and former smokers. *Nutr Cancer* 2005;**52**:15–21.
- Javed, Luqman M, Daud S, Raheem N, Ahmad J, Khan AUH. Risk factors for lung cancer in the Pakistani population. *Asian Pac J Cancer Prev* 2014;**15**:3035–39.
- Ganesh B, Sushama S, Monika S, Suvarna P. A case-control study of risk factors for lung cancer in Mumbai, India. *Asian Pac J Cancer Prev* 2011;**12**:357–62.
- Jacobsen BK, Bjelke E, Kvale G, Heuch I. Coffee drinking, mortality, and cancer incidence: results from a Norwegian prospective study. *J Natl Cancer Inst* 1986;**76**:823–31.
- Fu YY, Takezaki T, Tajima K. [Risk factors of lung cancer – follow-up studies in Nagoya Japan]. *Zhonghua Liu Xing Bing Xue Za Zhi* 1997;**18**:328–30.
- Takezaki T, Hirose K, Inoue M *et al*. Dietary factors and lung cancer risk in Japanese: with special reference to fish consumption and adenocarcinomas. *Br J Cancer* 2001;**84**:1199–206.
- Chow WH, Schuman LM, McLaughlin JK *et al*. A cohort study of tobacco use, diet, occupation, and lung cancer mortality. *Cancer Causes Control* 1992;**3**:247–54.
- Hu J, Mao Y, Dryer D, White K. Risk factors for lung cancer among Canadian women who have never smoked. *Cancer Detect Prev* 2002;**26**:129–38.
- Mettlin C. Milk drinking, other beverage habits, and lung cancer risk. *Int J Cancer* 1989;**43**:608–12.
- Mendilaharsu M, De Stefani E, Deneo-Pellegrini H, Carzoglio JC, Ronco A. Consumption of tea and coffee and the risk of lung cancer in cigarette-smoking men: a case-control study in Uruguay. *Lung Cancer* 1998;**19**:101–07.
- Nyberg F, Agrenius V, Svartengren K, Svensson C, Pershagen G. Dietary factors and risk of lung cancer in never-smokers. *Int J Cancer* 1998;**78**:430–36.
- Axelsson G, Liljeqvist T, Andersson L, Bergman B, Rylander R. Dietary factors and lung cancer among men in west Sweden. *Int J Epidemiol* 1996;**25**:32–39.
- Khan MM, Goto R, Kobayashi K *et al*. Dietary habits and cancer mortality among middle aged and older Japanese living in Hokkaido, Japan by cancer site and sex. *Asian Pac J Cancer Prev* 2004;**5**:58–65.
- Kubik A, Zatloukal P, Tomasek L *et al*. A case-control study of lifestyle and lung cancer associations by histological types. *Neoplasma* 2008;**55**:192–99.
- Sanikini H, Radoi L, Menvielle G *et al*. Coffee consumption and risk of lung cancer: the ICARE study. *Eur J Epidemiol* 2015;**30**: 81–85.
- Khuder SA. Effect of cigarette smoking on major histological types of lung cancer: a meta-analysis. *Lung Cancer* 2001;**31**:139–48.
- Subramanian J, Govindan R. Lung cancer in never smokers: a review. *J Clin Oncol* 2007;**25**:561–70.

31. Lofffield E, Freedman ND, Graubard BI *et al.* Coffee drinking and cutaneous melanoma risk in the NIH-AARP Diet and Health Study. *J Natl Cancer Inst* 2015;**107**(2). doi: 10.1093/jnci/dju421.
32. Gierach GL, Freedman ND, Andaya A *et al.* Coffee intake and breast cancer risk in the NIH-AARP diet and health study cohort. *Int J Cancer* 2012;**131**:452–60.
33. 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**:1119–25.
34. 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 Registry Manag* 2005;**32**:70–7.
35. World Health Organization. *International Classification of Diseases for Oncology. ICD-O-3*. 3rd edn. Geneva: WHO, 2010.
36. Brinton LA, Gierach GL, Andaya A *et al.* Reproductive and hormonal factors and lung cancer risk in the NIH-AARP Diet and Health Study Cohort. *Cancer Epidemiol Biomarkers Prev* 2011;**20**:900–11.
37. Freedman LS, Schatzkin A, Wax Y. The impact of dietary measurement error on planning sample size required in a cohort study. *Am J Epidemiol* 1990;**132**:1185–95.
38. Wang Y, Yu X, Wu Y, Zhang D. Coffee and tea consumption and risk of lung cancer: a dose-response analysis of observational studies. *Lung Cancer* 2012;**78**:169–70.
39. Rodenburg EM, Eijgelsheim M, Geleijnse JM *et al.* CYP1A2 and coffee intake and the modifying effect of sex, age, and smoking. *Am J Clin Nutr* 2012;**96**:182–87.
40. Cornelis MC, Monda KL, Yu K *et al.* Genome-wide meta-analysis identifies regions on 7p21 and 15q24 as determinants of habitual caffeine consumption. *PLoS Genet* 2011;**7**:e1002033.
41. Sulem P, Gudbjartsson DF, Geller F *et al.* Sequence variants at CYP1A1–CYP1A2 and AHR associate with coffee consumption. *Hum Mol Genet* 2011;**20**:2071–77.
42. Amin N, Byrne E, Johnson J *et al.* Genome-wide association analysis of coffee drinking suggests association with CYP1A1/CYP1A2 and NRCAM. *Mol Psychiatry* 2012;**17**:1116–29.
43. Gunes A, Dahl ML. Variation in CYP1A2 activity and its clinical implications: influence of environmental factors and genetic polymorphisms. *Pharmacogenomics* 2008;**9**:625–37.
44. Campbell ME, Spielberg SP, Kalow W. A urinary metabolite ratio that reflects systemic caffeine clearance. *Clin Pharmacol Ther* 1987;**42**:157–65.
45. Kalow W, Tang BK. Use of caffeine metabolite ratios to explore CYP1A2 and xanthine oxidase activities. *Clin Pharmacol Ther* 1991;**50**(5 Pt 1):508–19.
46. Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. *Am J Epidemiol* 2003;**158**:915–20.
47. Applebaum KM, Malloy EJ, Eisen EA. Left Truncation, Susceptibility, and Bias in Occupational Cohort Studies. *Epidemiology* 2011;**22**:599–606.
48. Hernan MA, Alonso A, Logan R *et al.* Observational studies analyzed like randomized experiments an application to postmenopausal hormone therapy and coronary heart disease. *Epidemiology* 2008;**19**:766–79.
49. Flanders WD, Eldridge RC, McClellan W. A nearly unavoidable mechanism for collider bias with index-event studies. *Epidemiology* 2014;**25**:762–64.
50. Schatzkin A, Kipnis V. Could exposure assessment problems give us wrong answers to nutrition and cancer questions? *J Natl Cancer Inst* 2004;**96**:1564–65.
51. 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 Registry Manag* 2005;**32**:70–5.