Environ Health Perspect

DOI: 10.1289/EHP8719

Note to readers with disabilities: *EHP* strives to ensure that all journal content is accessible to all readers. However, some figures and Supplemental Material published in *EHP* articles may not conform to 508 standards due to the complexity of the information being presented. If you need assistance accessing journal content, please contact ehp508@niehs.nih.gov. Our staff will work with you to assess and meet your accessibility needs within 3 working days.

Supplemental Material

Contribution of Long-Term Exposure to Outdoor Black Carbon to the Carcinogenicity of Air Pollution: Evidence regarding Risk of Cancer in the Gazel Cohort

Emeline Lequy, Jack Siemiatycki, Kees de Hoogh, Danielle Vienneau, Jean-François Dupuy, Valérie Garès, Ole Hertel, Jesper Heile Christensen, Sergey Zhivin, Marcel Goldberg, Marie Zins, and Bénédicte Jacquemin

Table of Contents

Figure S1. Flowchart of the selection of Gazel participants included in the analyses pertaining to all-site and lung cancer.

Figure S2. Mean and standard deviation of the synthetic values plotted against iteration number for the imputed data. Smokpac: smoking pack-years; fdep2009: deprivation index (put back as nonimputed values in the final database); smokpas: passive smoking at work or at home; alcoclass: alcohol consumption; family1: marital status; vegfr: fruit and vegetable intake; ses: socioeconomic status; educ1: education; density2: urban classification (put back as nonimputed values in the final database).

Figure S3. On the top tier, exposure-response relationship between cumulative air pollutant (x-axis: black carbon in 10^{-5}/m, $PM_{2.5}$ in µg/m^3, and NO_2 in µg/m^3) and all-site cancer risk in the Gazel cohort (using a 3-degree of freedom spline function); on the bottom tier, the corresponding distribution of the exposure among the 293,210 person-years for the all-site cancer analysis (3,711 cases). Hazard ratios and confidence intervals estimated by separate single-pollutant Cox model with attained age as underlying time-scale and time-dependent variables, adjusted for sex, cumulative smoking pack-years, passive smoking, alcohol use, BMI, education, socioeconomic status, family status, vegetable and fruit consumption, occupational exposure to lung carcinogens, age at inclusion and calendar time. Exposures were lagged 10 years. Participants were excluded from the analysis if they were diagnosed with cancer before 1999.

Figure S4. Maps of the relative change (%) between the black carbon exposure assessment for all available Gazel participants between 1995 and 2000 (left, reference year 1995) and 2000 and 2005 (right, reference year 2000). For example, the dark orange values indicate that the black carbon estimates were up to 50% higher in 2000 than in 1995 and the dark blue values that black carbon estimates were up to 70% lower in 2005 than in 2000.

Table S1. Cancer cases breakdown by year for all-site cancer and lung cancer in the study period (1999-2015), as numbers and cumulative percentages. Cases diagnosed during 1989–1998 were excluded, thereby limiting the study period to 1999–2015.

Table S2. Cancer cases breakdown by site for first-occurring cases of all-site cancer in the study period (1999-2015), as numbers and percentages. Cases diagnosed during 1989–1998 were excluded, thereby limiting the study period to 1999–2015.

Table S3. Associations between black carbon and its residuals and $PM_{2.5}$ (in separate single-pollutant models) with incident all-site and lung cancer as hazard ratios (HR) and their 95% confidence intervals (CI), implementing either a 10-year or a 2-year lag period.

Table S4. Associations between cumulative black carbon (top half) and $PM_{2.5}$ (bottom half) and all-site incident cancer in the main, sensitivity, and stratified analyses with the number of identified cancer cases among the number of participant-year over the follow-up, implementing a 10-year lag period. These numbers correspond to Figure 2.

Table S5. Associations between cumulative black carbon (top half) and $PM_{2.5}$ (bottom half) and lung incident cancer in the main and sensitivity analyses with the number of identified cancer cases among the number of participant-year over the follow-up. These numbers corresponding to Figure 3.

Supplementary Figures

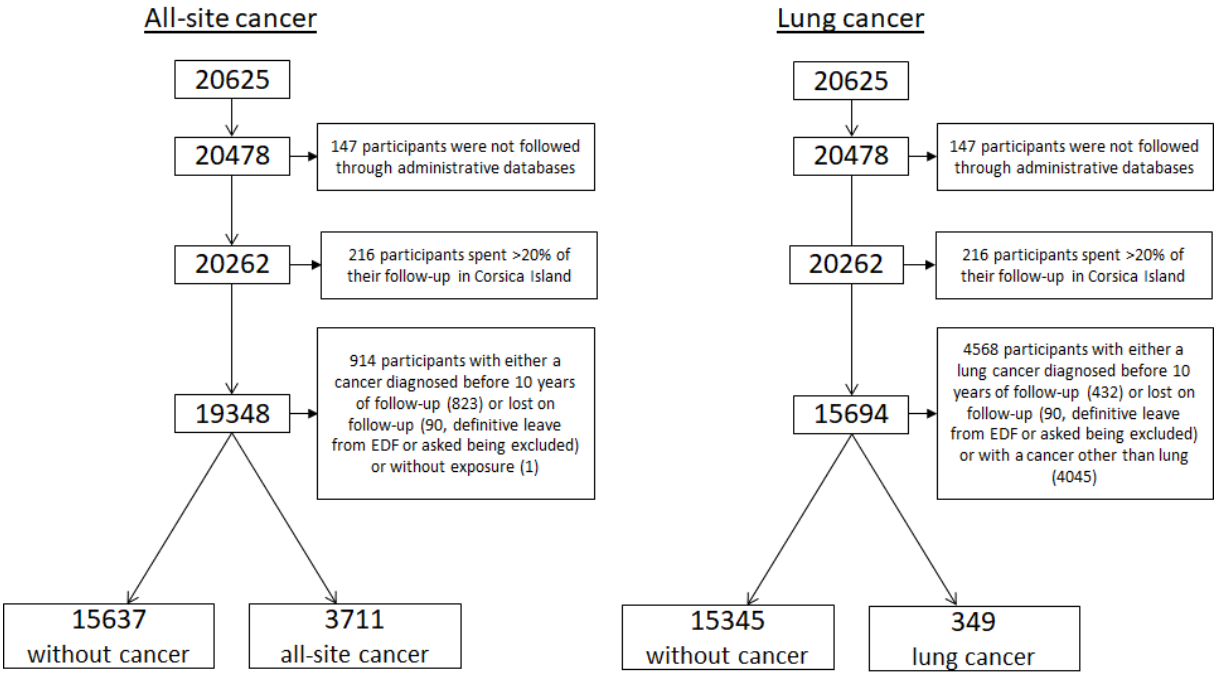

Figure S1: flowchart of the selection of Gazel participants included in the analyses pertaining to all-site and lung cancer.

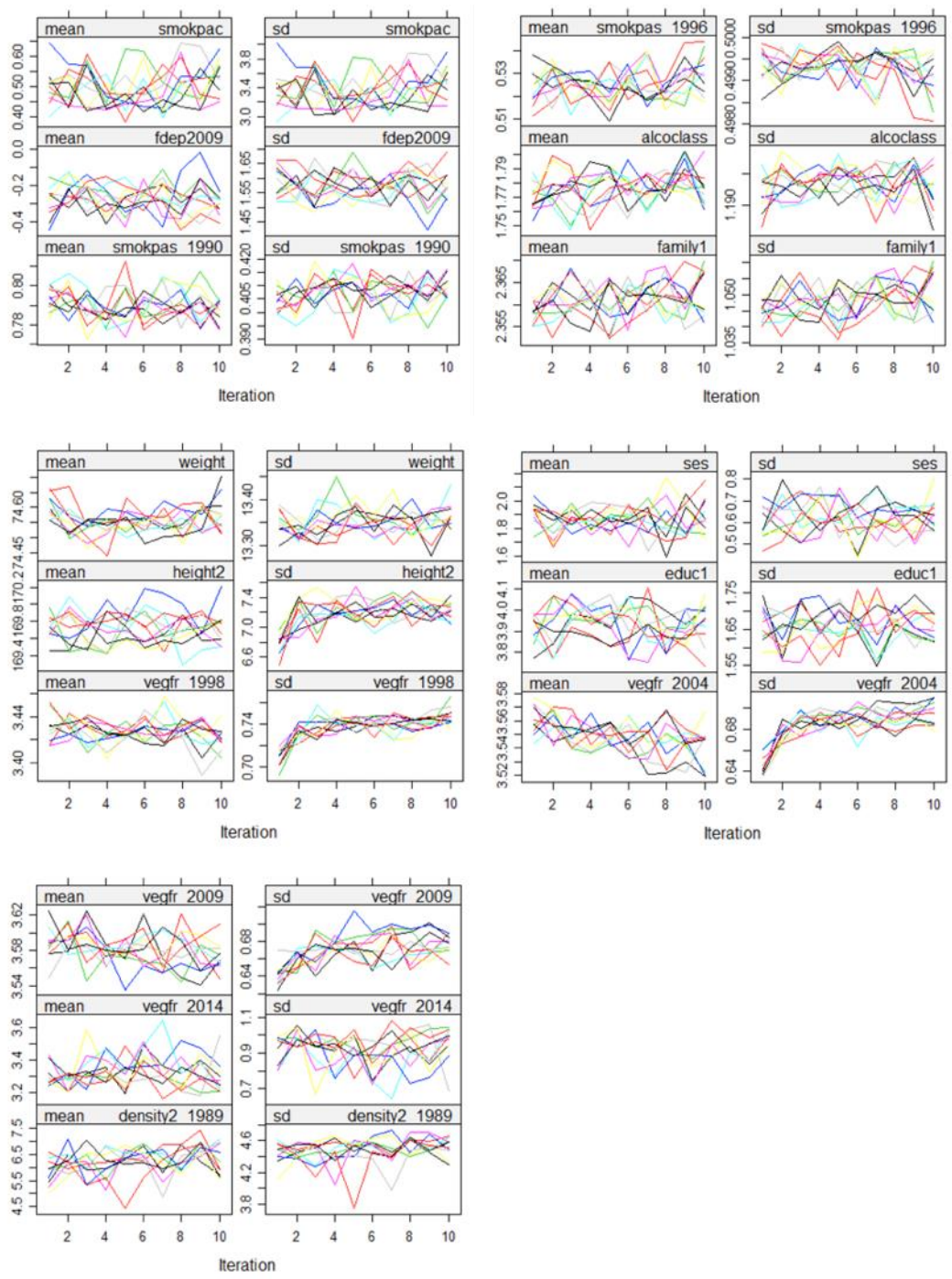

Figure S2: Mean and standard deviation of the synthetic values plotted against iteration number for the imputed data. Smokpac: smoking pack-years; fdep2009: deprivation index (put back as nonimputed values in the final database); smokpas: passive smoking at work or at home; alcoclass: alcohol consumption; family1: marital status; vegfr: fruit and vegetable intake; ses: socioeconomic status; educ1: education; density2: urban classification (put back as nonimputed values in the final database).

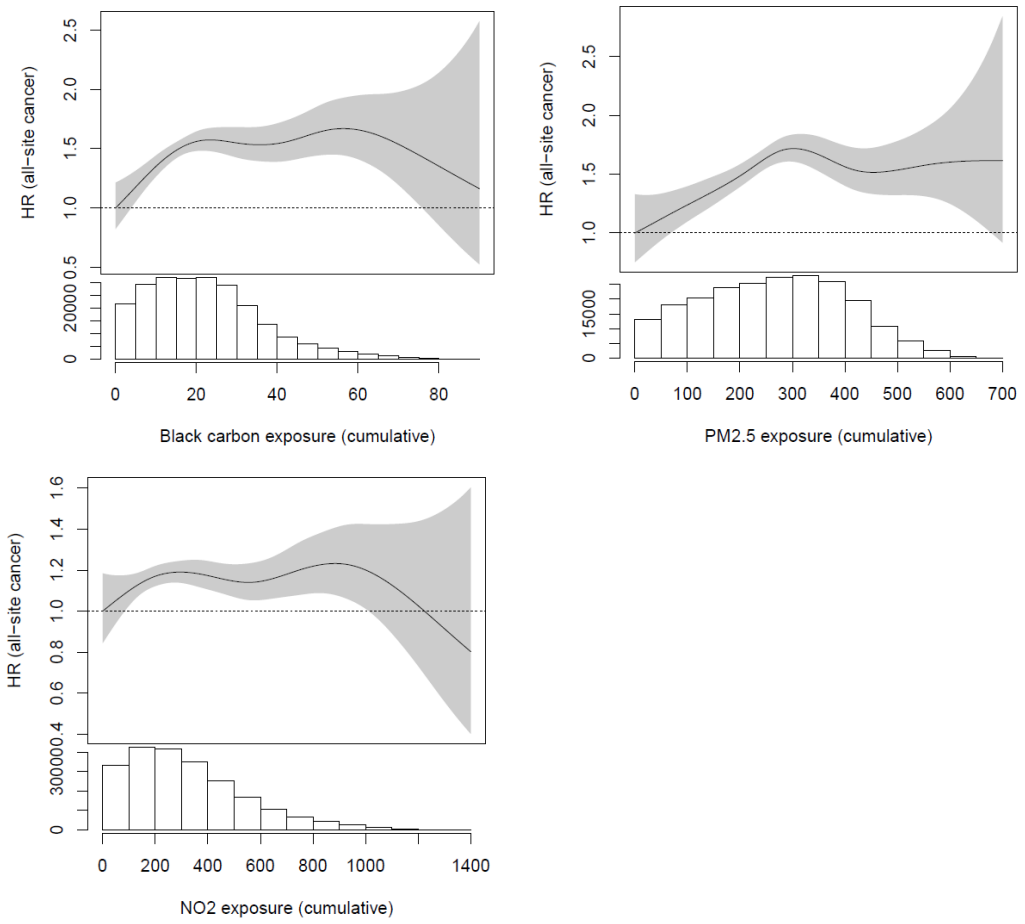

Figure S3: on the top tier, exposure-response relationship between cumulative air pollutant (x-axis: black carbon in 10^{-5}/m, PM2.5 in µg/m³, and NO2 in µg/m³) and all-site cancer risk in the Gazel cohort (using a 3-degree of freedom spline function); on the bottom tier, the corresponding distribution of the exposure among the 293,210 person-years for the all-site cancer analysis (3,711 cases). Hazard ratios and confidence intervals estimated by separate single-pollutant Cox model with attained age as underlying time-scale and time-dependent variables, adjusted for sex, cumulative smoking pack-years, passive smoking, alcohol use, BMI, education, socioeconomic status, family status, vegetable and fruit consumption, occupational exposure to lung carcinogens, age at inclusion and calendar time. Exposures were lagged 10 years. Participants were excluded from the analysis if they were diagnosed with cancer before 1999.

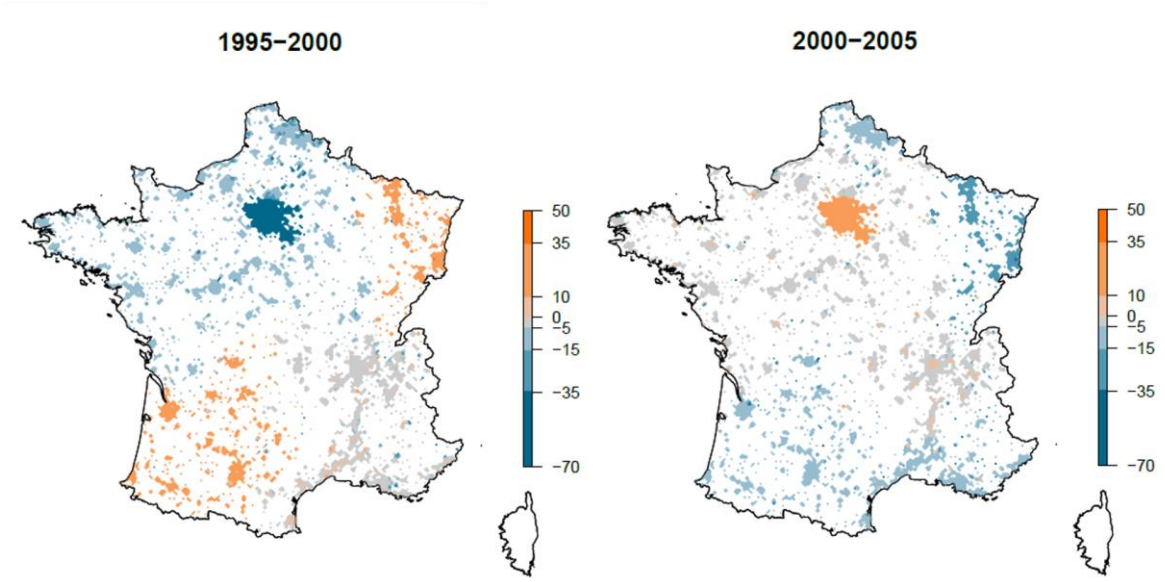

Figure S4: maps of the relative change (%) between the black carbon exposure assessment for all available Gazel participants between 1995 and 2000 (left, reference year 1995) and 2000 and 2005 (right, reference year 2000). For example, the dark orange values indicate that the black carbon estimates were up to 50% higher in 2000 than in 1995 and the dark blue values that black carbon estimates were up to 70% lower in 2005 than in 2000.

Supplementary Tables

Table S1: cancer cases breakdown by year for all-site cancer and lung cancer in the study period (1999-2015), as numbers and cumulative percentages. Cases diagnosed during 1989–1998 were excluded, thereby limiting the study period to 1999–2015

Cancer type/year	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
All-site (n)	78	88	121	157	137	171	181	234	328	299	284	274	305	291	282	255	307
(cumulative %)	2	4	8	12	15	20	25	31	39	47	55	62	70	78	85	92	100
Lung (n)	9	5	5	6	13	10	6	16	33	26	30	28	28	33	37	35	40
(cumulative %)	3	4	5	7	11	13	15	19	29	36	44	52	60	69	79	89	100

Table S2: cancer cases breakdown by site for first-occurring cases of all-site cancer in the study period (1999-2015), as numbers and percentages. Cases diagnosed during 1989–1998 were excluded, thereby limiting the study period to 1999–2015

Cancer site	N	%
Prostate	1301	34%
Breast	378	10%
Colorectal	362	9%
Hematopoietic	302	8%
Bronchus and lung	256	7%
Bladder	250	6%
Melanoma	145	4%
Kidney	136	4%
Female reproductive tract	72	2%

Table S3: associations between black carbon and its residuals and PM2.5 (in separate single-pollutant models) with incident all-site and lung cancer as hazard ratios (HR) and their 95% confidence intervals (CI), implementing either a 10-year or a 2-year lag period.

Lag (years)	Cancer site	Exposure variable	HR	CI	cases	person-years	p-value	AIC
10	All-site	Black carbon[a]	1.17	1.06-1.29	3,711	293,210	0.001	65978
		Black carbon residuals[b]	1.05	1.00-1.11			0.056	65984
		PM2.5[c]	1.20	1.06-1.34			0.003	65978
	Lung	Black carbon[a]	1.31	0.93-1.83	349	254,154	0.12	5842
		Black carbon residuals[b]	1.24	1.05-1.47			0.011	5838
		PM2.5[c]	1.01	0.68-1.51			0.942	5845
2	All-site	Black carbon[a]	1.17	1.06-1.29	3,708	292,664	0.002	65919
		Black carbon residuals[b]	1.05	0.99-1.10			0.09	65925
		PM2.5[c]	1.21	1.06-1.38			0.004	65920
	Lung	Black carbon[a]	1.21	0.85-1.71	348	253,646	0.293	5816
		Black carbon residuals[b]	1.14	0.96-1.35			0.141	5815
		PM2.5[c]	1.08	0.68-1.72			0.752	5817

AIC: Akaike Information Criterion.

All-site cancer cases were defined as the whole ICD10 chapter except C77-79 (secondary malignant neoplasms) and C44 (nonmelanoma skin cancers); we used C34 to identify lung cancer cases. Participants were excluded from the analysis if they were diagnosed with cancer before 1999. Exposures were lagged 10 years (top half of the table) or 2 years (bottom half of the table).

Cox model with attained age as underlying time-scale, adjusted for Model 3 covariates (sex, calendar time, age at inclusion, cumulative smoking pack-years, passive smoking, alcohol use, body mass index, education, socioeconomic status, family status, vegetable and fruit consumption, occupational exposure to nine lung carcinogens). Missing covariate data were imputed using MICE. Model-based estimates were pooled following Rubin's rules.

a HRs are for one IQR increase in ln-transformed cumulative black carbon (IQR = 1.0).

b HRs are for one IQR increase in ln-transformed cumulative black carbon residuals regressed on ln-transformed cumulative PM2.5 (IQR = 0.3).

c HRs are for an IQR increase in ln-transformed cumulative PM2.5 (IQR = 0.9 for the 10-year lag analyses and IQR = 1.0 for the 2-year lag analyses)

Table S4 : associations between cumulative black carbon (top half) and PM2.5 (bottom half) and all-site incident cancer in the main, sensitivity, and stratified analyses with the number of identified cancer cases among the number of participant-year over the follow-up, implementing a 10-year lag period. These numbers correspond to Figure 2.

Pollutant	Population/model	HR	CI-lower	CI-upper	p.value	cases	person-years
Black carbon	All (main model)	1.17	1.06	1.29	0.001	3711	293210
	All (main model further adjusted for deprivation)	1.17	1.06	1.29	0.002	3711	293210
	Using only address-level geocodes	1.17	1.00	1.36	0.056	1664	137092
	Participants with a follow-up longer than 20 years	1.40	1.20	1.64	0.000	1962	278907
	Using only complete cases data	1.02	0.87	1.19	0.809	1829	151037
	Considering missing data as a category	1.15	1.04	1.28	0.006	3711	293210
	Imputing missing data as the median/mode	1.19	1.08	1.31	0.001	3711	293210
	Female	1.16	0.96	1.41	0.129	755	81124
	Male	1.18	1.05	1.32	0.005	2956	212086
	Never Smoker	1.18	1.00	1.40	0.051	1252	113460
	Ever Smoker	1.17	1.04	1.33	0.012	2460	179750
	Nearest major road <500m	1.23	1.03	1.47	0.023	1028	71602
	Nearest major road >500m	1.16	1.02	1.31	0.024	2460	196254
	Urban	1.14	0.92	1.41	0.237	926	71915
	Semiurban	1.30	0.98	1.72	0.064	878	68923
	Rural	1.31	0.92	1.85	0.130	722	54668
PM2.5	All (main model)	1.20	1.06	1.34	0.003	3711	293210
	All (main model further adjusted for deprivation)	1.21	1.08	1.36	0.001	3711	293210
	Using only address-level geocodes	1.29	1.07	1.57	0.009	1664	137092
	Participants with a follow-up longer than 20 years	1.44	1.15	1.81	0.002	1962	278907
	Using only complete cases data	1.12	0.91	1.37	0.296	1829	151037
	Considering missing data as a category	1.19	1.04	1.35	0.009	3711	293210
	Imputing missing data as the median/mode	1.20	1.07	1.35	0.002	3711	293210
	Female	1.24	0.97	1.57	0.080	755	81124
	Male	1.18	1.03	1.35	0.014	2956	212086
	Never Smoker	1.23	1.00	1.52	0.055	1252	113460
	Ever Smoker	1.19	1.03	1.38	0.019	2460	179750
	Nearest major road <500m	1.26	1.01	1.56	0.039	1028	71602
	Nearest major road >500m	1.19	1.03	1.37	0.015	2460	196254
	Urban	1.22	0.96	1.56	0.118	926	71915
	Semiurban	1.17	0.91	1.50	0.231	878	68923
	Rural	1.12	0.86	1.46	0.398	722	54668

Hazard ratios and confidence intervals expressed for one IQR increase of cumulative exposure to black carbon or PM2.5 in separate single-pollutant Cox model with attained age as underlying time-scale and time-dependent variables, adjusted for sex, cumulative smoking pack-years, passive smoking, alcohol use, BMI, education, socioeconomic status, family status, fruit and vegetable consumption, occupational exposure to lung carcinogens, age at inclusion and calendar time. Exposures were lagged 10 years. Participants were excluded from the analysis if they were diagnosed with cancer before 1999. Unless specified otherwise, these model-based estimates were computed using MICE to address missing data, and were pooled following Rubin's rules.

Table S5: associations between cumulative black carbon (top half) and PM2.5 (bottom half) and lung incident cancer in the main and sensitivity analyses with the number of identified cancer cases among the number of participant-year over the follow-up. These numbers corresponding to Figure 3.

Pollutant	Population/model	HR	CI-lower	CI-upper	p-value	cases	person-years
Black carbon	All (main model)	1.31	0.93	1.83	0.118	349	254154
	All (main model further adjusted for deprivation)	1.36	0.98	1.89	0.070	349	254154
	Using all non lung cases as controls	1.33	0.94	1.87	0.103	349	318074
	Using only address-level geocodes	1.66	0.85	3.24	0.141	110	118980
	Participants with a follow-up longer than 20 years	1.67	1.05	2.66	0.030	225	250525
	Using only complete cases data	0.71	0.35	1.43	0.336	101	131264
	Considering missing data as a category	1.40	1.01	1.96	0.000	349	254154
	Imputing missing data as the median/mode	1.41	0.09	22.38	0.042	349	254154
PM2.5	All (main model)	1.01	0.68	1.51	0.942	349	254154
	All (main model further adjusted for deprivation)	1.14	0.77	1.68	0.523	349	254154
	Using all non lung cases as controls	1.04	0.69	1.55	0.863	349	318074
	Using only address-level geocodes	2.30	0.85	6.18	0.099	110	118980
	Participants with a follow-up longer than 20 years	1.56	0.79	3.11	0.202	225	250525
	Using only complete cases data	0.82	0.35	1.92	0.642	101	131264
	Considering missing data as a category	1.15	0.76	1.76	0.000	349	254154
	Imputing missing data as the median/mode	1.06	0.13	8.43	0.771	349	254154

Hazard ratios and confidence intervals expressed for one IQR increase in ln-transformed cumulative exposure to black carbon or PM2.5 in separate single-pollutant Cox model with attained age as underlying time-scale and time-dependent variables, adjusted for sex, cumulative smoking pack-years, passive smoking, alcohol use, BMI, education, socioeconomic status, family status, fruit and vegetable consumption, occupational exposure to lung carcinogens, age at inclusion and calendar time. Exposures were lagged 10 years. Participants were excluded from the analysis if they were diagnosed with cancer before 1999. Unless specified otherwise, these model-based estimates were computed using MICE to address missing data, and were pooled following Rubin's rules).