A Section 508–conformant HTML version of this article is available at https://doi.org/10.1289/EHP8719.

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,[1,2] Jack Siemiatycki,[2] Kees de Hoogh,[3,4] Danielle Vienneau,[3,4] Jean-François Dupuy,[5] Valérie Garès,[5] Ole Hertel,[6] Jesper Heile Christensen,[6] Sergey Zhivin,[1] Marcel Goldberg,[1] Marie Zins,[1] and Bénédicte Jacquemin[7]

[1]UMS 011, Institut national de la santé et de la recherché médicale (Inserm), Villejuif, France
[2]Centre de recherche du centre hospitalier de l'université de Montréal, Université de Montréal, Québec, Canada
[3]Swiss Tropical and Public Health Institute, Basel, Switzerland
[4]University of Basel, Basel, Switzerland
[5]UMR 6625 IRMAR, INSA, CNRS, Université de Rennes, Rennes, France
[6]Department of Environmental Science, Aarhus University, Roskilde, Denmark
[7]Univ Rennes, Inserm, EHESP, Irset (Institut de recherche en santé, environnement et travail) – UMR_S 1085, Rennes, France

Background: Black carbon (BC), a component of fine particulate matter [particles with an aerodynamic diameter ≤2.5 μm ($PM_{2.5}$)], may contribute to carcinogenic effects of air pollution. Until recently however, there has been little evidence to evaluate this hypothesis.

Objective: This study aimed to estimate the associations between long-term exposure to BC and risk of cancer. This study was conducted within the French Gazel cohort of 20,625 subjects.

Methods: We assessed exposure to BC by linking subjects' histories of residential addresses to a map of European black carbon levels in 2010 with back- and forward-extrapolation between 1989 and 2015. We used extended Cox models, with attained age as time-scale and time-varying cumulative exposure to BC, adjusted for relevant sociodemographic and lifestyle variables. To consider latency between exposure and cancer diagnosis, we implemented a 10-y lag, and as a sensitivity analysis, a lag of 2 y. To isolate the effect of BC from that of total $PM_{2.5}$, we regressed BC on $PM_{2.5}$ and used the residuals as the exposure variable.

Results: During the 26-y follow-up period, there were 3,711 incident cancer cases (all sites combined) and 349 incident lung cancers. Median baseline exposure in 1989 was 2.65 10^{-5}/m [interquartile range (IQR): 2.23–3.33], which generally slightly decreased over time. Using 10 y as a lag-time in our models, the adjusted hazard ratio per each IQR increase of the natural log-transformed cumulative BC was 1.17 (95% confidence interval: 1.06, 1.29) for all-sites cancer combined and 1.31 (0.93, 1.83) for lung cancer. Associations with BC residuals were also positive for both outcomes. Using 2 y as a lag-time, the results were similar.

Discussion: Our findings for a cohort of French adults suggest that BC may partly explain the association between $PM_{2.5}$ and lung cancer. Additional studies are needed to confirm our results and further disentangle the effects of BC, total $PM_{2.5}$, and other constituents. https://doi.org/10.1289/EHP8719

Introduction

Strong evidence over recent decades allowed classifying outdoor air pollution and fine particulate matter [fine particulate matter with an aerodynamic diameter ≤2.5 μm ($PM_{2.5}$)] as carcinogenic (Loomis et al. 2013; Pedersen et al. 2017; Raaschou-Nielsen et al. 2013; IARC 2016). Yet the separate effects of each $PM_{2.5}$ component (sulfates, nitrate, ammonium, organics, metals, etc.) are rarely quantified (Beelen et al. 2015; Ostro et al. 2011; Raaschou-Nielsen et al. 2016). Black carbon (BC), a component of $PM_{2.5}$, comes from incomplete combustion processes, mainly from anthropogenic sources such as fossil fuel or biomass burning (Chylek et al. 2015; Koelmans et al. 2006). The first health concerns about exposure to BC appeared decades ago (Mumford et al. 1990); since then, reports have accumulated linking exposure to BC with increased morbidity and mortality, including lung cancer mortality (Anenberg et al. 2012; Grahame et al. 2014; Hvidtfeldt et al. 2019; Yang et al. 2019), lower lung function and slower cognitive development in children (Paunescu et al. 2019; Sunyer et al. 2015), increased bone loss (Prada et al. 2017), and decreased cognitive functions in the elderly (Colicino et al. 2017; Wurth et al. 2018). Although evidence has accumulated on toxicity of BC, we still know little about the effects of chronic low-level exposure on cancer risk, partly because the paucity of available data on general population long-term exposure to BC left little opportunity for such studies. Recently, the ELAPSE project estimated annual outdoor BC concentrations between 1990 and 2015 at fine resolution over Europe (de Hoogh et al. 2018). In this study, we aimed to investigate the relationships between long-term exposure to BC and incident all-site and lung cancer in the population-based French cohort Gazel with a 26-y follow-up.

Material and Methods

Study Population

The Gazel cohort enrolled 20,625 participants in 1989 from the French national gas and energy company, Electricité-de-France Gaz-de-France (Goldberg et al. 2015). These participants, aged 35–50 y at enrollment, completed a baseline detailed self-administered questionnaire, then a follow-up questionnaire sent every year – with a high response rate during the follow-up (>80% from 1990 to 1992, and >70% from 1993 to 2015). Participants' histories of main residential addresses were collected and geocoded for each year since 1989. To minimize

Address correspondence to Emeline Lequy, UMS 011, Hôpital Paul Brousse, 16 avenue Paul Vaillant Couturier, 94807 VILLEJUIF CEDEX, France. Telephone number: +33 (0)1 77 74 74 19. Email: e.lequy@gmail.com and Bénédicte Jacquemin, Irset, 9 avenue du Prof. Léon Bernard, 35000 RENNES, France. Telephone number: +33 (0)2 23 23 7769. Email: benedicte.jacquemin@inserm.fr

Supplemental Material is available online (https://doi.org/10.1289/EHP8719).
The authors declare they have no actual or potential competing financial interests.
Received 25 November 2020; Revised 19 February 2021; Accepted 22 February 2021; Published 24 March 2021.

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 ehponline@niehs.nih.gov. Our staff will work with you to assess and meet your accessibility needs within 3 working days.

misclassification while retaining the largest possible number of participants, we excluded participants with the poorest exposure coverage [i.e., more than 20% of missing geocodes during their follow-up (due to stays abroad mainland France)]. This residential history was collected on an annual basis, so each address corresponds to a calendar year. We used last-observation carried forward to impute any missing addresses for the concerned participants. We identified the numbers of residential changes during the study period by choosing a threshold of 1-km difference in the geocodes to identify a substantial residential change; during the study period, we observed 13,981 of changes of more than 1 km, for 9,112 participants. Geocoding precision ranged from postal code (13%) to address level (48%).

We excluded 823 participants with any primary incident cancer diagnosed or censored before 1999, to take account of a potential 10-year lag between exposure and incidence/censoring (see "Statistical Analyses" section). This approach led to a slightly different study population for analyses on all-site and lung cancer. We also excluded 90 participants who were lost on follow-up (because they definitively left the company) or who asked to be removed from the study and their data to be deleted. Our study included participants who died during the follow-up without a diagnosis of cancer and who were censored at the date of death. Further, in the lung cancer analysis, we compared lung cancer cases to subjects not developing any cancer; thus we excluded participants with other cancers at any time during the study period (1999–2015) from the study population. Our study population included 19,348 and 15,694 participants for the analyses on primary incident all-site and primary incident lung cancer (Figure S1).

The Gazel study protocol was approved by the French authority for data confidentiality (Commission Nationale de l'Informatique et des Libertés No. 105,728) and by the Ethics Evaluation Committee of the Institut national de la santé et de la recherche médicale (Inserm, National Institute of Health and Medical Research) (IRB0000388, FWA00005831). The invitation to participate was sent by post to eligible persons, accompanied by a document detailing the project, the voluntary nature of their participation, the data collected, the conditions of security and confidentiality and the future use of the data. The subjects solicited were invited to complete a questionnaire indicating their consent.

Cancer Incidence

Incident cancer cases were ascertained from three sources: *a*) French national health administrative databases containing listings of incident cancers (more details below) during the study period (1999–2015); *b*) company records which have systematically recorded all cancers (except nonmelanoma skin cancers) diagnosed among their current employees [with pathology reports and the date of diagnosis, and coded according to the International Classification of Diseases (ICD)] (Goldberg et al. 1996); and *c*) cancer diagnoses self-reported by participants via the follow-up questionnaires from 2008 onward. Participants who gave consent were contacted for collecting medical information to obtain the date of diagnosis and the type of cancer. The linkage of participants to the French national health administrative databases that record each use of the health system allowed identifying cancer from data on hospitalizations and from the "Chronic Diseases" register (diseases including cancer for which all the treatment is reimbursed) with dates and diagnoses (Tuppin et al. 2017).

The ICD-10 classification system was used to code the type of cancer, with the whole ICD-10 chapter except C77–79 (secondary malignant neoplasms) and C44 (nonmelanoma skin cancers); we used C34 to identify lung cancer.

Exposure Assessment

For each subject of our study population, we estimated BC, $PM_{2.5}$, and nitrogen dioxide (NO_2) exposure in each year from 1989 to 2005, based on the subject's residential address linked to data from land use regression (LUR) models developed at a fine spatial scale (100×100 m) over Europe (de Hoogh et al. 2018). This linkage also accounts for any residential address change over the years (see "Study Population" section). $PM_{2.5}$ and BC measurement data came from two sources: $PM_{2.5}$ absorbance in samples collected in the ESCAPE project for BC (436 sites), and from regulatory monitoring data maintained in the AirBase European database for $PM_{2.5}$ (543 sites). For the year 2010, LUR models were developed by regressing the measured pollutant concentrations against a range of predictor variables (including land-use variables, road density, and altitude, as well as satellite-derived and chemical transport modeled pollutant estimates) followed for $PM_{2.5}$ only by universal kriging to explain spatial autocorrelation in the residuals. The full model (based on all monitoring sites) explained 72%, 54%, and 59% of $PM_{2.5}$, BC, and NO_2 respectively. Models were further validated, and shown to be robust, using a five hold out validation strategy which explained 66, 51, and 57% of the spatial variation in the respective measured $PM_{2.5}$, BC, and NO_2 concentrations (de Hoogh et al. 2018). Finally, the estimated concentrations for 2010 were rescaled annually for the years 1990–2015, by European Nomenclature of Territorial Units for Statistics -1 (NUTS-1) regions (i.e., European Union–defined administrative regions within countries) in France, using back- and forward extrapolation. This rescaling process was based on annual mean estimates (1990–2015) from the 26×26 km Danish Eulerian Hemispheric Model, downscaled from the original 50×50 km resolution using bilinear interpolation (Brandt et al. 2012). In addition, in this study, we further back-extrapolated $PM_{2.5}$ exposure to 1989. To visualize spatiotemporal differences in BC exposure over France, we mapped the differences in BC exposure over France for Gazel nonmover participants between periods 1995–2000 and 2000–2005. We calculated the relative change in %, for each nonmover participant for each pair of years. To improve the maps' readability, we averaged the results on a 5×5 km^2 grid.

Covariables

Based on recognized and suspected risk factors, we *a priori* selected the following sets of variables as potential confounders and/or effect modifiers:

Sociodemographic and occupational variables. Sex, education (attending school for 6–11 y, 12–13 y, 14–15 y, other secondary education, other diploma), and socioeconomic status (SES; low: blue-collar workers or clerks; medium: first-line supervisors or sales representatives; high: management), all at baseline. We also included a synthetic summary of occupational exposure to nine known lung carcinogens (asbestos, cadmium, chlorinated solvents, chromium, coal gasification, coal-tar pitch, creosotes, crystalline silica, and hydrazine) over the whole employment period [categorized into none, one, two, or at least three carcinogens, from each Gazel participant's career-long history linked to the French job-exposure matrix MATEX (Imbernon 1991)].

Lifestyle variables. Time-varying variables for tobacco (cumulative smoking pack-years), alcohol intake (abstinent, light drinker, moderate drinker, heavy drinker, unclear pattern), family status (single or not), body mass index (BMI, weight in kilograms divided by the square of height in meters). Some questions were only asked occasionally, such as passive smoking at home or at

work (yes or no) in 1990 and 1996, or fruit and vegetable intake (never or less than once a week; once or twice a week; more than twice a week but not every day; every day or almost) in 1998, 2004, 2009 and 2014. We processed these two variables to make them time-varying annually, attributing the data collected in 1990 and 1996 to each year between 1989–1995 and 1996–2015, respectively, for passive smoking, and the data collected in 1998, 2004, 2009, and 2014 to each year between 1989–1998, 1999–2004, 2005–2010, and 2011–2015, respectively, for fruit and vegetable intake.

Contextual variables. For all participants and every year, we calculated the distance to the nearest major road. At the municipality level, we obtained the population density in 1989, 2000, and 2010, from which we derived a urban classification: urban (high population density), semiurban (intermediate pop density), and rural (low population density). The population density cutoffs are based on the European urban/rural classification. To define whether the participants had lived solely in urban, semiurban, rural area, or in mixed areas over all the follow-up, we used the information from the 3 y for which we obtained such data. Also at the municipality level, we obtained the French deprivation index (Rey et al. 2009) calculated for 2009 for all participants who were still alive and therefore geocoded in 2009. To take into account participants who died before this variable could be calculated in 2009, we categorized the values from the French deprivation index into tertiles and added the missing values as a category so as to not lose any participant in the analyses because of this variable.

Imputations

For baseline variables, missing values ranged from 0% to 2.1% for sex and education, respectively. Throughout the follow-up, missing data ranged from 21% to 29% for alcohol consumption and BMI variables, respectively. We imputed all variables (except air pollution exposure and contextual variables) longitudinally for each participant using multiple imputations by chained equations from the R packages MICE and MICEADDS (van Buuren and Groothuis-Oudshoorn 2011), iterating 10 datasets 10 times with good convergence. All the variables described above were used as predictors. For the geographic variables (exposure to pollutants, deprivation index, distance to the road or urban classification), because the other predictors cannot predict them accurately, we recreated the initial missing values in our final dataset. We used the functions "2l.pmm" and "2l.only.pmm" for time-varying and time-independent variables, respectively. Following the MICE package manual (https://cran.r-project.org/web/packages/mice/mice.pdf), we assessed convergence visually using Figure S2; the streams are supposed to mingle well and without showing a clear trend. Only for fruit and vegetable intake did we observe a poor mingling of the streams, due to the correlation between the responses to this question asked in 1999, 2004, 2009 and 2014. Because we aimed at pooling these values into one time-dependent categorical variable, we discarded this issue. Model-based estimates were pooled following Rubin's rules.

Statistical Analyses

We used extended Cox models with BC as a time-dependent variable, with attained age as underlying time scale, to validate the proportional hazard assumption, while annually describing the exposure to air pollutants. We estimated the associations between incident cancer and BC, as a single pollutant, or taking $PM_{2.5}$ into account, following an approach based on residuals (Mostofsky et al. 2012). Since we have a rare opportunity to utilize a long time-series of BC exposure to study the long-term association with cancer incidence, we used cumulative exposures for each participant from baseline to incidence or censoring and adjusted for calendar time and age at inclusion (Pencina et al. 2007).

- 1. Single-pollutant approach: We included cumulative time-dependent black carbon exposure or cumulative time-dependent $PM_{2.5}$ exposure or cumulative time-dependent NO_2 exposure separately as a single pollutant in our main model. For these three pollutants, we used a spline function with 3 degrees of freedom (df) to test for nonlinearity. Based on visual inspection, the response to exposures approximately followed a natural logarithm-shaped curve for all-site cancer incidence (Figure S3), but not for cumulative NO_2 and lung cancer. To consider these nonlinear relationships, and to facilitate interpreting the results of the Cox models, we natural log-transformed the cumulative annual BC and $PM_{2.5}$ time-dependent exposures for both outcomes, and NO_2 time-dependent exposure for all-site cancer only.
- 2. Residual approach for black carbon: The Spearman's correlation coefficient between $PM_{2.5}$ and BC was 0.74. Including both of those exposure variables in a regression model can distort the true effects of one or both of those variables. As an attempt to isolate the effect of BC from that of $PM_{2.5}$, we followed the approach of Mostofsky et al. (2012), who suggested using the residuals of a regression between the constituent of $PM_{2.5}$ and $PM_{2.5}$ total mass. To do so, we first regressed BC (dependent variable) on $PM_{2.5}$ (independent variable). The residuals of this regression should represent the variations of BC independently of $PM_{2.5}$. When correctly specified, the residuals should be uncorrelated with $PM_{2.5}$. We included BC as natural log-transformed cumulative exposures in the regression, and cumulative $PM_{2.5}$ using a spline function with 4 df. This specification decorrelated the BC residuals from the cumulative $PM_{2.5}$ exposure. In Cox models using BC residuals as exposure, the coefficient represents the risk associated with higher levels of cumulative black carbon exposure, while holding cumulative $PM_{2.5}$ exposure constant. We did not further adjust for $PM_{2.5}$, because it would provide information to interpret the effect of the other constituents of $PM_{2.5}$, which was not the aim of this study. To test the independence of the effects of BC from those of NO_2, we did the same analysis using the residuals of a regression between BC and NO_2 because BC and NO_2 were also highly correlated (Spearman's correlation coefficient of 0.89) and precluded a bi-pollutant analysis with the two pollutants in the same model.

We provided hazard ratios for one interquartile range (IQR) of all variables of exposure to air pollution (after natural log-transforming cumulative BC and $PM_{2.5}$ for both outcomes, as well as for cumulative NO_2 for all-site cancer), and of the residual variables.

To consider cancer latency, it is customary, especially when using time-varying variables in the statistical analyses, to discount exposures that occurred recently, because these are unlikely to have affected the cancer risk (Rothman et al. 2008). Thus, we implemented a 10-y lag period during which exposure was not counted for any time-dependent variables except for passive smoking and for fruit and vegetable consumption, which already included a time-lag due to the way these variables were collected and interpolated (therefore using passive smoking values in 1996 only). The inclusion of contextual variables was explored only in sensitivity analyses. We used several levels of adjustment when using BC as exposure: *a*) Model 1 included sex, and calendar time and age at inclusion as continuous variables; *b*) Model 2 also included cumulative smoking pack-years and passive smoking (yes/no as defined in 1996); and *c*) Model 3 (main

model) was additionally adjusted for baseline education, SES, and occupational exposure to lung carcinogens and for time-varying alcohol consumption, family status, BMI, and fruit and vegetable consumption. We used the main model to derive estimates of the associations between $PM_{2.5}$, NO_2, or BC residuals and incident cancer. We calculated the Akaike Information Criterion (AIC) of each model, according to the level of adjustment or according to the exposure included.

We used multiple imputations by chained equations to conduct all the analyses (see section below) unless specified otherwise.

Sensitivity analyses were conducted for BC or $PM_{2.5}$ separately (single-pollutant models) by: *a*) implementing a 2-y lag period with the same study population as that of the analysis implementing the 10-y lag period; and, after again implementing a 10-y lag period; by *b*) further adjusting for the French deprivation index; *c*) restricting the study population to the participants with address-level geocodes throughout their follow-up; *d*) restricting the study population to the participants with at least 20 y of follow-up (thus an exposure window of at least 10 y); and *e*) considering three other ways to deal with missing data: conducting these analyses with complete cases only, considering missing values as a category (therefore categorizing continuous variables by quartiles), or imputing missing data as the median (for continuous variables) or the mode (for categorical variables). As a further sensitivity analysis to explore the nonlinear relationship of the association between BC exposure and all-site cancer, we developed a two-piece linear model, by including an interaction term between BC exposure and a Boolean variable with the most appropriate cut-off found out to be at 24 $10^{-5}/m$ (as visually observed in Figure S3 and with the maximum likelihood among values from 15 to 30 $10^{-5}/m$).

For all-site cancer, to search for any effect modification by sex, smoking status, urban classification, and distance to the nearest major road, we used the single-pollutant models for BC or $PM_{2.5}$ separately in single-pollutant models restricted to the following subsets: female or male, ever or never smokers, solely urban or solely semiurban or solely rural during the follow-up, and closer or farther than 500 m from the nearest major road over all the follow-up. Stratified analyses were not done for the analyses on lung cancer due to the small number of cases.

We conducted all the analyses with R (version 3.5.1; R Development Core Team 2018) with the SURVIVAL package (Therneau 2015; Therneau and Grambsch 2000).

Results

Study Population and Exposure

Study population characteristics are shown in Table 1. More than 70% were men. Mean age at baseline was 43.5, and median follow-up period was 27 y. Among those who eventually got cancer, the mean baseline age was 44.5 y, and the median follow-up until diagnosis was about 21 y. Most had 11 y or fewer of schooling, and most were employed in intermediate level jobs. Slightly fewer than half had been regular smokers at some time. Former and current smokers had cumulated 15.4 pack-years on average at baseline; those diagnosed with incident all-site or lung cancer cumulated 18 and 32 pack-years at baseline, respectively. We classified 71.4% participants as light drinkers, and 8.8% as heavy drinkers. Almost 70% of the participants ate fruit and vegetables almost every day and we calculated a median BMI of 24.3.

The exposure assessment yielded BC concentrations ranging between 0.7 and 8.9 $10^{-5}/m$ with a median of 1.9 between 1989 and 2015 (Table 2) with a modest decline over time (Figure 1) that differed slightly between regions (Supplementary Figure S4). The cumulative exposure ranged between 1.5 and 104.1 $10^{-5}/m$ with a median of 19.7. The exposure assessment yielded $PM_{2.5}$ concentrations ranging from 2.6 to 57.3 $\mu g/m^3$ with a median of 21.6 between 1989 and 2015, and a cumulative exposure ranging between 3.0 and 691.8 $\mu g/m^3$ with a median of 252.8 (Figure 1; Table 2).

Between 1989 and 2015, a total of 4,354 incident primary cancers (excluding nonmelanoma skin cancers) were diagnosed in the cohort, of which 410 were lung cancers. Our analysis lagged exposures by 10 y and thus included cancers diagnosed from 1999 onward only ($n = 3,711$ for all-site cancers; $n = 349$ for lung cancers.). (See Table S1 for diagnoses by calendar year.) All-site cancer cases included prostate (1,301 cases, 34%), breast (378 cases, 10%), and colorectal (362 cases, 9%); other cases counted for less than 8% each (Table S2).

Associations between Long-Term Exposure to Black Carbon and Cancer Incidence

All-site cancer. Using the single-pollutant model, BC exposure was significantly associated with increased cancer incidence in single-pollutant models (Table 3), with a hazard ratio (HR) of 1.17 [95% confidence interval (CI): 1.06, 1.29] for one IQR increase in ln-transformed cumulative BC exposure using the main model. Results were similar for models adjusted only for age, calendar time, and sex (Model 1), or models that also included smoking pack-years and passive smoking (Model 2). AICs were similar among the three models, but the smallest for the main model. Long-term $PM_{2.5}$ exposure was also significantly associated with all-site cancer, with a HR of 1.20 (95% CI: 1.06, 1.34) using the main model for one IQR increase in ln-transformed cumulative $PM_{2.5}$ exposure, but the association with long-term NO_2 exposure was smaller and statistically nonsignificant, with a HR of 1.05 (95% CI: 0.97, 1.14) for one IQR increase in ln-transformed cumulative NO_2. The BC residual approach (using the residuals of the cumulative BC exposure regressed on the cumulative $PM_{2.5}$ exposure) yielded a HR of 1.05 (95% CI: 1.00, 1.11) for an IQR increase of the residuals. This means that, holding total $PM_{2.5}$ constant, an increase in BC and closely linked constituents (and consequently, a decrease in other $PM_{2.5}$ constituents) was associated with a nonsignificant increase in the risk of incident all-site cancer. We estimated a similar and significant HR of 1.05 (95% CI: 1.02, 1.10) using the residuals of BC regressed on NO_2. AICs were similar for fully adjusted models of all-site cancer in relation to the different exposure variables, but slightly smaller for the model of cumulative BC exposure.

The association between cumulative BC and all-site cancer incidence (single-pollutant model) was similar when we implemented a 2-y lag period instead of a 10-y lag period (Table S3). The association between cumulative BC and all-site cancer was also similar in sensitivity analyses with further adjustment for the deprivation index and when restricting our analysis to address-level geocoded participants (1,664 cases vs. 3,711 cases), but stronger when restricting our study population to the participants followed at least 20 y (i.e., with a minimal exposure window of 10 y, 1,962 cases, HR 1.40; 95% CI: 1.20, 1.64 for one IQR increase) (Figure 2, Table S4). Using a missing data category or imputing missing data as the median/mode provided similar associations, whereas using the complete cases data did not yield any substantial association. Using a two-piece linear model, we found the most optimal cut-off at 24 $10^{-5}/m$ of cumulative black carbon; below and above this cut-off, we estimated HRs of 1.22 (95% CI: 1.07, 1.40) and of 1.08 (95% CI: 1.00, 1.17), respectively.

The main association between long-term BC exposure and all-site cancer was also stable in population subsets defined by sex, smoking status, and distance to the nearest major road (Figure 2; Table S4). We estimated slightly different associations

Table 1. Baseline characteristics in 1989 (unless stated otherwise) of the study population for all 19,348 participants and according to their all-site cancer status over the period 1999–2015.

	Individuals with a diagnosed cancer (all-site):		
	No	Yes	*p*-Value
Number	15,637	3,711	
Follow-up time (y)	27.0 [27.0, 27.0]	20.3 [17.1, 23.7]	<0.001
Age (y)	43.5 [41.0, 46.5]	44.5 [42.0, 47.0]	<0.001
Sex			<0.001
Male	11,147 (71.3)	2,956 (79.7)	
Female	4,490 (28.7)	755 (20.3)	
Smoking status			<0.001
Never smoker	6,859 (44.2)	1,425 (38.8)	
Former smoker	4,442 (28.6)	1,075 (29.2)	
Current smoker	4,210 (27.1)	1,177 (32.0)	
Unknown	126	34	
Cumulative pack-years[a]	15.0 [7.5, 25.7]	17.7 [8.5, 29.6]	<0.001
Unknown	361	92	
Passive smoking			0.93
Yes	10,623 (78.1)	2,507 (78.0)	
No	2,980 (21.9)	707 (22.0)	
Unknown	2,034	497	
Education			0.136
≤11 y	11,238 (73.5)	2,720 (74.8)	
12–13 y	1,144 (7.5)	227 (6.2)	
14–15 y	887 (5.8)	207 (5.7)	
Other secondary education	1,668 (10.9)	398 (10.9)	
Other diploma	357 (2.3)	83 (2.3)	
Unknown	343	76	
Occupational exposure[b]			<0.001
None	9,443 (60.4)	2,100 (56.6)	
One	1,319 (8.4)	328 (8.8)	
Two	1,697 (10.9)	448 (12.1)	
Three or more	3,178 (20.3)	835 (22.5)	
Alcohol use			<0.001
Abstinent	391 (2.5)	104 (2.8)	
Light drinker	11,170 (71.4)	2,421 (65.3)	
Moderate drinker	2,386 (15.3)	654 (17.6)	
Heavy drinker	1,370 (8.8)	448 (12.1)	
Unclear pattern	317 (2.0)	83 (2.2)	
Unknown	3	1	
Family status (not single)			0.09
Not single	13,722 (89.0)	3,288 (90.0)	
Single	1,696 (11.0)	366 (10.0)	
Unknown	219	57	
Socioeconomic status			0.011
Low (blue-collar workers or clerks)	2,762 (17.7)	621 (16.7)	
Intermediate (first-line supervisors)	9,179 (58.8)	2,130 (57.4)	
High (management)	3,677 (23.5)	958 (25.8)	
Unknown	19	2	
Body mass index[c]	24.3 [22.3, 26.3]	24.7 [22.8, 26.7]	<0.001
Unknown	2,253	533	
Fruit and vegetable consumption			0.123
Never or less than once a week	81 (0.7)	27 (1.0)	
Once or twice a week	794 (7.0)	211 (7.7)	
More than twice a week, not every day	2,562 (22.6)	641 (23.5)	
Every day or almost	7,912 (69.7)	1,851 (67.8)	
Unknown	4,288	981	
Distance to the nearest major road (km)[d]	0.8 [0.3, 1.6]	0.8 [0.3, 1.6]	0.407
Deprivation index[e]			0.983
High	4,994 (33.3)	1,211 (33.5)	
Intermediate	4,986 (33.3)	1,199 (33.1)	
Low	5,007 (33.4)	1,209 (33.4)	
Unknown	650	92	
Urban classification[f]			0.720
Solely urban	3,911 (25.0)	926 (25.0)	
Solely semiurban	3,713 (23.7)	878 (23.7)	
Solely rural	2,921 (18.7)	722 (19.5)	
Mixed	5,092 (32.6)	1,185 (31.9)	

Note: Number (percentage) or median (25th and 75th percentiles). All-site cancer cases were defined as the whole ICD-10 chapter except C77–79 (secondary malignant neoplasms) and C44 (nonmelanoma skin cancers). Participants were excluded from the analysis if they were diagnosed with cancer before 1999.
[a]Calculated only for current and former smokers.
[b]To nine lung carcinogens over the whole employment period.
[c]In 1990.
[d]Updated annually. Median distance over the follow-up.
[e]Calculated only for the residence in 2009 and for participants who were still alive then. Participants with "unknown" status were those who died before 2009.
[f]Urban classification obtained only in 1989, 2000, and 2010.

Table 2. Exposure characteristics of the study population for all 19,348 participants of the Gazel cohort and according to their cancer status (all-site cancer) over the period 1999–2015.

Number of participants		All individuals				Individuals with a diagnosed cancer (all-site):								
		19,348				No				Yes				
						15,637				3,711				
Annual mean		Median	25th percentile	75th percentile	IQR	Median	25th percentile	75th percentile	IQR	Median	25th percentile	75th percentile	IQR	p-Value[a]
Baseline	Black carbon (10^{-5}/m)	2.65	2.23	3.33	1.10	2.65	2.23	3.34	1.11	2.63	2.22	3.27	1.05	0.351
	$PM_{2.5}$ ($\mu g/m^3$)	33.3	30.1	37.8	7.7	33.3	30.1	37.8	7.7	33.2	29.9	37.7	7.8	0.075
	NO_2 ($\mu g/m^3$)	39.3	28.4	52.4	24.0	39.5	28.4	52.6	24.2	38.9	28.2	51.8	23.6	0.073
	Black carbon residuals													
	Regressed on $PM_{2.5}$	0.00	−0.13	0.13	0.26	0.00	−0.13	0.13	0.26	−0.01	−0.13	0.13	0.26	0.858
	Regressed on NO_2	−0.20	−0.28	−0.13	0.15	−0.20	−0.28	−0.13	0.15	−0.20	−0.27	−0.12	0.15	0.037
At censoring	Black carbon (10^{-5}/m)	1.82	1.52	2.34	0.82	1.77	1.49	2.26	0.77	2.04	1.67	2.61	0.94	<0.001
	$PM_{2.5}$ ($\mu g/m^3$)	21.1	18.9	24.7	5.7	20.8	18.7	23.8	5.1	23.9	20.1	28.8	8.7	<0.001
	NO_2 ($\mu g/m^3$)	26.0	18.3	36.2	17.9	25.3	17.9	35.0	17.1	29.5	20.2	41.5	21.3	<0.001
	Black carbon residuals													
	Regressed on $PM_{2.5}$	−0.04	−0.17	0.09	0.26	−0.05	−0.18	0.09	0.27	−0.01	−0.15	0.14	0.29	<0.001
	Regressed on NO_2	0.07	−0.03	0.17	0.20	0.08	−0.01	0.17	0.18	0.02	−0.09	0.13	0.22	<0.001

Note: All-site cancer cases were defined as the whole ICD-10 chapter except C77–79 (secondary malignant neoplasms) and C44 (nonmelanoma skin cancers). Participants were excluded from the analysis if they were diagnosed with cancer before 1999. Exposures were lagged 10 y. IQR: Interquartile range.
[a]p-Value of Wilcoxon tests for differences between participants diagnosed with all-site cancer or not.

according to the population density with higher HRs among rural and semiurban than among urban participants. For $PM_{2.5}$, the main association was also stable in population subsets, with a slightly decreasing values from urban to rural participants, contrarily to the results of BC.

Lung cancer. We estimated a positive association between long-term BC exposure and lung incident cancer with a HR of 1.31 (95% CI: 0.93, 1.83) for one IQR increase in ln-transformed cumulative BC exposure, supported by the BC residual approach that yielded a significant HR of 1.24 (95% CI: 1.05, 1.47) for one IQR increase (Table 3). This means that, holding total $PM_{2.5}$ constant, we estimated an increased risk of incident lung cancer for an increased content in BC. In comparison with the main model estimate, the association between lung cancer and one IQR increase in ln-transformed cumulative BC was positive but weaker when adjusted only for age, time, and sex (Model 1, HR 1.19; 95% CI: 0.87, 1.63) and was weakest for Model 2 (with additional adjustment for cumulative pack-years of smoking and secondhand smoking, HR 1.07; 95% CI: 0.77, 1.47) (Table 3). AICs were similar, but smallest for the main model. Long-term $PM_{2.5}$ exposure was not associated with lung cancer, with a HR of 1.01 (95% CI: 0.68, 1.51) for one IQR increase in ln-transformed cumulative $PM_{2.5}$ exposure; long-term NO_2 exposure was also not associated with lung cancer, with an HR of 1.06 (95% CI: 0.88, 1.27) for one IQR increase of the cumulative NO_2 exposure. However, the residuals of BC regressed on NO_2 were associated with an increased risk of lung cancer, with a HR of 1.14 (95% CI: 1.01, 1.28). This means that, holding NO_2 levels constant, we estimated an increased risk of incident lung cancer for increased levels of BC. AICs were similar for fully adjusted models of lung cancer in relation to the different exposure variables, but AICs were smallest for the models of BC residuals.

Associations between lung cancer and cumulative BC and BC residuals were positive but closer to the null when exposures were lagged for 2 y instead of 10 y (e.g., for black carbon: HR 1.21; 95% CI: 0.85, 1.71, compared with HR of 1.31 for a 10-y lag) (Table S3). For cumulative $PM_{2.5}$, the association with lung cancer was stronger but still nonsignificant when exposures were lagged for 2 y (HR 1.08; 95% CI: 0.68, 1.72, compared with HR 1.01 for a 10-y lag). The estimated associations between lung cancer and cumulative BC or cumulative $PM_{2.5}$ were similar to the main model when additionally adjusted for the deprivation index, but HRs were stronger when the analysis was restricted to participants with address-level geocodes (110 vs. 349 cases, HR 1.66; 95% CI: 0.85, 3.24 for BC) and to those with $\geq$20 y of follow-up (meaning a minimal exposure window of 10 y) (225 cases, HR 1.67; 95% CI: 1.05, 2.66 for BC) (Figure 3, Table S5). Associations were also stronger when missing covariate data were modeled using a missing indicator category and when missing data were imputed as the mean value, but the association became inverse when based on a complete-case analysis (101 cases, HR 0.71; 95% CI: 0.35, 1.43 for BC).

Discussion

In the French Gazel cohort, a predominantly male population of French adults who were employees of the national gas and energy company, long-term BC exposure was positively associated with incident all-site cancer and lung cancer based on single-pollutant models of cumulative exposure, and on models of BC residuals used to estimate the effect of BC as a $PM_{2.5}$ constituent, while holding the effect of total $PM_{2.5}$ constant. Associations were also positive for BC residuals regressed on NO_2. In general, results were robust to sensitivity analyses, and we found no substantial effect modification by sex, smoking, or urban classification.

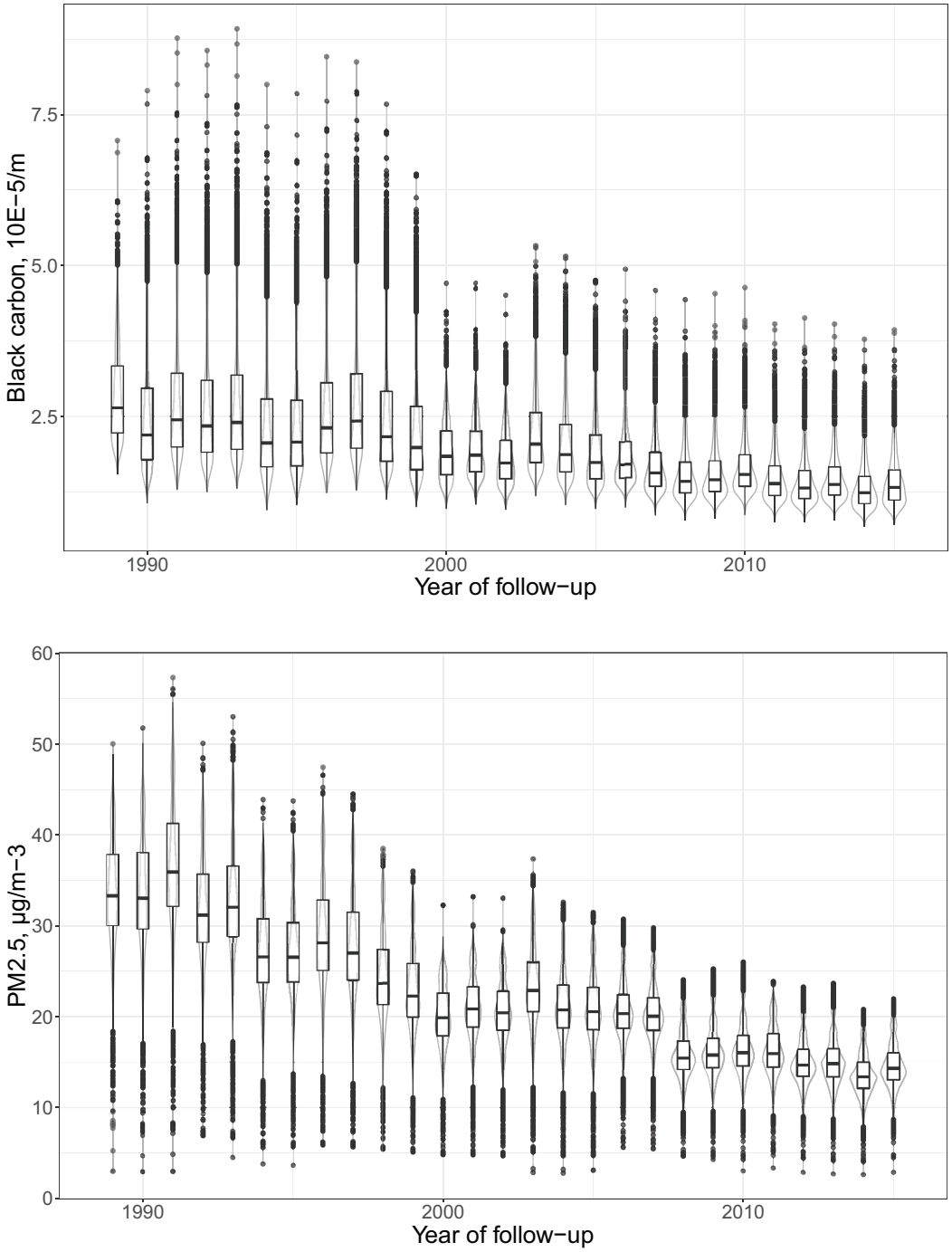


Figure 1. Black carbon (top) and $PM_{2.5}$ (bottom) concentrations between 1989 and 2015 at residential addresses of 19,348 participants of the French Gazel cohort. Black carbon (10^{-5}/m) and $PM_{2.5}$ ($\mu g/m^3$) are depicted by yearly boxplots in black (minimum, 25th percentile, median, 75th percentile, outliers calculated as 75th percentile plus 1.5 times the interquartile range, and maximum) and violin plots in gray (two rotated kernel density plots depicting the probability of each exposure level and informing on the skewedness of the distribution).

The Gazel cohort provides a relatively large and well characterized study population with many variables collected throughout the follow-up. BC, $PM_{2.5}$, and NO_2 exposures were estimated using a validated LUR model, at fine scale and over a long time period using an accepted methodology to back- and forward extrapolate exposures between 1990 and 2015. We used time-dependent Cox models. We handled confounding by total $PM_{2.5}$ exposure or by NO_2 via a residual approach that yielded findings corroborating the inference that BC could at least partially explain the health effects of total $PM_{2.5}$ and also seemingly independently from NO_2. Though not always statistically significant, the analyses generally yielded point estimates above the null, and the analysis using the residuals of black carbon regressed on $PM_{2.5}$ or on NO_2 provided significant associations. Further, associations with all cancers were similar when we implemented a 2-y lag instead of a 10-y lag, but HRs for lung cancer and BC (cumulative and residuals) were closer to the null when exposures were lagged by only 2 y, and confidence intervals were wider than for corresponding estimates using a 10-y lag. This 2-y lag could have induced more misclassification, therefore yielding point estimates closer to the null; we hypothesize that

Table 3. Associations between black carbon, black carbon residuals, and $PM_{2.5}$ with incident all-site and lung cancer in the Gazel cohort as HR and their 95% CI.

Exposure	HR	CI	Cases	Person-years	*p*-Value	AIC
All cancers						
Cumulative black carbon[a]						
Model 1	1.16	1.06, 1.27	3,711	293,210	0.002	66,056
Model 2	1.15	1.05, 1.26			0.003	65,991
Model 3	1.17	1.06, 1.29			0.001	65,978
Black carbon residuals[b]						
Regressed on $PM_{2.5}$	1.05	1.00, 1.11			0.056	65,984
Regressed on NO_2	1.05	1.02, 1.10			0.005	65,980
Cumulative $PM_{2.5}$[c]	1.20	1.06, 1.34			0.003	65,978
Cumulative NO_2[d]	1.05	0.97, 1.14			0.206	65,986
Lung cancer						
Cumulative black carbon[a]						
Model 1	1.19	0.87, 1.63	349	254,154	0.275	6,140
Model 2	1.07	0.77, 1.47			0.694	5,880
Model 3	1.31	0.93, 1.83			0.118	5,842
Black carbon residuals[b]						
Regressed on $PM_{2.5}$	1.24	1.05, 1.47			0.011	5,838
Regressed on NO_2	1.14	1.01, 1.28			0.030	6,013
Cumulative $PM_{2.5}$[c]	1.01	0.68, 1.51			0.942	5,845
Cumulative NO_2[e]	1.06	0.88, 1.27			0.568	6,018

Note: All-site cancer cases were defined as the whole ICD-10 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 y. Cox model with attained age as underlying time-scale. Missing covariate data were imputed using MICE. Model-based estimates were pooled following Rubin's rules. Model 1: adjusted for sex, calendar time, age at inclusion. Model 2: Model 1 plus cumulative smoking pack-years and passive smoking. Model 3 (main model): Model 2 plus alcohol use, body mass index, education, socioeconomic status, family status, fruit and vegetable consumption, occupational exposure to nine lung carcinogens. AIC. Akaike Information Criterion; CI, confidence interval; HR, hazard ratio.
[a]HRs are for one IQR increase in ln-transformed cumulative black carbon (IQR = 1.0).
[b]HRs are for IQR increases in ln-transformed cumulative black carbon residuals regressed on ln-transformed cumulative $PM_{2.5}$ (IQR = 0.3) or on ln-transformed cumulative NO_2 (IQR = 0.2), respectively. Adjusted for Model 3 covariates.
[c]HRs are for an IQR increase in ln-transformed cumulative $PM_{2.5}$ (IQR = 0.9) adjusted for Model 3 covariates.
[d]HRs are for an IQR increase in ln-transformed cumulative NO_2 (IQR = 1.2) adjusted for Model 3 covariates.
[e]HRs are for an IQR increase in cumulative NO_2 (IQR = 300 $\mu g/m^3$) adjusted for Model 3 covariates.

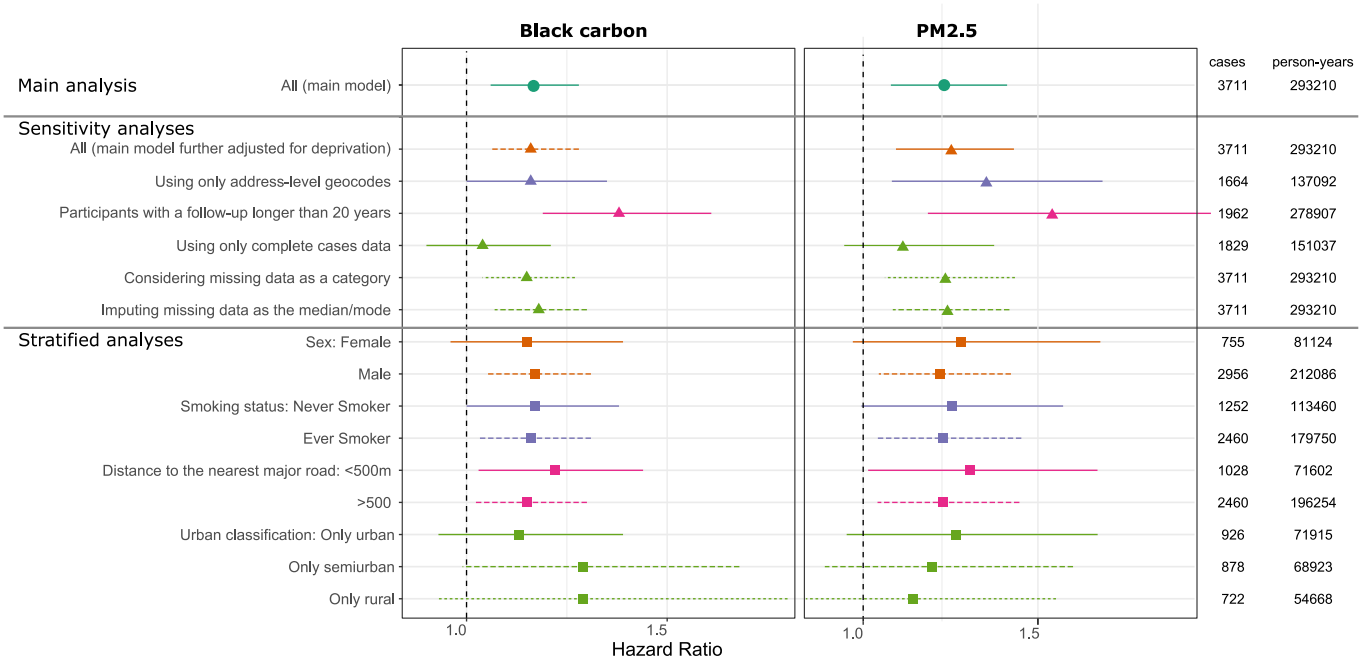


Figure 2. Associations between cumulative black carbon (left) and $PM_{2.5}$ (right) and all-site incident cancer in the main, sensitivity, and stratified analyses in the Gazel cohort, with the number of identified cancer cases among the number of participant-year over the follow-up. Hazard ratios and confidence intervals expressed for one IQR increase in ln-transformed cumulative exposure to black carbon or $PM_{2.5}$ 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 y. Participants were excluded from the analysis if they were diagnosed with cancer before 1999. See Table S4 for corresponding numeric data. Unless specified otherwise, these model-based estimates were computed using MICE to address missing data and were pooled following Rubin's rules. Note: BMI, body mass index; IQR, interquartile range.

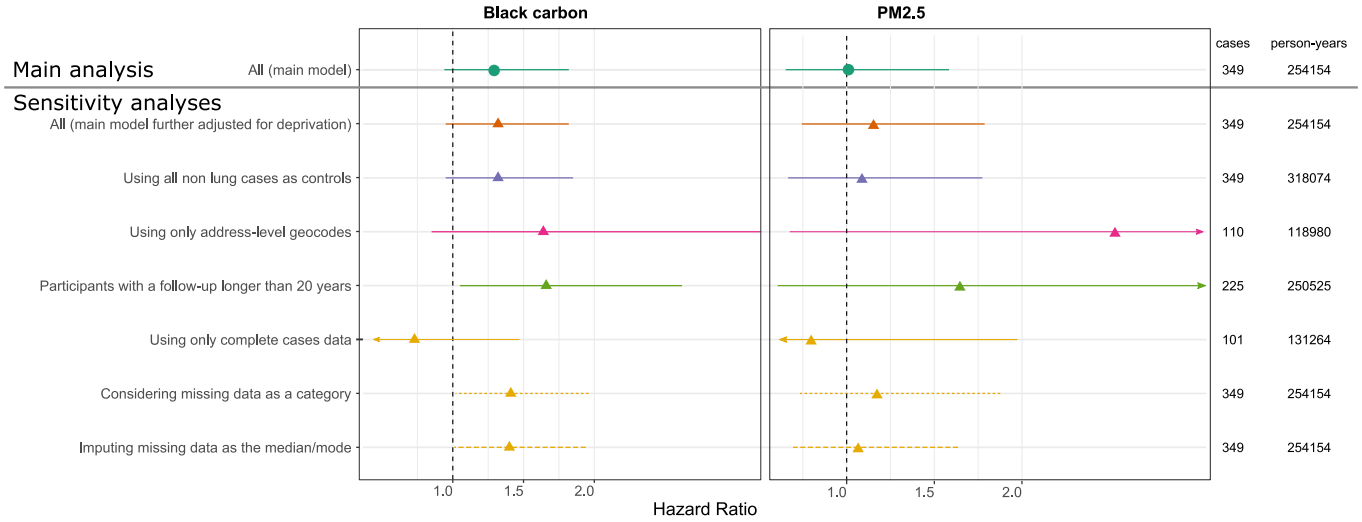


Figure 3. Associations between cumulative black carbon (left) and $PM_{2.5}$ (right) and lung incident cancer in the main and sensitivity analyses in the Gazel cohort, with the number of identified cancer cases among the number of participant-year over the follow-up. Hazard ratios and confidence intervals expressed for one IQR increase in ln-transformed cumulative exposure to black carbon or $PM_{2.5}$ 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 y. Participants were excluded from the analysis if they were diagnosed with cancer before 1999. See Table S5 for corresponding numeric data. Unless specified otherwise, these model-based estimates were computed using MICE to address missing data and were pooled following Rubin's rules. Note: BMI, body mass index; IQR, interquartile range.

2 y could be too short a latency period for lung cancer. In addition, the residual approach needs a careful interpretation: higher levels of BC mean higher levels of components correlated with BC, and thus, at constant total $PM_{2.5}$, this implies lower levels of all other components that are not correlated with BC. Considering this, for all-site cancer, we found that the risk estimate yielded using the residuals regressed on $PM_{2.5}$ was not significant, contrarily to the one using BC. This suggests that the association using BC was probably confounded by total $PM_{2.5}$ or by co-occurring components in $PM_{2.5}$. On the contrary, in the lung analysis, unexpectedly we found no association with $PM_{2.5}$, but we did find an association with BC (although still nonsignificant), and a significant and substantial association using the residuals of BC; one hypothesis would be that BC and its co-occurring components might be suitable chemicals to at least partly explain the association between outdoor air pollution and lung cancer. These co-occurring chemicals may come from the same sources as BC and constitute primary components of $PM_{2.5}$ such as metals, as opposed to secondary components, such as secondary aerosols made of, e.g., sulfates or nitrates. We found significant associations using the residuals of BC regressed on NO_2, suggesting that the associations found for BC were not confounded by NO_2. Despite providing novel evidence on BC exposure and incident cancer, our approaches still did not allow for fully disentangling the effects of BC, its correlated pollutants, and total $PM_{2.5}$.

The BC and $PM_{2.5}$ measurements used in the LUR model came from two different measurement networks. Yet we think this would not likely hamper our findings, because we found a strong correlation between the two exposure estimates—as it is expected from literature—and reasonable levels for BC. The association between cumulative BC and all cancers did not appear to differ by distance to the nearest major road (<500 m vs. ≥500 m). A slightly stronger association among semiurban and rural dwellers than among urban dwellers might reflect differences in BC sources, particle sizes, or coexposures, but may also be explained by random variation.

Research on health impacts of ambient BC is quite recent. In 2013, the World Health Organization's (WHO) review of evidence on health aspects of air pollution (WHO 2013) concluded that annual exposure to BC could be involved in cardiovascular diseases and cardiopulmonary mortality. Studies published after this review have corroborated the above-mentioned conclusions, notably for cardiovascular diseases (Kirrane et al. 2019), and started unraveling the effects of long-term exposure to BC on respiratory health (Aaron et al. 2019; Mordukhovich et al. 2015; Rice et al. 2016; Wang et al. 2019; Yang et al. 2018). In their review on the likeliness of $PM_{2.5}$ constituents to cause adverse health effects, Yang et al. (2019) found only two studies dealing with long-term exposures to BC or organic carbon and respiratory mortality, which prevented the authors from conducting a meta-analysis. Indeed, only a few epidemiological studies focused on residential outdoor long-term exposure to BC. To our knowledge, none dealt with incident cancer. In their review, Grahame et al. (2014) suggested BC was related to lung cancer mortality, according to studies investigating diesel effects. Hvidtfeldt et al. (2019) found no statistically significant association between lung cancer mortality and BC with a point estimate close to the null in a population-based and gender-balanced Danish cohort of approximately 50,000 participants with 1,223 cases of lung cancer out of 10,193 total cases. In most articles and as highlighted by Luben et al. (2017) for cardiovascular mortality, the main challenge consists of disentangling the effects of BC and those of total $PM_{2.5}$ (and we think this remains true whatever the health outcome of interest). Because BC is part of $PM_{2.5}$, the above-described associations with health outcomes intertwine with those of total $PM_{2.5}$. Additionally, the WHO warned that the epidemiological findings on BC may not reflect BC's effects by itself, but that BC estimates could proxy harmful co-occurring combustion-derived chemicals (WHO Regional Office for Europe 2013). Indeed, BC is more influenced by small-scale sources than $PM_{2.5}$ is and thereby more linked to local traffic and wood burning.

BC could trigger oxidative stress, inflammation [especially lung inflammation as described in ANSES (2019)], and DNA methylation (Niranjan and Thakur 2017), by itself or due to its co-occurrence with other forms of carbon or with metals. Chemical analyses of different sizes of $PM_{2.5}$ revealed that BC was found in all size bins (Bein et al. 2005); yet black carbon

seems mostly concentrated in the finer fractions, that is particulate matter with a diameter <1μm or even ultrafine particles with a diameter <0.1μm (Li et al. 2003; Pérez et al. 2010), those which go deeper into the lungs. Diesel ultrafine particles include mostly BC, which can enter lung cells *in vitro* (Steiner et al. 2016). These very fine particles may reach blood circulation; but the capability of BC to do so remains only partly understood, as shown in observational and experimental studies (Saenen et al. 2017; Shimada et al. 2006). On the other hand, metals contained in such tiny particles (which are often correlated with BC) may also enter this circulation and trigger the biological effects above mentioned (Nakane 2012).

More measurements and fine scale exposure models are needed to support studies such as ours in other areas, to increase understanding of how BC may affect health and cancer, and to provide data for policy makers. As suggested by the WHO, BC could act as a better indicator of exposure to combustion sources than $PM_{2.5}$ does. As many other studies do, this study emphasizes the need for disentangling the effects of $PM_{2.5}$ as a mass, as the sum of various constituents such as BC, and as the sum of various size bins (such as ultrafine particles). In particular, we need more tools to do so, including methods complementing the use of residuals.

In Gazel, as in most observational studies on outdoor air pollution, we could not obtain exposure data for individuals at work or during their commutes. But even if we could not obtain a complete lifetime exposure with left-censored participants, we still obtained up to 16 y of air pollution exposure at residential addresses, taking into account any residential address change during the follow-up. In addition, in our sensitivity using participants with at least 20 y of follow-up (i.e., a minimal exposure window of 10 y), we estimated associations between cumulative BC and both all-site and lung cancer with higher point estimates and lower *p*-values. Although we used three different sources to identify cancer cases, 85% were identified through one of them: the French national health administrative databases. Furthermore, 80% of cases were identified after 2007, when there was overlap among all sources, and the case identification was consistent between the three sources (national databases, company health records, and self-reported diagnoses). The number of all-site cancer cases allowed for a satisfying statistical power, but the estimated associations with air pollutants may be driven by the associations between air pollutants and the most frequently identified cancer sites in our study population (prostate and breast). The small number of lung cancer cases considerably reduced our statistical power in the sensitivity analyses and prevented us from conducting stratified analyses for lung cancer. In addition, for lung cancer, the model adjusted for age, sex, time, and smoking characteristics yielded a smaller and nonsignificant association. Smoking is a major risk factor for lung cancer yet can be confounded by other covariates such as SES, which may explain this inconsistent result. For both cancer types, due to the amount of missing data in each of the covariates included in the main model, using complete cases only led to selecting half of the study population, yielding very different results as the models using imputed data (with three different imputation methods) and very likely biased estimates. We are more confident in the results using multiple imputations, as generally recommended (Raghunathan 2004; White and Carlin 2010). Although the Gazel cohort recruited participants with a range of occupations and incomes, it was not representative of the general population. Additional studies are needed to confirm the findings of this study in other populations and locations.

Conclusions

In this study on BC and the risk of incident cancer in a predominantly male population of French adults, we found positive associations between BC and incident all-site and lung cancers that were generally consistent across sensitivity analyses. Further, we used a residual approach to theoretically isolate the effect of BC from that of $PM_{2.5}$, and it also provided positive associations. Therefore, this study suggests BC could at least partly explain the carcinogenic effects of outdoor air pollution. Further studies on the long-term associations between cancer risk and BC exposure are warranted and, in addition to seeking to replicate our findings, should include more approaches to disentangle the effects of BC from its co-occurring chemicals and total $PM_{2.5}$.

Acknowledgments

The authors thank the ARC Foundation, which supported this study (OCAPOL, CANCAIR-201501234). The authors are grateful to S. Bonenfant (INSERM), who reconstituted the residential histories of Gazel participants, and to V. Wagner (Santé publique France), who helped design the multiple imputations.

References

Aaron CP, Hoffman EA, Kawut SM, Austin JHM, Budoff M, Michos ED, et al. 2019. Ambient air pollution and pulmonary vascular volume on computed tomography: the MESA air pollution and lung cohort studies. Eur Respir J 53(6):1802116, PMID: 31167881, https://doi.org/10.1183/13993003.02116-2018.

Anenberg SC, Schwartz J, Shindell D, Amann M, Faluvegi G, Klimont Z, et al. 2012. Global air quality and health co-benefits of mitigating near-term climate change through methane and black carbon emission controls. Environ Health Perspect 120(6):831–839, PMID: 22418651, https://doi.org/10.1289/ehp.1104301.

ANSES. 2019. Particulate matter in ambient air health effects according to components, sources and particle size. https://www.anses.fr/en/system/files/AIR2014SA0156RaEN.pdf [accessed 31 January 2020].

Bein KJ, Zhao Y, Wexler AS, Johnston MV. 2005. Speciation of size-resolved individual ultrafine particles in Pittsburgh, Pennsylvania. J Geophys Res Atmos 110, https://doi.org/10.1029/2004JD004708.

Beelen R, Hoek G, Raaschou-Nielsen O, Stafoggia M, Andersen ZJ, Weinmayr G, et al. 2015. Natural-cause mortality and long-term exposure to particle components: an analysis of 19 European cohorts within the multi-center ESCAPE project. Environ Health Perspect 123(6):525–533, PMID: 25712504, https://doi.org/10.1289/ehp.1408095.

Brandt J, Silver JD, Frohn LM, Geels C, Gross A, Hansen AB, et al. 2012. An integrated model study for Europe and North America using the Danish Eulerian Hemispheric Model with focus on intercontinental transport of air pollution. Atmos Environ 53:156–176, https://doi.org/10.1016/j.atmosenv.2012.01.011.

Chylek P, Jennings SG, Pinnick R. 2015. AEROSOLS | Soot. In: *Encyclopedia of Atmospheric Sciences*. North GR, Pyle J, Zhang F, eds. 2nd ed. Oxford, UK: Academic Press, 86–91.

Colicino E, Wilson A, Frisardi MC, Prada D, Power MC, Hoxha M, et al. 2017. Telomere length, Long-term black carbon exposure, and cognitive function in a cohort of older men: the VA Normative Aging Study. Environ Health Perspect 125(1):76–81, PMID: 27259001, https://doi.org/10.1289/EHP241.

de Hoogh K, Chen J, Gulliver J, Hoffmann B, Hertel O, Ketzel M, et al. 2018. Spatial $PM_{2.5}$, NO_2, O_3 and BC models for Western Europe - evaluation of spatiotemporal stability. Environ Int 120:81–92, PMID: 30075373, https://doi.org/10.1016/j.envint.2018.07.036.

Goldberg M, Chevalier A, Imbernon E, Coing F, Pons H. 1996. The epidemiological information system of the French national electricity and gas company: the SIEPI project. Med Lav 87(1):16–28, PMID: 8699978.

Goldberg M, Leclerc A, Zins M. 2015. Cohort profile update: the GAZEL cohort study. Int J Epidemiol 44(1):77–77g, PMID: 25422284, https://doi.org/10.1093/ije/dyu224.

Grahame TJ, Klemm R, Schlesinger RB. 2014. Public health and components of particulate matter: the changing assessment of black carbon. J Air Waste Manag Assoc 64(6):620–660, PMID: 25039199, https://doi.org/10.1080/10962247.2014.912692.

Hvidtfeldt UA, Sørensen M, Geels C, Ketzel M, Khan J, Tjønneland A, et al. 2019. Long-term residential exposure to $PM_{2.5}$, black carbon, NO_2, and ozone and mortality in a Danish cohort. Environ Int 123:265–272, PMID: 30551059, https://doi.org/10.1016/j.envint.2018.12.010.

IARC. 2016. IARC Monographs on The Evaluation of Carcinogenic Risks to Humans, Volume 109, 2016. http://monographs.iarc.fr/ENG/Monographs/vol109/index.php [accessed 22 January 2020].

Imbernon E. 1991. MATEX: une matrice emplois-expositions destinée à la surveillance épidémiologique des travailleurs d'une grande entreprise (E.D.F.-G.D.F.). Archives Des Maladies Professionnelles et de Médecine du Travail 52:559–566.

Kirrane EF, Luben TJ, Benson A, Owens EO, Sacks JD, Dutton SJ, et al. 2019. A systematic review of cardiovascular responses associated with ambient black carbon and fine particulate matter. Environ Int 127:305–316, PMID: 30953813, https://doi.org/10.1016/j.envint.2019.02.027.

Koelmans AA, Jonker MTO, Cornelissen G, Bucheli TD, Van Noort PCM, Gustafsson Ö. 2006. Black carbon: the reverse of its dark side. Chemosphere 63(3):365–377, PMID: 16226291, https://doi.org/10.1016/j.chemosphere.2005.08.034.

Li N, Sioutas C, Cho A, Schmitz D, Misra C, Sempf J, et al. 2003. Ultrafine particulate pollutants induce oxidative stress and mitochondrial damage. Environ Health Perspect 111(4):455–460, PMID: 12676598, https://doi.org/10.1289/ehp.6000.

Loomis D, Grosse Y, Lauby-Secretan B, Ghissassi FE, Bouvard V, Benbrahim-Tallaa L, et al. 2013. The carcinogenicity of outdoor air pollution. Lancet Oncol 14(13):1262–1263, PMID: 25035875, https://doi.org/10.1016/S1470-2045(13)70487-X.

Luben TJ, Nichols JL, Dutton SJ, Kirrane E, Owens EO, Datko-Williams L, et al. 2017. A systematic review of cardiovascular emergency department visits, hospital admissions and mortality associated with ambient black carbon. Environ Int 107:154–162, PMID: 28735152, https://doi.org/10.1016/j.envint.2017.07.005.

Mordukhovich I, Lepeule J, Coull BA, Sparrow D, Vokonas P, Schwartz J. 2015. The effect of oxidative stress polymorphisms on the association between long-term black carbon exposure and lung function among elderly men. Thorax 70(2):133–137, PMID: 25414198, https://doi.org/10.1136/thoraxjnl-2014-206179.

Mostofsky E, Schwartz J, Coull BA, Koutrakis P, Wellenius GA, Suh HH, et al. 2012. Modeling the association between particle constituents of air pollution and health outcomes. Am J Epidemiol 176(4):317–326, PMID: 22850792, https://doi.org/10.1093/aje/kws018.

Mumford JL, Helmes CT, Lee XM, Seidenberg J, Nesnow S. 1990. Mouse skin tumorigenicity studies of indoor coal and wood combustion emissions from homes of residents in Xuan Wei, China with high lung cancer mortality. Carcinogenesis 11(3):397–403, PMID: 2311182, https://doi.org/10.1093/carcin/11.3.397.

Nakane H. 2012. Translocation of particles deposited in the respiratory system: a systematic review and statistical analysis. Environ Health Prev Med 17(4):263–274, PMID: 22101916, https://doi.org/10.1007/s12199-011-0252-8.

Niranjan R, Thakur AK. 2017. The toxicological mechanisms of environmental soot (black carbon) and carbon black: focus on oxidative stress and inflammatory pathways. Front Immunol 8: PMID: 28713383, https://doi.org/10.3389/fimmu.2017.00763.

Ostro B, Reynolds P, Goldberg D, Hertz A, Burnett RT, Shin H, et al. 2011. Assessing long-term exposure in the California Teachers Study. Environ Health Perspect 119(6):A242–A243, https://doi.org/10.1289/ehp.119-3114832.

Paunescu A-C, Casas M, Ferrero A, Pañella P, Bougas N, Beydon N, et al. 2019. Associations of black carbon with lung function and airway inflammation in schoolchildren. Environ Int 131:104984, PMID: 31301585, https://doi.org/10.1016/j.envint.2019.104984.

Pedersen M, Andersen ZJ, Stafoggia M, Weinmayr G, Galassi C, Sørensen M, et al. 2017. Ambient air pollution and primary liver cancer incidence in four European cohorts within the ESCAPE project. Environ Res 154:226–233, PMID: 28107740, https://doi.org/10.1016/j.envres.2017.01.006.

Pencina MJ, Larson MG, D'Agostino RB. 2007. Choice of time scale and its effect on significance of predictors in longitudinal studies. Stat Med 26(6):1343–1359, PMID: 16955538, https://doi.org/10.1002/sim.2699.

Pérez N, Pey J, Cusack M, Reche C, Querol X, Alastuey A, et al. 2010. Variability of particle number, black carbon, and PM10, and PM1 levels and speciation: influence of road traffic emissions on urban air quality. Aerosol Sci Technol 44(7):487–499, https://doi.org/10.1080/02786821003758286.

Prada D, Zhong J, Colicino E, Zanobetti A, Schwartz J, Dagincourt N, et al. 2017. Association of air particulate pollution with bone loss over time and bone fracture risk: analysis of data from two independent studies. Lancet Planet Health 1(8):e337–e347, PMID: 29527596, https://doi.org/10.1016/S2542-5196(17)30136-5.

Raaschou-Nielsen O, Andersen ZJ, Beelen R, Samoli E, Stafoggia M, Weinmayr G, et al. 2013. Air pollution and lung cancer incidence in 17 European cohorts: prospective analyses from the European Study of Cohorts for Air Pollution Effects (ESCAPE). Lancet Oncol 14(9):813–822, PMID: 23849838, https://doi.org/10.1016/S1470-2045(13)70279-1.

Raaschou-Nielsen O, Beelen R, Wang M, Hoek G, Andersen ZJ, Hoffmann B, et al. 2016. Particulate matter air pollution components and risk for lung cancer. Environ Int 87:66–73, PMID: 26641521, https://doi.org/10.1016/j.envint.2015.11.007.

Raghunathan TE. 2004. What do we do with missing data? Some options for analysis of incomplete data. Annu Rev Public Health 25:99–117, PMID: 15015914, https://doi.org/10.1146/annurev.publhealth.25.102802.124410.

Rey G, Jougla E, Fouillet A, Hémon D. 2009. Ecological association between a deprivation index and mortality in France over the period 1997–2001: variations with spatial scale, degree of urbanicity, age, gender and cause of death. BMC Public Health 9:33, PMID: 19161613, https://doi.org/10.1186/1471-2458-9-33.

Rice MB, Rifas-Shiman SL, Litonjua AA, Oken E, Gillman MW, Kloog I, et al. 2016. Lifetime exposure to ambient pollution and lung function in children. Am J Respir Crit Care Med 193(8):881–888, PMID: 26575800, https://doi.org/10.1164/rccm.201506-1058OC.

Rothman KJ, Greenland S, Lash TL. 2008. *Modern Epidemiology*. 3rd ed. Philadelphia, PA: Lippincott Williams & Wilkins.

Saenen ND, Bové H, Steuwe C, Roeffaers MBJ, Provost EB, Lefebvre W, et al. 2017. Children's urinary environmental carbon load. A novel marker reflecting residential ambient air pollution exposure? Am J Respir Crit Care Med 196(7):873–881, PMID: 28686472, https://doi.org/10.1164/rccm.201704-0797OC.

Shimada A, Kawamura N, Okajima M, Kaewamatawong T, Inoue H, Morita T. 2006. Translocation pathway of the intratracheally instilled ultrafine particles from the lung into the blood circulation in the mouse. Toxicol Pathol 34(7):949–957, PMID: 17178695, https://doi.org/10.1080/01926230601080502.

Steiner S, Bisig C, Petri-Fink A, Rothen-Rutishauser B. 2016. Diesel exhaust: current knowledge of adverse effects and underlying cellular mechanisms. Arch Toxicol 90(7):1541–1553, PMID: 27165416, https://doi.org/10.1007/s00204-016-1736-5.

Sunyer J, Esnaola M, Alvarez-Pedrerol M, Forns J, Rivas I, López-Vicente M, et al. 2015. Association between traffic-related air pollution in schools and cognitive development in primary school children: a prospective cohort study. PLoS Med 12(3):e1001792, PMID: 25734425, https://doi.org/10.1371/journal.pmed.1001792.

Therneau TM. 2015. *A Package for Survival Analysis in S*.

Therneau TM, Grambsch PM. 2000. *Modeling Survival Data: Extending the Cox Model*. New York, NY: Springer.

Tuppin P, Rudant J, Constantinou P, Gastaldi-Ménager C, Rachas A, de Roquefeuil L, et al. 2017. Value of a national administrative database to guide public decisions: from the système national d'information interrégimes de l'Assurance Maladie (SNIIRAM) to the système national des données de santé (SNDS) in France. Rev Epidemiol Sante Publique 65:S149–S167, PMID: 28756037, https://doi.org/10.1016/j.respe.2017.05.004.

van Buuren S, Groothuis-Oudshoorn K. 2011. Mice: multivariate imputation by chained equations in R. J Stat Soft 45(3):1–67, https://doi.org/10.18637/jss.v045.i03.

Wang M, Aaron CP, Madrigano J, Hoffman EA, Angelini E, Yang J, et al. 2019. Association between long-term exposure to ambient air pollution and change in quantitatively assessed emphysema and lung function. JAMA 322(6):546–556, PMID: 31408135, https://doi.org/10.1001/jama.2019.10255.

White IR, Carlin JB. 2010. Bias and efficiency of multiple imputation compared with complete-case analysis for missing covariate values. Stat Med 29(28):2920–2931, PMID: 20842622, https://doi.org/10.1002/sim.3944.

WHO (World Health Organization). 2013. Review of evidence on health aspects of air pollution – REVIHAAP Project: Technical Report. Copenhagen, Denmark: WHO Regional Office for Europe.

Wurth R, Kioumourtzoglou M-A, Tucker KL, Griffith J, Manjourides J, Suh H. 2018. Fine particle sources and cognitive function in an older Puerto Rican cohort in greater Boston. Environ Epidemiol 2: PMID: 30506018, https://doi.org/10.1097/EE9.0000000000000022.

Yang Y, Ruan Z, Wang X, Yang Y, Mason TG, Lin H, et al. 2019. Short-term and long-term exposures to fine particulate matter constituents and health: a systematic review and meta-analysis. Environ Pollut 247:874–882, PMID: 30731313, https://doi.org/10.1016/j.envpol.2018.12.060.

Yang Y, Tang R, Qiu H, Lai P-C, Wong P, Thach T-Q, et al. 2018. Long term exposure to air pollution and mortality in an elderly cohort in Hong Kong. Environ Int 117:99–106, PMID: 29730535, https://doi.org/10.1016/j.envint.2018.04.034.