



**HAL**  
open science

## **Influence of exposure assessment methods on associations between long-term exposure to outdoor fine particulate matter and risk of cancer in the French cohort Gazel**

Emeline Lequy, Mohammad Javad Zare Sakhvidi, Danielle Vienneau, Kees de Hoogh, Jie Chen, Jean-François Dupuy, Valérie Garès, Emilie Burte, Olivier Bouaziz, Sergey Zhivin, et al.

► **To cite this version:**

Emeline Lequy, Mohammad Javad Zare Sakhvidi, Danielle Vienneau, Kees de Hoogh, Jie Chen, et al.. Influence of exposure assessment methods on associations between long-term exposure to outdoor fine particulate matter and risk of cancer in the French cohort Gazel. *Science of the Total Environment*, 2022, 820, pp.article n°153098. 10.1016/j.scitotenv.2022.153098 . hal-03576594

**HAL Id: hal-03576594**

**<https://hal.science/hal-03576594v1>**

Submitted on 1 Apr 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

1 Influence of exposure assessment methods on associations between long-term exposure to  
2 outdoor fine particulate matter and risk of cancer in the French cohort Gazel

3 Authors: Emeline Lequy<sup>1,2\*</sup>, Mohammad Javad Zare Sakhvidi<sup>3</sup>, Danielle Vienneau<sup>4,5</sup>, Kees de  
4 Hoogh<sup>4,5</sup>, Jie Chen<sup>6</sup>, Jean-François Dupuy<sup>7</sup>, Valérie Garès<sup>7</sup>, Emilie Burte<sup>3</sup>, Olivier Bouaziz<sup>8</sup>,  
5 Alain Le Tertre<sup>9</sup>, Vèrène Wagner<sup>9</sup>, Ole Hertel<sup>10</sup>, Jesper Heile Christensen<sup>10</sup>, Sergey Zhivin<sup>1</sup>,  
6 Jack Siemiatycki<sup>2</sup>, Marcel Goldberg<sup>1</sup>, Marie Zins<sup>1</sup>, Bénédicte Jacquemin<sup>3</sup>

7 1: Unité "Cohortes en Population" UMS 011 Inserm/Université de Paris/Université Paris  
8 Saclay/UVSQ, Villejuif, France

9 2: Centre de recherche du centre hospitalier de l'université de Montréal, Québec, Canada

10 3: Univ Rennes, Inserm, EHESP, Irset (Institut de recherche en santé, environnement et travail)  
11 – UMR\_S 1085, F-35000 Rennes, France

12 4: Swiss Tropical and Public Health Institute, Basel, Switzerland.

13 5: University of Basel, Basel, Switzerland.

14 6: Utrecht University, Utrecht, the Netherlands

15 7: Univ Rennes, INSA, CNRS, IRMAR - UMR 6625, F-35000 Rennes, France

16 8 : MAP5, UMR CNRS 8145, Université de Paris, Paris, France

17 9: Santé publique France, Saint-Maurice, France

18 10: Dep. Env. Science, Aarhus University, Frederiksborgvej 399, 4000 Roskilde, Denmark

19

20 \*Address correspondence to:

21 Emeline Lequy, UMS 011, Hôpital Paul Brousse, 16 avenue Paul Vaillant Couturier, 94807

22 VILLEJUIF CEDEX, France. Email: [e.lequy@gmail.com](mailto:e.lequy@gmail.com). Phone number: +33 (0)1 77 74 74 23.

23

24 **Declaration of competing financial interests:**

25 The authors declare they have no actual or potential competing financial interests.

26 **Abstract**

27 **Background:** Many studies investigated the relationship between outdoor fine particulate  
28 matter (PM<sub>2.5</sub>) and cancer. While they generally indicated positive associations, results have  
29 not been fully consistent, possibly because of the diversity of methods used to assess exposure.

30 **Objectives:** To investigate how using different PM<sub>2.5</sub> exposure assessment methods influences  
31 risk estimates in the large French general population-based Gazel cohort (20,625 participants  
32 at enrollment) with a 26-year follow-up with complete residential histories.

33 **Methods:** We focused on two cancer incidence outcomes: all-sites combined and lung. We used  
34 two distinct exposure assessment methods: a western European land use regression (LUR),  
35 and a chemistry-dispersion model (Gazel-Air) for France, each with a time series  $\geq 20$ -years  
36 annual concentrations. Spearman correlation coefficient between the two estimates of PM<sub>2.5</sub>  
37 was 0.71 across all person-years; the LUR tended to provide higher exposures. We used  
38 extended Cox models with attained age as time-scale and time-dependent cumulative  
39 exposures, adjusting for a set of confounders including sex and smoking, to derive hazard  
40 ratios (HRs) and their 95% confidence interval, implementing a 10-year lag between exposure  
41 and incidence/censoring.

42 **Results:** We obtained similar two-piece linear associations for all-sites cancer (3,711 cases),  
43 with a first slope of HRs of 1.53 (1.24-1.88) and 1.43 (1.19-1.73) for one IQR increase of  
44 cumulative PM<sub>2.5</sub> exposure for the LUR and the Gazel-Air models respectively, followed by a  
45 plateau at around 1.5 for both exposure assessments. For lung cancer (349 cases), the HRs  
46 from the two exposure models were less similar, with largely overlapping confidence limits.

47 **Conclusion:** Our findings using long-term exposure estimates from two distinct exposure  
48 assessment methods corroborate the association between air pollution and cancer risk.

49 **Keywords:** air pollution; PM<sub>2.5</sub>; lung cancer; survival analysis; epidemiology

## 50 Introduction

51

52 Outdoor air pollution including fine particulate matter (with a diameter  $<2.5\mu\text{m}$  or  $\text{PM}_{2.5}$ ) has  
53 been classified carcinogenic by the International Agency for Research on Cancer (Loomis et  
54 al., 2013). Indeed, a large body of studies found positive associations between exposure to  
55 different air pollutants, including  $\text{PM}_{2.5}$ , and cancer risk of different sites (Pope et al., 2002;  
56 Raaschou-Nielsen et al., 2016, 2013; Turner et al., 2017). Most studies relied on a single  
57 exposure assessment method, which varied across the studies. Further, not all studies obtained  
58 decade-long time-series data that allow to generate long-term exposures. While early studies  
59 used ground-based measurement (such as Dockery et al. (1993)), the computation power and  
60 the available data enabled later studies to use advanced modeling of air pollutant exposure  
61 estimates, such as land use regression, kriging, satellite-based estimates, or chemistry-  
62 dispersion models. These rely on different source data, assumptions and equations, and come  
63 with their own strengths and limitations in terms of: the granularity of the spatial resolution  
64 (from less than 100 meters to more than ten kilometers); spatial coverage (from regional to  
65 global scale); temporal coverage (from daily to annual time-step and covering up to more than  
66 10 years); and the ability to capture spatial or temporal variations at different scales. This  
67 methodological heterogeneity in exposure assessment among studies could partly explain the  
68 heterogeneity of the epidemiologic findings. There have been a few attempts to investigate  
69 whether the choice of exposure assessment method influences the results of risk analyses. Most  
70 studies focused on nitrogen oxides and coarse particulate matter (Cohen et al., 2019; Korek et  
71 al., 2017; Sellier et al., 2014), and only a few included  $\text{PM}_{2.5}$  (Jerrett et al., 2017, 2017; Jin et  
72 al., 2019; Klompaker et al., 2021; Kulhánová et al., 2018; McGuinn et al., 2017), among which  
73 only one focused on risk of cancer (Cheng et al., 2019).

74 To fill this gap, we aim to compare estimated exposures and the resulting risk estimates in a  
75 study of cancer using two distinct exposure assessment methods for  $\text{PM}_{2.5}$ . In a large  
76 population-based French cohort with 26 years of follow-up, exposure was assigned using land

77 use regression and chemistry-dispersion modeling, subsequently used to estimate associations  
78 with two outcomes: all-type cancers and lung cancer.

## 79 **Material and methods**

### 80 **Study population**

81 The Gazel cohort included at inception, in 1989, 20,625 participants, aged 35-50, from the  
82 French national gas and energy company Electricité-de-France (Goldberg et al., 2015).  
83 Participants completed a baseline detailed self-administered questionnaire in 1989, and a  
84 follow-up questionnaire every subsequent year– with a response rate over 80% during the first  
85 three years, and over 70% since. The addresses of participants were collected 1) between 1989  
86 and 2006, by residential calendars completed by the participants in 2006, and 2) from 2006  
87 onward, by registration of each participant’s precise residential address (Figure S1). All  
88 addresses were geocoded and precision quantified: from less precise (at postal code, 26%) to  
89 the most precise (address-level with the full address including street name and number, 36%).  
90 We did not observe any time trend in the percentage of precision of the geocodes.

91 By implementing a 10-year latency between exposure and incidence/censoring (see exposure  
92 assessment and statistical analyses sections), we excluded participants with incident primary  
93 cancers (all-sites or lung) diagnosed before 1999, or who died or were lost on follow-up before  
94 1999. In addition, for the analyses on lung cancer, we compared lung cancer cases to the  
95 participants who did not develop any cancer. Participants who were lost to follow-up or died  
96 during the follow-up after 1999, without a diagnosis of cancer, were censored at date of exit or  
97 at the date of death. We excluded participants with more than 20% of follow up due to missing  
98 addresses, being located outside mainland France, or located on Corsica Island (for which we  
99 could not attribute exposures with one of the models). In our study period (1999-2015), the  
100 study populations included 19,348 and 15,694 participants for the respective analyses on  
101 incident primary all-site and incident primary lung incident cancer (Figure S2).

102 The Gazel study protocol, including a written consent of the cohort participants, was approved  
103 by the French authority for data confidentiality (Commission Nationale Informatique et  
104 Liberté #105728) and by the Ethics Evaluation Committee of Inserm (IRB0000388,  
105 FWA00005831).

106

## 107 **Cancer incidence**

108 We identified any incident cancer in the study period (1999-2015), by means of several sources  
109 described below: the register of the gas and energy company's (Electricité-de-France) medical  
110 department, a validated in-house procedure based on the questionnaires and medical records  
111 for retired participants (Goldberg et al., 2007), and the French national medico-administrative  
112 database of the national health insurance.

113 The company registered all cancers (except non-melanoma skin cancers) occurring during the  
114 employment period of all Gazel participants. Each notified cancer case was documented with  
115 pathology reports and the date of diagnosis, and coded according to the International  
116 Classification of Diseases. The company did not register cancers occurring after retirement. To  
117 address this limitation, from 2008 we obtained dates of diagnosis and types of cancer (ICD-10  
118 and ICD-O-2) as follows: first by recording cancer diagnosis that Gazel participants self-  
119 reported in the annual follow-up questionnaires, and then by contacting all of the Gazel  
120 participants who gave consent for collecting medical information for validation purposes. We  
121 also used data from the national health insurance database that records all data on the use of  
122 the French health system leading to reimbursement (from medical appointments to the  
123 purchase of prescribed drugs and medical examinations). We linked this database to all Gazel  
124 participants who gave consent. This database allowed identifying cancer from data on  
125 hospitalizations with dates and diagnoses, and from the "chronic diseases" register (an official  
126 health insurance's list of 30 serious chronic diseases, including cancer) (Tuppin et al., 2017).

127 In these cancer data sources, we used ICD-10 codes to identify primary incident cases of all-  
128 site cancer (the whole neoplasms chapter except secondary malignant neoplasms C77-79) and

129 lung cancer (C34). Right censoring was applied at incident primary cancers (all-sites or lung,  
130 depending on analysis), date of departure from the cohort or end of follow up, whichever  
131 occurred first.

## 132 **Exposure assessment**

133 We assessed PM<sub>2.5</sub> exposure in the Gazel cohort with two conceptually different exposure  
134 assessment methods: a land use regression model developed for western Europe covering the  
135 1990-2015 period (de Hoogh et al., 2018), and a chemistry-transport model developed for  
136 France covering the 1989-2008 period (Bentayeb et al., 2014) – these time spans are fully  
137 compatible with our study period (1999-2015) since we implement a 10-year lag.

138 Land use regression (LUR): annual mean PM<sub>2.5</sub> concentrations for 2010 were estimated at a  
139 fine spatial resolution (100x100 m) using a supervised stepwise linear regression approach  
140 regressing measured PM<sub>2.5</sub> concentrations (543 sites from the European AIRBASE network)  
141 against a range of predictor variables, including satellite derived and chemical transport  
142 modelled PM<sub>2.5</sub> estimates, road density, land use variables and altitude. Residual variation at  
143 the monitoring sites was explained by ordinary kriging. The model explained 66% of the spatial  
144 variation in measured PM<sub>2.5</sub> concentrations (de Hoogh et al., 2018). The estimated PM<sub>2.5</sub>  
145 concentrations for 2010 were rescaled for years 1990-2015, by European NUTS-1 region (eight  
146 NUTS-1 regions in mainland France), using back- and forward- extrapolation with the ratio  
147 method. Annual mean PM<sub>2.5</sub> estimates (1990 – 2015) from the 26x26 km Danish Eulerian  
148 Hemispheric Model (DEHM) (Brandt et al., 2012) were used for this purpose. In addition, in  
149 this study, we further back-extrapolated PM<sub>2.5</sub> exposure to 1989.

150 Chemistry-dispersion model (Gazel-Air): the chemistry-dispersion model, already used in the  
151 previous Gazel-Air study (Bentayeb et al., 2014), assessed PM<sub>2.5</sub> concentrations each year  
152 between 1989 and 2008 following these steps: (i) the Chimere chemistry-transport model was  
153 used to compute annual mean concentrations over Europe (on a 30x30 km grid) and France  
154 (on a 10x10 km grid). This model was fed by the European Monitoring and Evaluation Program  
155 emissions inventories, point sources, road emissions, and biogenic emissions. (ii) Mesh

156 refinement and data assimilation with geostatistical analyses were then applied to estimate  
157 concentrations on a 2x2 km grid.

158 For each participant and each available year, we estimated PM<sub>2.5</sub> exposure as the value at the  
159 residential address for both exposure assessment methods.

## 160 **Covariables**

161 Based on previous literature, we selected *a priori* the following variables as confounders or  
162 potential effect modifiers:

163 Sociodemographic and occupational variables: sex, education (9-11 years, 12-13 years, 14-15  
164 years, other secondary education, other diploma), and socio-economic status (low: blue-collar  
165 workers or clerks; medium: first-line supervisors or sales representatives; high: management),  
166 all at baseline. We also used Gazel participants' occupational histories linked to a French job  
167 exposure matrix (Imbernon, 1991) to include a synthetic summary of occupational exposure to  
168 nine known lung carcinogens (asbestos, cadmium, chlorinated solvents, chromium, coal  
169 gasification, coal-tar pitch, creosotes, crystalline silica, and hydrazine) over the whole  
170 employment period in the company and categorized into none, one, two, or at least three  
171 carcinogens .

172 Lifestyle variables: time-varying variables for tobacco (cumulative smoking pack-years and  
173 smoking status), alcohol intake (abstinent, non-daily drinker, light drinker, moderate drinker,  
174 heavy drinker, unclear pattern), family status (single or not), body mass index (BMI in kg/m<sup>2</sup>).  
175 Two variables were included only at specific time points: passive smoking at home or at work  
176 (yes or no) in 1990 and 1996, and fruit-vegetable intake (never or less than once a week; once  
177 or twice a week; more than twice a week but not every day; every day or almost) in 1998, 2004,  
178 2009 and 2014. We computed these variables as time-varying, attributing the data collected in  
179 1990 and 1996 to each year of the periods 1989-1995 and 1996-2015 for passive smoking, and  
180 the data collected in 1998, 2004, 2009 and 2014 to each year of the periods 1989-1998, 1999-  
181 2004, 2005-2010, 2011-2015 for fruit and vegetable consumption.



182 **Contextual variables:** For all participants, we calculated the distance to the nearest major road  
183 and assigned European NUTS-1 region (Nomenclature of Territorial Units for Statistics, which  
184 are areas defined by the European Union) every year. We built an “urban classification”  
185 variable based on the population density obtained from the French National Institute for  
186 Statistics, for every municipality, which was weighted according to the inhabited surface of  
187 each municipality and classified as high (urban), moderate (semi-urban), and low (rural). We  
188 further specified whether urban participants were living in the Paris region or not. We obtained  
189 such values for the years 1990, 2000, and 2010 and assessed a “long term urban classification”  
190 as solely urban in the Paris region, solely urban out of the Paris region, solely semi-urban,  
191 solely rural, or mixed during these three years. We also obtained the French deprivation index  
192 (Rey et al., 2009) calculated in 2009 for all participants having survived until then (categorized  
193 into three tertiles as low, middle, and high deprivation – with a missing value for participants  
194 who died before 2009, considered as a category so not to lose 650 participants when using this  
195 variable in the analyses).

## 196 **Imputations**

197 Our dataset included some missing values (from 21% for passive smoking to 29% for BMI). We  
198 imputed missing values longitudinally, considering we had several years of follow-up by  
199 participant. To do so, we chose a 2-level imputation model with participants as clusters using  
200 the packages MICE (van Buuren and Groothuis-Oudshoorn, 2011) and MICEADDS (Robitzsch  
201 et al., 2018). We imputed all variables except exposures to PM<sub>2.5</sub>, and contextual variables (but  
202 we still used them to impute all the other variables) using the functions “2l.pmm” and  
203 “2l.only.pmm” for time-varying and time-independent variables, respectively. We obtained a  
204 satisfactory convergence after 10 iterations for the 10 datasets (Figure S3).

205

## 206 **Statistics**

207 Comparing estimates of exposure to PM<sub>2.5</sub>:

208 Estimates from the two exposure assessment methods, applied to each participant for each  
209 year, were compared as follows. First, we calculated Spearman's correlation coefficients a) for  
210 the complete dataset and b) on subsets according to the year and region to check for any  
211 differences. We plotted Bland-Altman graphs to check for any bias between the two methods,  
212 again using the complete dataset and subsets. Since the two exposure assessment methods  
213 considered the influence of traffic and roads differently, we also explored correlations in  
214 population subsets defined by distance to the nearest major road. We defined subsets as (i) the  
215 European NUTS-1 regions, and (ii) we divided the distance to the major road into quartiles,  
216 and further subdivided the first quartile at 150m to explore near-road exposures.

217 To visualize the differences between the two exposure models over France in 1990 and 2008,  
218 we mapped the differences in exposures (LUR minus Gaze-Air) for participants available at  
219 the two time points. To improve the maps' readability, we averaged the results on a 5x5km  
220 grid.

221 Associations between PM<sub>2.5</sub> and cancer, using exposure derived from the two models:

222 Since cancers are believed to develop over a lengthy period of time, it is customary to discount  
223 exposures that occurred recently as these are unlikely to have affected the disease risk  
224 (Rothman et al., 2008). In the main analysis, we thus implemented a 10-year lag period  
225 (Supplementary Methods). We thus considered only cases diagnosed between 1999 and 2015,  
226 with corresponding exposures 10 years before incidence/censoring.

227 We used Cox regressions to analyze the associations between PM<sub>2.5</sub> exposure and incident  
228 primary cancer (all-site and lung cancer cases). Specifically, we built extended Cox models with  
229 time-varying variables to satisfy the proportional hazards assumption (Supplementary  
230 material). We used attained age as underlying time scale and a time-varying cumulative  
231 exposure to PM<sub>2.5</sub> as previously used (Lequy et al., 2021), adjusted for age at enrollment

232 (Pencina et al., 2007); we also adjusted for calendar time dichotomized with a cut-off in 2007  
233 (Supplementary Material). We used a spline function with three degrees of freedom to test for  
234 non-linearity. Due to non-linearity (Figure S4), we modelled the relationship with a two-piece  
235 linear model by including an interaction term between PM<sub>2.5</sub> exposure and a Boolean variable  
236 at the specified percentile (Lequy et al., 2021), and provide resulting hazard ratios (HR) and  
237 95% confidence intervals accordingly (Supplementary Material). As an alternative analysis to  
238 address non-linearity and check for the robustness of the results, we natural log-transformed  
239 the cumulative annual PM<sub>2.5</sub> time-dependent exposure (Lequy et al., 2021).

240 For both the two-piece linear and the log-transformation approach, we adjusted for the  
241 following a priori variables: sex, tobacco smoking (cumulative pack-years), passive smoking,  
242 alcohol consumption, socioeconomic status, education, marital status, BMI, occupational  
243 exposure to lung carcinogens, and fruit and vegetable consumption, each with a 10-year lag  
244 (except passive smoking and fruit-vegetable consumption, because of the way they were  
245 imputed, see “Imputations” section). We assumed a linear relationship between continuous  
246 variables and each outcome. Sex was included as a strata variable as it did not meet the  
247 proportional hazards assumption.

248 For all-site cancer incidence, we conducted sensitivity analyses by (i) further adjusting for the  
249 area-level French deprivation index in addition to individual SES (Temam et al., 2017), (ii)  
250 restricting the study population to the participants with only the most precise geocodes  
251 throughout their follow-up (i.e. addresses-level geocodes, see “Study population”), and  
252 implementing several alternative ways to handle missing data by (iii) considering missing data  
253 as a category, (iv) imputing missing data as median/mode, or (v) conducting complete-case  
254 analyses. To detect any effect modification by sex, smoking status, distance to the nearest  
255 major road, and population density, we restricted our analyses to the following groups: female  
256 or male, ever or never smokers, nearest major road within or beyond the median distance of  
257 500m over the follow-up, and following the urban classification at municipality-level (see

258 covariables): solely urban (separately in the Paris region vs out of the Paris region) or solely  
259 semi-urban or solely rural.

260 For lung cancer incidence, due to the small number of cases, we did not perform any sensitivity  
261 or stratified analyses.

262 We conducted all analyses with the R software version 3.5.1 (R Core Team, 2018) and the  
263 'survival' package (Therneau, 2015; Therneau and Grambsch, 2000). Unless specified  
264 otherwise, we used multiple imputations by chained equations to conduct all the analyses and  
265 pooled estimates using Rubin's rules.

## 266 **Results**

### 267 **Characteristics of Gazel's participants**

268 Of the eligible 19,348 participants, we identified 3,711 incident primary cancers (all-sites) and  
269 349 incident primary lung cancer cases diagnosed between 1999 and 2015. Our study sample  
270 included more than 70% males and the median age of participants was 43.5 years at enrollment  
271 (Table 1). At enrollment, most had never smoked or had quit smoking, and the cumulative  
272 pack-years for ever-smokers was 15. More than 40% were light alcohol drinkers. Around 60%  
273 were not exposed to any of the nine occupational lung carcinogens. The high number of study  
274 participants leads to statistically significant differences between participants diagnosed with  
275 cancer (all-sites) and participants without cancer during the observation period. However,  
276 these differences were rather small, except for sex and pack-years of tobacco smoking: cancer  
277 cases were more likely to be men and reported more cumulative pack-years.

### 278 **Comparing the LUR and Gazel-Air exposure models**

279 The common period in which the two models provided exposures spanned from 1989 to 2008.  
280 The median exposure to PM<sub>2.5</sub> in 1989 was 33.3 µg.m<sup>-3</sup> [interquartile range: 30.1-37.8] and  
281 17.0 µg.m<sup>-3</sup> [14.5-20.5] for the LUR and the Gazel-Air model, respectively. We observed a

282 decreasing trend over time for both models, although more pronounced for the LUR model  
283 (Figure 1).

284 *Temporal comparison:* the raw annual estimates from the two exposure assessment methods  
285 correlated highest in 2008 (Spearman's correlation coefficient of 0.68) the year closest to 2010  
286 for which the LUR was originally developed and for which we have Gazel-Air estimates.  
287 Correlations ranged from 0.62 to 0.79 across years (0.71 combining all available years).  
288 Exposure estimates diverged when going back in time, with almost systematically higher LUR  
289 estimates. Further, the Bland-Altman graphs indicated an overestimation of the LUR model  
290 compared to the Gazel-Air model already noticeable in 2008 and increasing when going back  
291 in time (Figure S5).

292 *Spatial comparison:* when stratifying by European NUTS-1 region, we observed correlations  
293 ranging from 0.59 to 0.75 (Table 2). Furthermore, within each European NUTS-1 region and  
294 for the specific years 1990, 2000, and 2008, we observed quite stable correlations (although  
295 slightly weaker) in these regions over time (Table 2).

296 *Spatio-temporal comparison:* the maps revealed differences between the two models that  
297 evolved differentially according to the region (Figure 2). For example, the LUR model  
298 estimated both higher and lower levels of PM<sub>2.5</sub> in 2008, but estimated almost only higher  
299 PM<sub>2.5</sub> concentrations across France in 1990. Therefore, the spatial differences did not increase  
300 similarly all over France when going back in time.

301 We did not detect any substantial effect of the distance to the nearest major road on the  
302 correlations, with a magnitude of 0.7 for all years combined (Table 2). We also checked  
303 specifically in 1990, 2000, and 2008 and observed no substantial effect of the distance to the  
304 nearest major road.

## 305 **Associations between PM<sub>2.5</sub> exposure and cancer using two exposure** 306 **assessment methods**

307 All-sites combined cancer incidence:

308 Using the two-piece linear approach with one interquartile range (IQR – 216  $\mu\text{g}/\text{m}^3$  for the  
309 LUR, 127  $\mu\text{g}/\text{m}^3$  for Gazel-Air) as the unit of exposure, we estimated associations for each  
310 exposure assessment method separately.  $\text{PM}_{2.5}$  response approximately followed nonlinear  
311 curve with a positive slope until roughly the 65<sup>th</sup> percentile of the exposure distribution before  
312 a plateau (Figure S4). The exposure estimates derived from the LUR model led to an HR of  
313 1.53 (95% Confidence interval: 1.24-1.88) and those from the Gazel-Air model led to an HR of  
314 1.43 (1.19-1.73) below the 65<sup>th</sup> percentile of exposure (Table 3), then a plateau at HR around  
315 1.5 for both exposure assessments (Figure 3, Table 3). Using the natural log-transformed  
316 approach, we estimated HR of 1.21 (1.07-1.37) per 1-unit increase of natural log-transformed  
317 cumulative  $\text{PM}_{2.5}$  for the LUR approach and 1.19 (1.07-1.33) for the Gazel-Air approach (Table  
318 S1).

319 The sensitivity analyses yielded mostly positive and similar point-estimates using the two  
320 exposure assessment methods (Table 4, Table S1, Figure S6). Across population subsets, we  
321 estimated similar associations between incident cancer and exposure to  $\text{PM}_{2.5}$ , whether using  
322 the LUR or the Gazel-Air exposures (Table S1, Table S2, Figure S6).

323 The estimated associations between important risk factors such as smoking and alcohol intake  
324 and all-site combined cancer incidence showed expected directions and magnitude (Table S3).

#### 325 Lung cancer incidence:

326  $\text{PM}_{2.5}$  response approximately followed nonlinear curve with a positive slope until roughly the  
327 73<sup>rd</sup> percentile of the exposure distribution before a plateau (Figure S4). Our main analysis  
328 provided discrepant HRs between the two exposure models, but statistically non-significant  
329 and with largely overlapping confidence intervals (Figure S4, Table 3). The exposure estimates  
330 derived from the LUR model led to an HR of 1.15 (0.58-2.29) and those from the Gazel-Air  
331 model led to an HR of 1.79 (0.94-3.40) below the 73<sup>th</sup> percentile of exposure, then a plateau at  
332 around 1.1 and 2.0 for the LUR and the Gazel-Air exposure assessment respectively. Using the  
333 log-transformation approach, we estimated, per 1-unit increase of log-transformed cumulative

334  $PM_{2.5}$ , an HR of 1.00 (0.65-1.56) for the LUR estimates and 1.14 (0.77-1.69) for the Gazel-Air  
335 estimates per 1-unit increase of natural log-transformed cumulative  $PM_{2.5}$ .

## 336 Discussion

337 This study aimed at comparing the associations between long-term exposure to  $PM_{2.5}$  and  
338 incident cancer risk using two conceptually different exposure assessment methods. Our  
339 findings confirmed the known association between long-term exposure to  $PM_{2.5}$  and all-site  
340 cancer risk. Furthermore, we found very similar associations between long-term cumulative  
341  $PM_{2.5}$  exposure and incident cancer using a land-use regression and a chemistry-dispersion  
342 model. For lung cancer, our results were less consistent and statistically non-significant.

343 Regarding the exposure assessment, we observed higher exposure estimates for the LUR in  
344 2008 and these differences between the LUR and Gazel-Air models increased back in time  
345 from 2008 to 1989. These differences did not depend on a major  $PM_{2.5}$  determinant such as  
346 the distance to the nearest major road and could be at least partly explained by the differences  
347 between Gazel-Air and the DEHM model used to extrapolate the LUR estimates between 1990-  
348 2015, such as the underlying equations, the covered area (France or Europe), and the emission  
349 data. Yet the exposure estimates from both methods were highly correlated.

350 When comparing the associations obtained using separately these two exposure assessment  
351 methods, we noticed generally similar associations in the case of all-site cancer for which the  
352 analyses had a great statistical power. For all-site cancer, the only striking discrepancy in our  
353 findings pertains to the population subset “urban out of Paris region” for which using the LUR  
354 model provided a positive but statistically non-significant HR, while using the Gazel-Air model  
355 provided a null association. While this difference could be due to chance, this could also be due  
356 to the exposure assessment methodology: as already mentioned, Gazel-Air had a resolution of  
357 2x2 km (larger than the LUR) and did not model  $PM_{2.5}$  in urban areas as thoroughly in other  
358 cities as in Paris, which may lead to misclassification for participants living in such urban areas.  
359 On the contrary, the LUR estimates were derived from a single European model: the advantage

360 is that these estimates are harmonized all over France, even though these estimates may be  
361 less specific of French PM<sub>2.5</sub> measurements than the Gazel-Air estimates. The associations on  
362 lung cancer were very different but the wide confidence intervals largely overlapped. Still,  
363 despite statistically non-significant, the point estimates were positive, in agreement with the  
364 current literature. With only 349 lung cancer cases, these analyses clearly lacked statistical  
365 power and interpreting these findings requires great caution. There are many methods of PM<sub>2.5</sub>  
366 exposure assessment, and for large study areas and populations models based on  
367 measurements are needed. There is, however, no gold-standard; in this study, for outcomes  
368 with sufficient statistical power, the two validated methods led to similar risk estimates. More  
369 research is warranted on the influence of different exposure assessment methods on risk  
370 estimates, in particular using long time-series. Such time-series should also try including  
371 trends in PM<sub>2.5</sub> composition to further refine the association between PM<sub>2.5</sub> exposure and risk  
372 of cancer, and evaluate the effects of composition changes.

373 This study relies on very detailed epidemiological and environmental characterization, from a  
374 large population-based cohort with a 26-year follow-up. We could assess exposure for each  
375 available address and year and take into account all moves over the follow-up period. This  
376 ensured the most accurate exposure assignment considering our data (home addresses and  
377 annual concentrations of PM<sub>2.5</sub>), a quite rare feature in the literature especially when studying  
378 the influence of the exposure assessment method on risk estimates. Both exposure models  
379 provided spatial exposure contrasts. We used a time-dependent Cox model which controls for  
380 as many time-dependent covariables as possible, and took into account the possible bias linked  
381 to an aging cohort and the corresponding increasing cancer risk. These models estimated  
382 expected associations with all covariables (e.g. positive associations with smoking, negative  
383 with increasing fruit-vegetable intake). Like all similar studies so far, we have only obtained  
384 residential addresses, so we could not consider commuting or pollution at the workplace or in  
385 leisure activities. We also had no data on indoor air pollution. We obtained exposure data at  
386 an annual time-step, which should minimize misclassification since we focused on long-term  
387 exposure and took into account participants' moves. The MATEX job-exposure matrix did not



388 allow for distinguishing exposure to specific compounds such as polycyclic aromatic  
389 hydrocarbon; yet adding these exposures would probably not affect the point estimates of the  
390 association between PM<sub>2.5</sub> and cancer incidence, as the estimated associations between cancer  
391 incidence and occupational exposure to carcinogens in this study were not statistically  
392 significant (Table S3). Finally, we had to proceed to multiple imputations due to the missing  
393 data in the follow-up – since we adjusted for several covariables, the percentage of incomplete  
394 data was rather high in the final Cox model. But we used 10 sets of imputed data with good  
395 convergence across these 10 datasets, therefore we are confident in our results using the  
396 imputed data.

397 To conduct our analyses, and to meet the proportional hazard assumption, we had to specify a  
398 time-dependent Cox model using cumulative exposure as main variable. This unusual method  
399 allowed considering the fact PM<sub>2.5</sub> exposure levels decreased over a 26-year period during  
400 which the study population aged and was more at risk to develop a cancer; we can expect that  
401 studies on the long-term health effects of air pollution exposure are going to face this issue  
402 more and more often, at least in countries in which air pollution has decreased. Besides,  
403 current levels of developed countries decreased to low levels, whose effects might be difficult  
404 to detect. To consider these two issues, it is necessary to develop suitable methods. In our  
405 study, both of the evaluated exposure variables (i) led to violating the proportional hazard  
406 assumption in a non-time-dependent Cox model, (ii) followed a similar non-linear relationship  
407 with cancer risk, and (iii) provided similar associations with all-site cancer risk. For all these  
408 advantages, this methodology yielded associations that can be difficult to directly compare with  
409 the literature.

410 Many different exposure assessment methods have been used in air pollution epidemiology.  
411 Only a few studies have used multiple methods to derive estimates of PM<sub>2.5</sub> exposure with a  
412 view to comparing the results generated by the different methods. Jerrett et al. (2017) and Jin  
413 et al. (2019) studied mortality risk and burden, respectively, and used PM<sub>2.5</sub> exposures from  
414 different modeling techniques (based on remote sensing and ground-level data, at different

415 resolutions from 0.01° to 12x12km). Both studies found high correlations between the exposure  
416 assessment methods, ranging from 0.54 to 0.99 and from 0.65 to 0.90, and associations with  
417 a similar magnitude. Klompmaker et al. (2021) also studied mortality risk using three types of  
418 modelled PM<sub>2.5</sub> exposure (at baseline) – including the same LUR as that used in this study –  
419 with correlations between 0.61-0.72, and generally positive associations with mortality  
420 especially for respiratory outcomes (respiratory diseases and lung cancer mortality), with  
421 weaker associations using the LUR exposures. Jin and colleagues (2019) had observed a larger  
422 uncertainty due to the epidemiological data larger than the one due to the exposure assessment  
423 method. McGuinn et al. (2017) studied cardiovascular pathologies and exposures from five  
424 approaches providing correlated variables (coefficient between 0.60 and 0.88). They also  
425 found similar odds. Cheng et al. (2019) studied breast cancer incidence and used exposures to  
426 PM<sub>2.5</sub> from ordinary kriging and dispersion models: they did not find similar associations with  
427 breast cancer risk but pointed out different spatial and temporal resolutions for the two  
428 exposure models. In particular, their kriging model relied on regional rather than local sources,  
429 while the dispersion model focused on local sources. This was not the case for our two exposure  
430 models. In general, our findings are consistent with those of previous studies, both in terms of  
431 the exposure correlation coefficient (0.71 in our study) and in the sense that we found similar  
432 magnitude of the associations from the two exposure models. This indicates a reassuring  
433 robustness of the findings reported so far, especially for main analyses, but stresses the need  
434 for giving due consideration to the selection of exposure models more often.

## 435 **Conclusions**

436 This study compared two distinct and often used exposure assessment methods in  
437 epidemiological analyses on air pollution exposure and cancer risk. We found similar  
438 associations between PM<sub>2.5</sub> exposure and all-site cancer using both exposure assessment  
439 methods. Although the associations with lung cancer were less similar, results from both  
440 exposure assessment methods were positive. While both methods are considered fit for  
441 purpose, without a gold-standard exposure assessment method, we cannot say which point

442 estimate is more accurate; however, our findings suggest that the epidemiological findings  
443 related to PM<sub>2.5</sub> and the investigated cancers are robust.

#### 444 **Acknowledgments**

445 The Gazel Cohort Study was funded by EDF-GDF, and received grants from CAMIEG (Caisse  
446 d'assurance maladie des industries électrique et gazière) and the "Comité de coordination des  
447 CMCAS". This study was supported by the ARC French Foundation for cancer research  
448 (OCAPOL, CANCAIR-201501234). The authors want to thank Sébastien Bonenfant who  
449 reconstituted the residential histories of Gazel participants.

#### 450 **References**

- 451 Bentayeb, M., Stempfelet, M., Wagner, V., Zins, M., Bonenfant, S., Songeur, C., Sanchez, O.,  
452 Rosso, A., Brulfert, G., Rios, I., Chaxel, E., Virga, J., Armengaud, A., Rossello, P.,  
453 Riviere, E., Bernard, M., Vasbien, F., Deprost, R., 2014. Retrospective modeling  
454 outdoor air pollution at a fine spatial scale in France, 1989-2008. *Atmos. Environ.* 92,  
455 267–279. <https://doi.org/10.1016/j.atmosenv.2014.04.019>
- 456 Brandt, J., Silver, J.D., Frohn, L.M., Geels, C., Gross, A., Hansen, A.B., Hansen, K.M.,  
457 Hedegaard, G.B., Skjøth, C.A., Villadsen, H., Zare, A., Christensen, J.H., 2012. An  
458 integrated model study for Europe and North America using the Danish Eulerian  
459 Hemispheric Model with focus on intercontinental transport of air pollution.  
460 *Atmospheric Environment, AQMEII: An International Initiative for the Evaluation of*  
461 *Regional-Scale Air Quality Models - Phase 1* 53, 156–176.  
462 <https://doi.org/10.1016/j.atmosenv.2012.01.011>
- 463 Cheng, I., Tseng, C., Wu, J., Yang, J., Conroy, S.M., Shariff-Marco, S., Li, L., Hertz, A.,  
464 Gomez, S.L., Le Marchand, L., Whittemore, A.S., Stram, D.O., Ritz, B., Wu, A.H.,  
465 2019. Association between ambient air pollution and breast cancer risk: The  
466 multiethnic cohort study. *International journal of cancer.*  
467 <https://doi.org/10.1002/ijc.32308>

468 Cohen, G., Steinberg, D.M., Yuval, Levy, I., Chen, S., Kark, J.D., Levin, N., Witberg, G.,  
469 Bental, T., Broday, D.M., Kornowski, R., Gerber, Y., 2019. Cancer and mortality in  
470 relation to traffic-related air pollution among coronary patients: Using an ensemble of  
471 exposure estimates to identify high risk individuals. *Environ. Res.* 176, 108560.  
472 <https://doi.org/10.1016/j.envres.2019.108560>

473 de Hoogh, K., Chen, J., Gulliver, J., Hoffmann, B., Hertel, O., Ketznel, M., Bauwelinck, M., van  
474 Donkelaar, A., Hvidtfeldt, U.A., Katsouyanni, K., Klompmaker, J., Martin, R.V.,  
475 Samoli, E., Schwartz, P.E., Stafoggia, M., Bellander, T., Strak, M., Wolf, K., Vienneau,  
476 D., Brunekreef, B., Hoek, G., 2018. Spatial PM<sub>2.5</sub>, NO<sub>2</sub>, O<sub>3</sub> and BC models for  
477 Western Europe - Evaluation of spatiotemporal stability. *Environ Int* 120, 81–92.  
478 <https://doi.org/10.1016/j.envint.2018.07.036>

479 Dockery, D.W., Pope, C.A., Xu, X., Spengler, J.D., Ware, J.H., Fay, M.E., Ferris, B.G.Jr.,  
480 Speizer, F.E., 1993. An Association between Air Pollution and Mortality in Six U.S.  
481 Cities. *New England Journal of Medicine* 329, 1753–1759.  
482 <https://doi.org/10.1056/NEJM199312093292401>

483 Goldberg, M., Leclerc, A., Bonenfant, S., Chastang, J.F., Schmaus, A., Kaniewski, N., Zins, M.,  
484 2007. Cohort profile: the GAZEL Cohort Study. *Int J Epidemiol* 36, 32–39.  
485 <https://doi.org/10.1093/ije/dyl247>

486 Goldberg, M., Leclerc, A., Zins, M., 2015. Cohort Profile Update: The GAZEL Cohort Study.  
487 *Int J Epidemiol* 44, 77–77g. <https://doi.org/10.1093/ije/dyu224>

488 Imbernon, E., 1991. Matex : une matrice emplois-expositions destinée à la surveillance  
489 épidémiologique des travailleurs d'une grande entreprise (E.D.F.-G.D.F.). *Archives*  
490 *des maladies professionnelles et de médecine du travail* 52, 559–566.

491 Jerrett, M., Turner, M.C., Beckerman, B.S., Pope III, C.A., van Donkelaar, A., Martin, R.V.,  
492 Serre, M., Crouse, D., Gapstur, S.M., Krewski, D., Diver, W.R., Coogan, P.F.,  
493 Thurston, G.D., Burnett, R.T., 2017. Comparing the Health Effects of Ambient  
494 Particulate Matter Estimated Using Ground-Based versus Remote Sensing Exposure

495 Estimates. *Environmental Health Perspectives* 125, 552–559.  
496 <https://doi.org/10.1289/EHP575>

497 Jin, X., Fiore, A.M., Civerolo, K., Bi, J., Liu, Y., van Donkelaar, A., Martin, R.V., Al-Hamdan,  
498 M., Zhang, Y., Insaif, T.Z., Kioumourtzoglou, M.-A., He, M.Z., Kinney, P.L., 2019.  
499 Comparison of multiple PM<sub>2.5</sub> exposure products for estimating health benefits of  
500 emission controls over New York State, USA. *Environ. Res. Lett.* 14, 084023.  
501 <https://doi.org/10.1088/1748-9326/ab2dcb>

502 Klompaker, J.O., Janssen, N., Andersen, Z.J., Atkinson, R., Bauwelinck, M., Chen, J., de  
503 Hoogh, K., Houthuijs, D., Katsouyanni, K., Marra, M., Oftedal, B., Rodopoulou, S.,  
504 Samoli, E., Stafoggia, M., Strak, M., Swart, W., Wesseling, J., Vienneau, D.,  
505 Brunekreef, B., Hoek, G., 2021. Comparison of associations between mortality and air  
506 pollution exposure estimated with a hybrid, a land-use regression and a dispersion  
507 model. *Environment International* 146, 106306.  
508 <https://doi.org/10.1016/j.envint.2020.106306>

509 Korek, M., Johansson, C., Svensson, N., Lind, T., Beelen, R., Hoek, G., Pershagen, G.,  
510 Bellander, T., 2017. Can dispersion modeling of air pollution be improved by land-use  
511 regression? An example from Stockholm, Sweden. *J Expo Sci Environ Epidemiol* 27,  
512 575–581. <https://doi.org/10.1038/jes.2016.40>

513 Kulhánová, I., Morelli, X., Le Tertre, A., Loomis, D., Charbotel, B., Medina, S., Ormsby, J.-N.,  
514 Lepeule, J., Slama, R., Soerjomataram, I., 2018. The fraction of lung cancer incidence  
515 attributable to fine particulate air pollution in France: Impact of spatial resolution of  
516 air pollution models. *Environ Int.* <https://doi.org/10.1016/j.envint.2018.09.055>

517 Lequy, E., Siemiatycki, J., de Hoogh, K., Vienneau, D., Dupuy, J.-F., Garès, V., Hertel, O.,  
518 Christensen, J.H., Zhivin, S., Goldberg, M., Zins, M., Jacquemin, B., 2021.  
519 Contribution of Long-Term Exposure to Outdoor Black Carbon to the Carcinogenicity  
520 of Air Pollution: Evidence regarding Risk of Cancer in the Gazel Cohort. *Environ*  
521 *Health Perspect* 129, 37005. <https://doi.org/10.1289/EHP8719>

522 Loomis, D., Grosse, Y., Lauby-Secretan, B., Ghissassi, F.E., Bouvard, V., Benbrahim-Tallaa,  
523 L., Guha, N., Baan, R., Mattock, H., Straif, K., 2013. The carcinogenicity of outdoor air  
524 pollution. *The Lancet Oncology* 14, 1262–1263. [https://doi.org/10.1016/S1470-](https://doi.org/10.1016/S1470-2045(13)70487-X)  
525 [2045\(13\)70487-X](https://doi.org/10.1016/S1470-2045(13)70487-X)

526 McGuinn, L.A., Ward-Caviness, C., Neas, L.M., Schneider, A., Di, Q., Chudnovsky, A.,  
527 Schwartz, J., Koutrakis, P., Russell, A.G., Garcia, V., Kraus, W.E., Hauser, E.R.,  
528 Cascio, W., Diaz-Sanchez, D., Devlin, R.B., 2017. Fine particulate matter and  
529 cardiovascular disease: Comparison of assessment methods for long-term exposure.  
530 *Environ. Res.* 159, 16–23. <https://doi.org/10.1016/j.envres.2017.07.041>

531 Pencina, M.J., Larson, M.G., D’Agostino, R.B., 2007. Choice of time scale and its effect on  
532 significance of predictors in longitudinal studies. *Stat Med* 26, 1343–1359.  
533 <https://doi.org/10.1002/sim.2699>

534 Pope, C.A., Burnett, R.T., Thun, M.J., Calle, E.E., Krewski, D., Ito, K., Thurston, G.D., 2002.  
535 Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate  
536 air pollution. *Jama* 287, 1132–41.

537 R Core Team, 2018. *R: A Language and Environment for Statistical Computing*. R  
538 Foundation for Statistical Computing, Vienna, Austria.

539 Raaschou-Nielsen, O., Andersen, Z.J., Beelen, R., Samoli, E., Stafoggia, M., Weinmayr, G.,  
540 Hoffmann, B., Fischer, P., Nieuwenhuijsen, M.J., Brunekreef, B., Xun, W.W.,  
541 Katsouyanni, K., Dimakopoulou, K., Sommar, J., Forsberg, B., Modig, L., Oudin, A.,  
542 Oftedal, B., Schwarze, P.E., Nafstad, P., De Faire, U., Pedersen, N.L., Ostenson, C.-G.,  
543 Fratiglioni, L., Penell, J., Korek, M., Pershagen, G., Eriksen, K.T., Sørensen, M.,  
544 Tjønneland, A., Ellermann, T., Eeftens, M., Peeters, P.H., Meliefste, K., Wang, M.,  
545 Bueno-de-Mesquita, B., Key, T.J., de Hoogh, K., Concin, H., Nagel, G., Vilier, A.,  
546 Grioni, S., Krogh, V., Tsai, M.-Y., Ricceri, F., Sacerdote, C., Galassi, C., Migliore, E.,  
547 Ranzi, A., Cesaroni, G., Badaloni, C., Forastiere, F., Tamayo, I., Amiano, P.,  
548 Dorronsoro, M., Trichopoulou, A., Bamia, C., Vineis, P., Hoek, G., 2013. Air pollution  
549 and lung cancer incidence in 17 European cohorts: prospective analyses from the

550 European Study of Cohorts for Air Pollution Effects (ESCAPE). *Lancet Oncol.* 14,  
551 813–822. [https://doi.org/10.1016/S1470-2045\(13\)70279-1](https://doi.org/10.1016/S1470-2045(13)70279-1)

552 Raaschou-Nielsen, O., Pedersen, M., Stafoggia, M., Weinmayr, G., Andersen, Z.J., Galassi, C.,  
553 Sommar, J., Forsberg, B., Olsson, D., Oftedal, B., Krog, N.H., Aasvang, G.M., Pyko, A.,  
554 Pershagen, G., Korek, M., De Faire, U., Pedersen, N.L., Östenson, C.-G., Fratiglioni,  
555 L., Sørensen, M., Eriksen, K.T., Tjønneland, A., Peeters, P.H., Bueno-de-Mesquita, B.,  
556 Plusquin, M., Key, T.J., Jaensch, A., Nagel, G., Föger, B., Wang, M., Tsai, M.-Y.,  
557 Grioni, S., Marcon, A., Krogh, V., Ricceri, F., Sacerdote, C., Migliore, E., Tamayo, I.,  
558 Amiano, P., Dorronsoro, M., Sokhi, R., Kooter, I., de Hoogh, K., Beelen, R., Eeftens,  
559 M., Vermeulen, R., Vineis, P., Brunekreef, B., Hoek, G., 2016. Outdoor air pollution  
560 and risk for kidney parenchyma cancer in 14 European cohorts. *Int. J. Cancer.*  
561 <https://doi.org/10.1002/ijc.30587>

562 Rey, G., Jougl, E., Fouillet, A., Hémon, D., 2009. Ecological association between a  
563 deprivation index and mortality in France over the period 1997 - 2001: variations with  
564 spatial scale, degree of urbanicity, age, gender and cause of death. *BMC Public Health*  
565 9, 33. <https://doi.org/10.1186/1471-2458-9-33>

566 Robitzsch, A., Grund, S., Henke, T., 2018. miceadds: Some additional multiple imputation  
567 functions, especially for mice.

568 Rothman, K.J., Greenland, S., Lash, T.L., 2008. *Modern Epidemiology*. Lippincott Williams  
569 & Wilkins.

570 Sellier, Y., Galineau, J., Hulin, A., Caini, F., Marquis, N., Navel, V., Bottagisi, S., Giorgis-  
571 Allemand, L., Jacquier, C., Slama, R., Lepeule, J., 2014. Health effects of ambient air  
572 pollution: Do different methods for estimating exposure lead to different results?  
573 *Environment International* 66, 165–173.  
574 <https://doi.org/10.1016/j.envint.2014.02.001>

575 Temam, S., Burte, E., Adam, M., Antó, J.M., Basagaña, X., Bousquet, J., Carsin, A.-E.,  
576 Galobardes, B., Keidel, D., Künzli, N., Le Moual, N., Sanchez, M., Sunyer, J., Bono, R.,  
577 Brunekreef, B., Heinrich, J., de Hoogh, K., Jarvis, D., Marcon, A., Modig, L., Nadif, R.,

578 Nieuwenhuijsen, M., Pin, I., Siroux, V., Stempfelet, M., Tsai, M.-Y., Probst-Hensch,  
579 N., Jacquemin, B., 2017. Socioeconomic position and outdoor nitrogen dioxide (NO<sub>2</sub>)  
580 exposure in Western Europe: A multi-city analysis. *Environ Int* 101, 117–124.  
581 <https://doi.org/10.1016/j.envint.2016.12.026>

582 Therneau, T.M., 2015. A Package for Survival Analysis in S.

583 Therneau, T.M., Grambsch, P.M., 2000. Modeling Survival Data: Extending the Cox Model.  
584 Springer, New York.

585 Tuppin, P., Rudant, J., Constantinou, P., Gastaldi-Ménager, C., Rachas, A., de Roquefeuil, L.,  
586 Maura, G., Caillol, H., Tajahmady, A., Coste, J., Gissot, C., Weill, A., Fagot-Campagna,  
587 A., 2017. Value of a national administrative database to guide public decisions: From  
588 the système national d'information interrégimes de l'Assurance Maladie (SNIIRAM)  
589 to the système national des données de santé (SNDS) in France. *Revue*  
590 *d'Épidémiologie et de Santé Publique*, Réseau REDSIAM 65, S149–S167.  
591 <https://doi.org/10.1016/j.respe.2017.05.004>

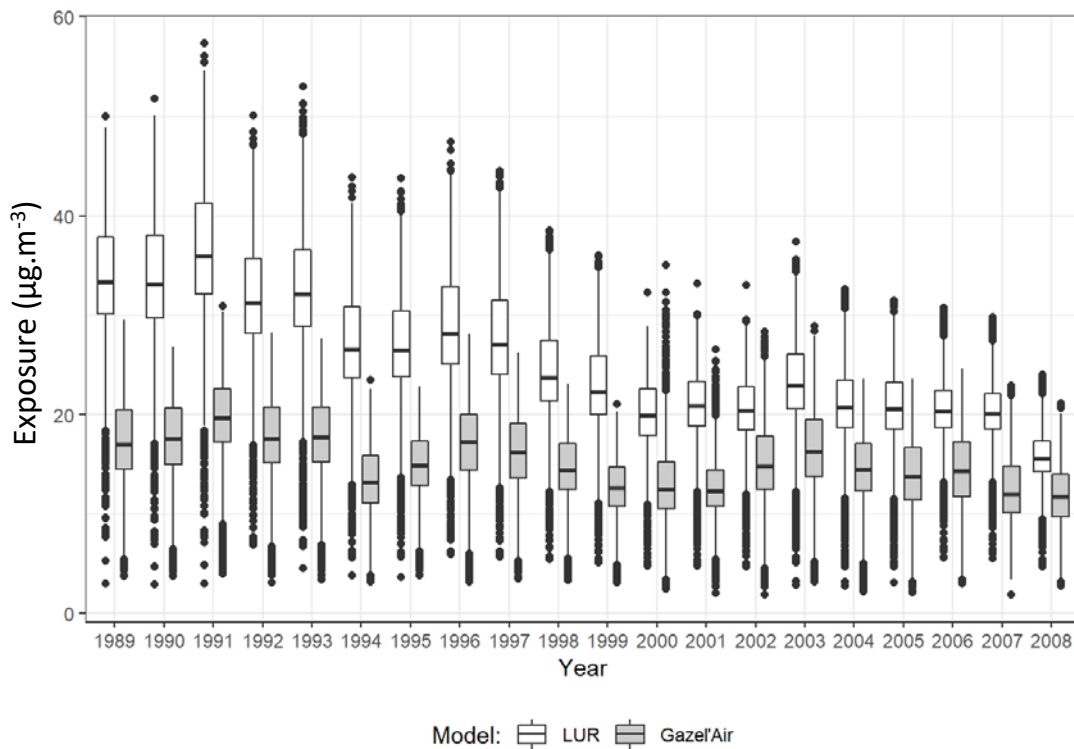
592 Turner, M.C., Krewski, D., Diver, W.R., Pope, C.A., Burnett, R.T., Jerrett, M., Marshall, J.D.,  
593 Gapstur, S.M., 2017. Ambient Air Pollution and Cancer Mortality in the Cancer  
594 Prevention Study II. *Environ. Health Perspect.* 125, 087013.  
595 <https://doi.org/10.1289/EHP1249>

596 van Buuren, S., Groothuis-Oudshoorn, K., 2011. mice: Multivariate Imputation by Chained  
597 Equations in R. *Journal of Statistical Software* 45, 1–67.  
598  
599



600 **List of figures**

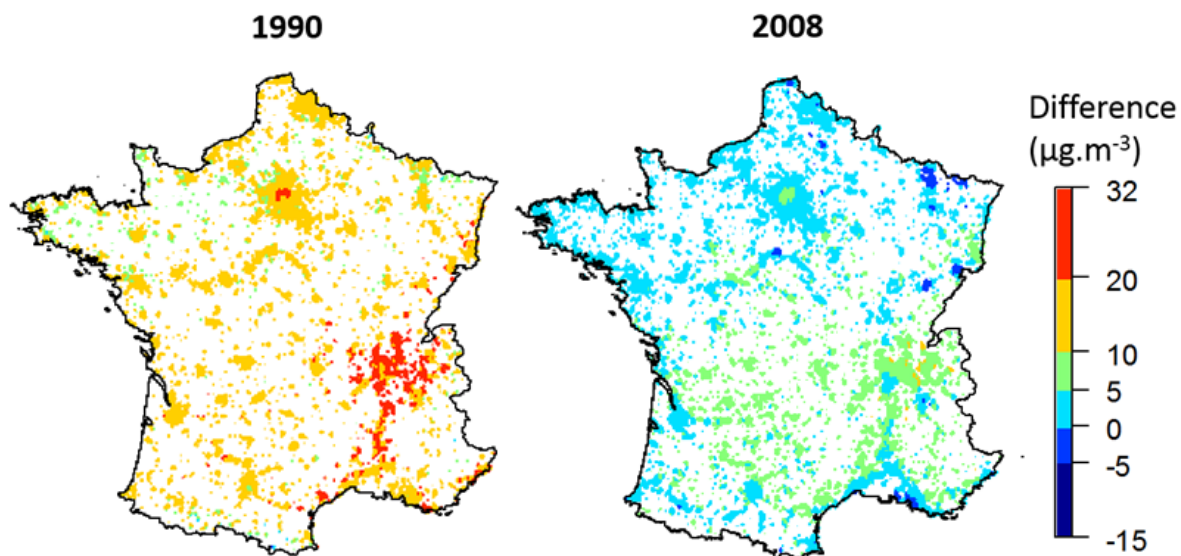
601



602

603 **Figure 1: distribution of PM<sub>2.5</sub> concentrations for Gazel participants (19,348)**  
604 **from 1989 to 2008 (without any time lag) according to the LUR (white) or the**  
605 **Gazel-Air (grey) exposure assessment (the boxes depict the 1<sup>st</sup> and 3<sup>rd</sup> quartiles with the**  
606 **median as the internal dash, the lines stretch to the median±1.5 interquartile range and the**  
607 **points represent values beyond these cutoffs).**

608



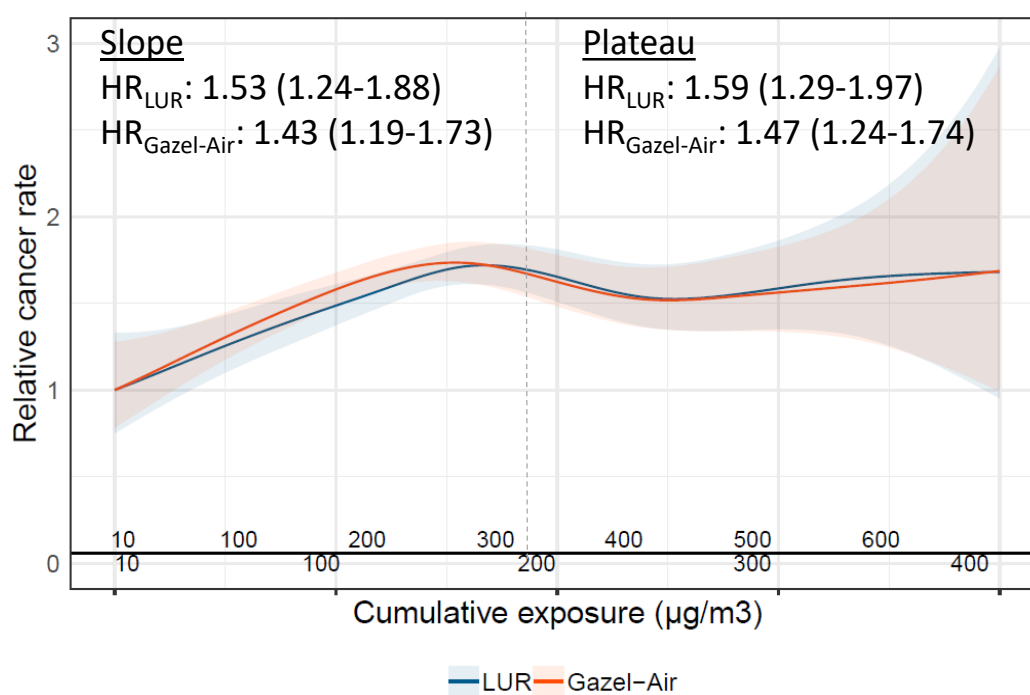
609

610 **Figure 2: Differences between the LUR and the Gazel-Air exposure assessment**  
 611 **(with LUR estimates as the reference) for all available Gazel participants in**  
 612 **1990 (left) and 2008 (right).** For example, the red values that the LUR estimates were up  
 613 to  $32\mu\text{g.m}^{-3}$  higher than the Gazel-Air estimates (e.g. in 1990 in the Southeastern region), and  
 614 the dark blue values indicate that the LUR estimates were up to  $15\mu\text{g.m}^{-3}$  lower than the  
 615 Gazel-Air estimates (e.g. in small areas of the Northeastern region in 2008).

616

617

618



619

620 **Figure 3: Associations between cumulative  $PM_{2.5}$  estimated by the LUR or the**  
621 **Gazel-Air exposure assessments and all-site incident cancer in the Gazel cohort,**  
622 **with 293,188 person-years and 3,711 incident cancer cases.** The y-axis represents the  
623 Hazard ratios (HR) and confidence intervals with the lowest exposure as the reference for each  
624 exposure assessment method using Cox models with a spline function for the exposure, and  
625 with attained age as underlying time-scale and time-dependent variables, adjusted for sex,  
626 cumulative smoking pack-years, passive smoking, alcohol use, BMI, education, socioeconomic  
627 status, family status, fruit and vegetable consumption, occupational exposure to nine lung  
628 carcinogens, age at inclusion and calendar time. The “Slope” HRs on the left represent the slope  
629 of the curve below the 65th percentile exposure (vertical dashed line, 315  $\mu g/m^3$  for the LUR,  
630 185  $\mu g/m^3$  for Gazel-Air) and are expressed per one IQR increase in cumulative exposure to  
631  $PM_{2.5}$ . The “Plateau” HRs on the right represent the plateau above the 65th percentile of  
632 exposure, calculated for the 80th percentile of each exposure assessment method (380  $\mu g/m^3$  for  
633 the LUR, 215  $\mu g/m^3$  for Gazel-Air). IQR= 216  $\mu g/m^3$  for the LUR, 127  $\mu g/m^3$  for Gazel-Air. The x-

26

634 axis represents the cumulative exposure for both exposure assessment methods, the LUR at  
635 the top of the axis and Gazel-Air at the bottom.

636

## 637 **List of tables**

638 *Table 1: Sociodemographic characteristics of the 19,348 Gazel participants included in this*  
639 *study, stratified by their diagnosis of incident primary cancer (all-sites). Number*  
640 *(percentage) of participants or median [interquartile range].*

	Incident cancer during follow-up period	
	No	Yes
	15637	3711
<b>Follow-up time</b> (years)	27.0 [27.0, 27.0]	20.3 [17.0, 23.7]
<b>Age at enrollment</b> (years)	43.5 [41.0, 46.5]	44.5 [42.0, 47.0]
<b>Sex</b> (male)	11147 (71.3)	2956 (79.7)
<b>Smoking status</b>		
Nonsmoker	6859 (44.2)	1425 (38.8)
Former smoker	4442 (28.6)	1075 (29.2)
Current smoker	4210 (27.1)	1177 (32.0)
<b>Smoking - cumulative pack-years*</b>	15.0 [7.5, 25.7]	17.7 [8.5, 29.6]
<b>Passive smoking</b> (yes)	10623 (78.1)	2507 (78.0)
<b>Alcohol use</b>		
Abstinent	391 (2.5)	104 (2.8)
Light drinker	11170 (71.4)	2421 (65.3)
Moderate drinker	2386 (15.3)	654 (17.6)
Heavy drinker	1370 (8.8)	448 (12.1)
Unclear pattern	317 (2.0)	83 (2.2)

**Occupational exposure\*\***

None	9443 (60.4)	2100 (56.6)
One	1319 (8.4)	328 (8.8)
Two	1697 (10.9)	448 (12.1)
Three or more	3178 (20.3)	835 (22.5)

**Education**

9-11 years	11238 (73.5)	2720 (74.8)
12-13 years	1144 (7.5)	227 (6.2)
14-15 years	887 (5.8)	207 (5.7)
Other secondary education	1668 (10.9)	398 (10.9)
Other diploma	357 (2.3)	83 (2.3)

**Socioeconomic status**

Low	2762 (17.7)	621 (16.7)
Intermediate	9179 (58.8)	2130 (57.4)
High	3677 (23.5)	958 (25.8)

<b>Family status</b> (not single)	13722 (89.0)	3288 (90.0)
-----------------------------------	--------------	-------------

<b>Body mass index</b>	25.0 [23.2, 27.0]	26.3 [22.6, 30.0]
------------------------	-------------------	-------------------

**Vegetable & fruit consumption**

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 everyday	2562 (22.6)	641 (23.5)
Every day or almost	7912 (69.7)	1851 (67.8)

---

\*: pack-years for ever-smokers only; \*\*: to nine selected lung carcinogens

642 *Table 2: Spearman's correlation coefficients between the LUR and the Gazel-Air exposure*  
 643 *estimates assessed for the 19,348 eligible Gazel participants across years, then stratified by*  
 644 *European NUTS-1 region or by distance to the nearest major road, over the full follow-up or*  
 645 *during specific years.*

<b>Variable</b>	<b>Category</b>	<b>All years</b>	<b>1990</b>	<b>2000</b>	<b>2008</b>
All combined		0.71	0.70	0.70	0.68
NUTS-1	FR1 - Région parisienne	0.71	0.78	0.54	0.58
region	FR2 - Bassin parisien	0.75	0.51	0.36	0.48
	FR3 - Nord	0.73	0.75	0.77	0.71
	FR4 - Est	0.59	0.23	0.36	0.23
	FR5 - Ouest	0.71	0.63	0.58	0.53
	FR6 - Sud-Ouest	0.63	0.76	0.61	0.62
	FR7 - Centre-Est	0.73	0.80	0.77	0.73
	FR8 - Méditerranée	0.59	0.72	0.67	0.73
Distance to	<150m	0.73	0.77	0.70	0.73
the nearest	150-365m	0.70	0.69	0.67	0.69
major road	365-830m	0.70	0.69	0.70	0.66
	830-1730m	0.71	0.70	0.68	0.66
	>1730m	0.70	0.61	0.65	0.62

European NUTS-1 regions refer to areas defined by the European Union.

The number of participants slightly decreased from 1990 onwards as some of them have been censored, and participants moved to other NUTS-1 regions during the follow-up.

646

647

648 *Table 3: main associations between cumulative long-term exposure to PM<sub>2.5</sub> and first-*  
 649 *occurring all-site and lung incident cancer in the Gazel cohort using two exposure models.*  
 650 *Hazard Ratios (HR) and their 95% Confidence Interval (CI) correspond to the increased risk*  
 651 *of cancer below the 65<sup>th</sup> percentile of cumulative exposure for all-site cancer and 73<sup>rd</sup> for lung*  
 652 *cancer (“slope”) per one IQR increase of cumulative exposure, and above these percentiles*  
 653 *(“plateau”).*

654

Cancer	Cumulative exposure	LUR		Gazel-Air	
		HR	CI	HR	CI
All-site	<p65* (slope)	1.53	1.24-1.88	1.43	1.19-1.73
	>p65 **(plateau)	1.59	1.29-1.97	1.47	1.24-1.74
Lung	<p73* (slope)	1.25	0.58-2.29	1.79	0.94-3.40
	>p73 **(plateau)	1.04	0.07-16.1	1.94	0.25-15.29

Extended Cox model with attained age as time axis, and with time-varying exposure to cumulative exposure to PM<sub>2.5</sub>, adjusted for sex (included with a strata function), age at enrollment, calendar time (time-varying, dichotomized at year 2007), cumulative pack-years (time-varying), passive smoking, alcohol intake (time-varying), socioeconomic status, marital status (time-varying), body mass index (time-varying), occupational exposure to lung carcinogens, and consumption of fruit and vegetable (time-varying). All-site cancer: results obtained with 3,711 cases of all-site cancer and 293,188 person-years. Lung cancer: results obtained with 349 cases of all-site cancer and 293,188 person-years. \*p65 is the 65<sup>th</sup> percentile of cumulative exposure and corresponds to 315 µg/m<sup>3</sup> and to 185 µg/m<sup>3</sup> for the LUR and the Gazel-Air exposure assessments, respectively; p73 is the 73<sup>th</sup> percentile of cumulative exposure and corresponds to 355 µg/m<sup>3</sup> and to 200 µg/m<sup>3</sup> for the LUR and the Gazel-Air exposure assessments, respectively. IQR= 216 µg/m<sup>3</sup> for the LUR, 127 µg/m<sup>3</sup> for Gazel-Air.\*\*: the HR and CI were calculated for the 80<sup>th</sup> percentile (380 µg/m<sup>3</sup> for the LUR, 215 µg/m<sup>3</sup> for Gazel-Air).

655

656

657 *Table 4: Sensitivity analyses for associations between long-term exposure to PM<sub>2.5</sub> and first-*  
658 *occurring all-site incident cancer in the Gazel cohort using two exposure models. Hazard*  
659 *Ratios (HR) and their 95% Confidence Interval (CI) correspond to the increased risk of cancer*  
660 *below the 65<sup>th</sup> percentile of exposure (“slope”) per one IQR increase of cumulative exposure)*  
661 *and above this 65<sup>th</sup> percentile (“plateau”).*

Exposure	Cases	LUR				Gazel-Air			
		<p65* (slope)		>p65 (plateau) **		<p65 (slope)*		>p65 (plateau)**	
Sensitivity analyses	Cases	HR	95% CI	HR	95%CI	HR	95% CI	HR	95%CI
Further adjusted for area-level deprivation	3,711	1.59	1.29-1.95	1.67	1.35-2.06	1.47	1.22-1.77	1.50	1.27-1.78
Using address-level geocodes	1,014	1.59	1.03-2.46	1.43	0.59-3.51	1.37	0.93-2.01	1.32	0.67-2.62
Complete-case analysis	1,881	1.40	1.03-1.90	1.33	0.84-2.10	1.48	1.13-1.93	1.44	1.03-2.05
Missing data as category	3,711	1.53	1.23-1.89	1.61	1.28-2.01	1.41	1.17-1.71	1.45	1.22-1.73

Extended Cox model with attained age as time axis, and with time-varying exposure to cumulative exposure to PM<sub>2.5</sub>, adjusted for sex (included with a strata function), age at enrollment, calendar time (time-varying, dichotomized at year 2007), cumulative pack-years (time-varying), passive smoking, alcohol intake (time-varying), socioeconomic status, marital status (time-varying), body mass index (time-varying), occupational exposure to lung carcinogens, and consumption of fruit and vegetable (time-varying). \*p65 is the 65<sup>th</sup> percentile of cumulative exposure and corresponds to 315 µg/m<sup>3</sup> and to 185 µg/m<sup>3</sup> for the LUR and the Gazel-Air exposure assessments, respectively. IQR= 216 µg/m<sup>3</sup> for the LUR, 127 µg/m<sup>3</sup> for Gazel-Air.\*\*: the HR and CI were calculated for the 80<sup>th</sup> percentile (380 µg/m<sup>3</sup> for the LUR, 215 µg/m<sup>3</sup> for Gazel-Air).

662

663

664

665