



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-03576594>

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

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

Authors: Emeline Lequy^{1,2*}, Mohammad Javad Zare Sakhvidi³, Danielle Vienneau^{4,5}, Kees de Hoogh^{4,5}, Jie Chen⁶, Jean-François Dupuy⁷, Valérie Garès⁷, Emilie Burte³, Olivier Bouaziz⁸, Alain Le Tertre⁹, Vèrène Wagner⁹, Ole Hertel¹⁰, Jesper Heile Christensen¹⁰, Sergey Zhivin¹, Jack Siemiatycki², Marcel Goldberg¹, Marie Zins¹, Bénédicte Jacquemin³

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

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

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

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

5: University of Basel, Basel, Switzerland.

6: Utrecht University, Utrecht, the Netherlands

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

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

9: Santé publique France, Saint-Maurice, France

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

*Address correspondence to:

Emeline Lequy, UMS 011, Hôpital Paul Brousse, 16 avenue Paul Vaillant Couturier, 94807 VILLEJUIF CEDEX, France. Email: e.lequy@gmail.com. Phone number: +33 (0)1 77 74 74 23.

Declaration of competing financial interests:

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

Abstract

Background: Many studies investigated the relationship between outdoor fine particulate matter (PM_{2.5}) and cancer. While they generally indicated positive associations, results have not been fully consistent, possibly because of the diversity of methods used to assess exposure.

Objectives: To investigate how using different PM_{2.5} exposure assessment methods influences risk estimates in the large French general population-based Gazel cohort (20,625 participants at enrollment) with a 26-year follow-up with complete residential histories.

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

Results: We obtained similar two-piece linear associations for all-sites cancer (3,711 cases), 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 cumulative PM_{2.5} exposure for the LUR and the Gazel-Air models respectively, followed by a plateau at around 1.5 for both exposure assessments. For lung cancer (349 cases), the HRs from the two exposure models were less similar, with largely overlapping confidence limits.

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

Keywords: air pollution; PM_{2.5}; lung cancer; survival analysis; epidemiology

Introduction

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

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

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

Material and methods

Study population

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

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

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

Cancer incidence

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

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

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

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

Exposure assessment

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

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

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

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

For each participant and each available year, we estimated PM_{2.5} exposure as the value at the residential address for both exposure assessment methods.

Covariables

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

Sociodemographic and occupational variables: sex, education (9-11 years, 12-13 years, 14-15 years, other secondary education, other diploma), and socio-economic status (low: blue-collar workers or clerks; medium: first-line supervisors or sales representatives; high: management), all at baseline. We also used Gazel participants' occupational histories linked to a French job exposure matrix (Imbernon, 1991) to include 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 in the company and categorized into none, one, two, or at least three carcinogens .

Lifestyle variables: time-varying variables for tobacco (cumulative smoking pack-years and smoking status), alcohol intake (abstinent, non-daily drinker, light drinker, moderate drinker, heavy drinker, unclear pattern), family status (single or not), body mass index (BMI in kg/m²). Two variables were included only at specific time points: passive smoking at home or at work (yes or no) in 1990 and 1996, and fruit-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 computed these variables as time-varying, attributing the data collected in 1990 and 1996 to each year of the periods 1989-1995 and 1996-2015 for passive smoking, and the data collected in 1998, 2004, 2009 and 2014 to each year of the periods 1989-1998, 1999-2004, 2005-2010, 2011-2015 for fruit and vegetable consumption.

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

Imputations

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

Statistics

Comparing estimates of exposure to PM_{2.5}:

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

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

Associations between PM_{2.5} and cancer, using exposure derived from the two models:

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

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

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

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

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

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

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

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

Results

Characteristics of Gazel's participants

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

Comparing the LUR and Gazel-Air exposure models

The common period in which the two models provided exposures spanned from 1989 to 2008. The median exposure to PM_{2.5} in 1989 was 33.3 µg.m⁻³ [interquartile range: 30.1-37.8] and 17.0 µg.m⁻³ [14.5-20.5] for the LUR and the Gazel-Air model, respectively. We observed a

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

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

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

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

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

Associations between PM_{2.5} exposure and cancer using two exposure assessment methods

All-sites combined cancer incidence:

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

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

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

Lung cancer incidence:

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

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 estimates per 1-unit increase of natural log-transformed cumulative PM_{2.5}.

Discussion

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

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

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

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

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

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

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

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

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

Conclusions

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

estimate is more accurate; however, our findings suggest that the epidemiological findings related to PM_{2.5} and the investigated cancers are robust.

Acknowledgments

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

References

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

- Cohen, G., Steinberg, D.M., Yuval, Levy, I., Chen, S., Kark, J.D., Levin, N., Witberg, G., Bental, T., Broday, D.M., Kornowski, R., Gerber, Y., 2019. Cancer and mortality in relation to traffic-related air pollution among coronary patients: Using an ensemble of exposure estimates to identify high risk individuals. *Environ. Res.* 176, 108560. <https://doi.org/10.1016/j.envres.2019.108560>
- de Hoogh, K., Chen, J., Gulliver, J., Hoffmann, B., Hertel, O., Ketzel, M., Bauwelinck, M., van Donkelaar, A., Hvidtfeldt, U.A., Katsouyanni, K., Klompmaker, J., Martin, R.V., Samoli, E., Schwartz, P.E., Stafoggia, M., Bellander, T., Strak, M., Wolf, K., Vienneau, D., Brunekreef, B., Hoek, G., 2018. Spatial PM_{2.5}, NO₂, O₃ and BC models for Western Europe - Evaluation of spatiotemporal stability. *Environ Int* 120, 81–92. <https://doi.org/10.1016/j.envint.2018.07.036>
- Dockery, D.W., Pope, C.A., Xu, X., Spengler, J.D., Ware, J.H., Fay, M.E., Ferris, B.G.Jr., Speizer, F.E., 1993. An Association between Air Pollution and Mortality in Six U.S. Cities. *New England Journal of Medicine* 329, 1753–1759. <https://doi.org/10.1056/NEJM199312093292401>
- Goldberg, M., Leclerc, A., Bonenfant, S., Chastang, J.F., Schmaus, A., Kaniewski, N., Zins, M., 2007. Cohort profile: the GAZEL Cohort Study. *Int J Epidemiol* 36, 32–39. <https://doi.org/10.1093/ije/dyl247>
- Goldberg, M., Leclerc, A., Zins, M., 2015. Cohort Profile Update: The GAZEL Cohort Study. *Int J Epidemiol* 44, 77–77g. <https://doi.org/10.1093/ije/dyu224>
- 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.
- Jerrett, M., Turner, M.C., Beckerman, B.S., Pope III, C.A., van Donkelaar, A., Martin, R.V., Serre, M., Crouse, D., Gapstur, S.M., Krewski, D., Diver, W.R., Coogan, P.F., Thurston, G.D., Burnett, R.T., 2017. Comparing the Health Effects of Ambient Particulate Matter Estimated Using Ground-Based versus Remote Sensing Exposure

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

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

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

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

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

Lequy, E., Siemiatycki, J., de Hoogh, K., Vienneau, D., Dupuy, J.-F., Garès, V., Hertel, O., Christensen, J.H., Zhivin, S., Goldberg, M., Zins, M., Jacquemin, B., 2021. Contribution of Long-Term Exposure to Outdoor Black Carbon to the Carcinogenicity of Air Pollution: Evidence regarding Risk of Cancer in the Gazel Cohort. *Environ 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
 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

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

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

Rey, G., Jougl, 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. <https://doi.org/10.1186/1471-2458-9-33>

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

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

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

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

- Nieuwenhuijsen, M., Pin, I., Siroux, V., Stempfelet, M., Tsai, M.-Y., Probst-Hensch, N., Jacquemin, B., 2017. Socioeconomic position and outdoor nitrogen dioxide (NO₂) exposure in Western Europe: A multi-city analysis. *Environ Int* 101, 117–124.
<https://doi.org/10.1016/j.envint.2016.12.026>
- Therneau, T.M., 2015. A Package for Survival Analysis in S.
- Therneau, T.M., Grambsch, P.M., 2000. Modeling Survival Data: Extending the Cox Model. Springer, New York.
- Tuppin, P., Rudant, J., Constantinou, P., Gastaldi-Ménager, C., Rachas, A., de Roquefeuil, L., Maura, G., Caillol, H., Tajahmady, A., Coste, J., Gissot, C., Weill, A., Fagot-Campagna, A., 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. *Revue d'Épidémiologie et de Santé Publique*, Réseau REDSIAM 65, S149–S167.
<https://doi.org/10.1016/j.respe.2017.05.004>
- Turner, M.C., Krewski, D., Diver, W.R., Pope, C.A., Burnett, R.T., Jerrett, M., Marshall, J.D., Gapstur, S.M., 2017. Ambient Air Pollution and Cancer Mortality in the Cancer Prevention Study II. *Environ. Health Perspect.* 125, 087013.
<https://doi.org/10.1289/EHP1249>
- van Buuren, S., Groothuis-Oudshoorn, K., 2011. mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software* 45, 1–67.

List of figures

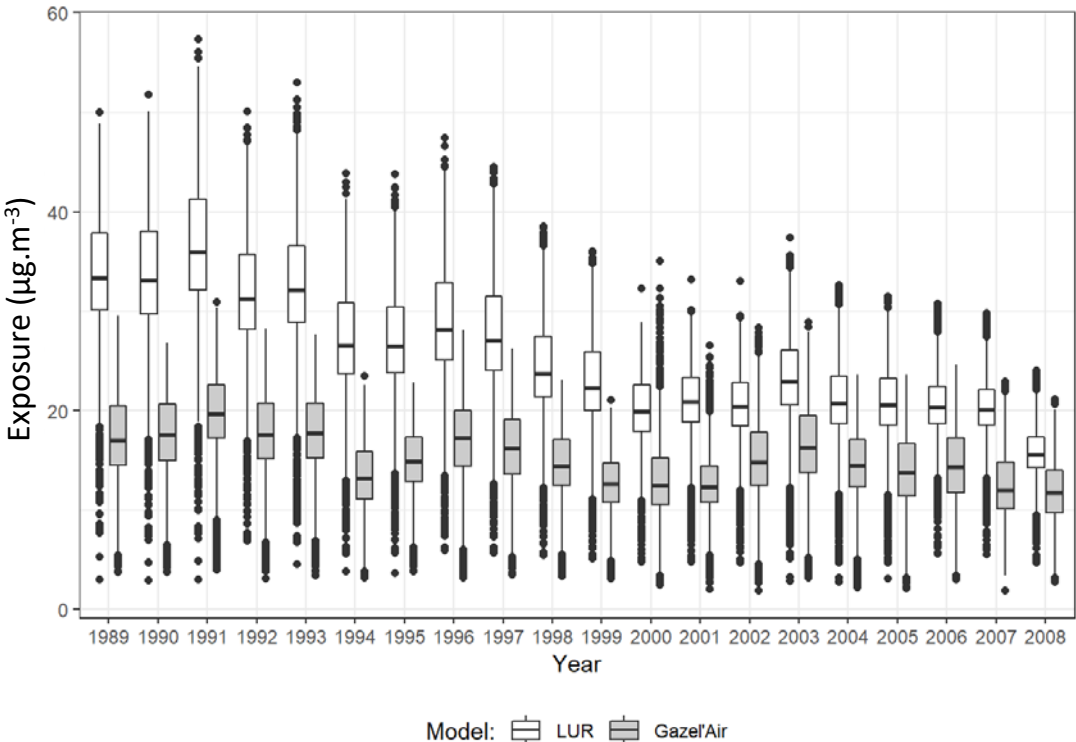


Figure 1: distribution of PM_{2.5} concentrations for Gazel participants (19,348) from 1989 to 2008 (without any time lag) according to the LUR (white) or the Gazel-Air (grey) exposure assessment (the boxes depict the 1st and 3rd quartiles with the median as the internal dash, the lines stretch to the median \pm 1.5 interquartile range and the points represent values beyond these cutoffs).

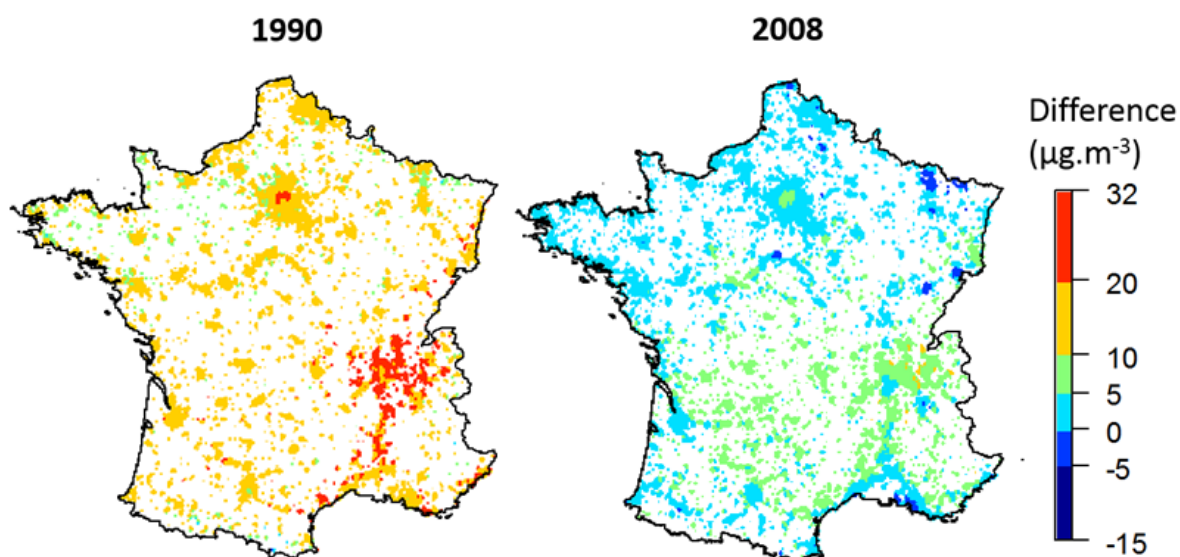


Figure 2: Differences between the LUR and the Gazel-Air exposure assessment (with LUR estimates as the reference) for all available Gazel participants in 1990 (left) and 2008 (right). For example, the red values that the LUR estimates were up to $32\mu\text{g.m}^{-3}$ higher than the Gazel-Air estimates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

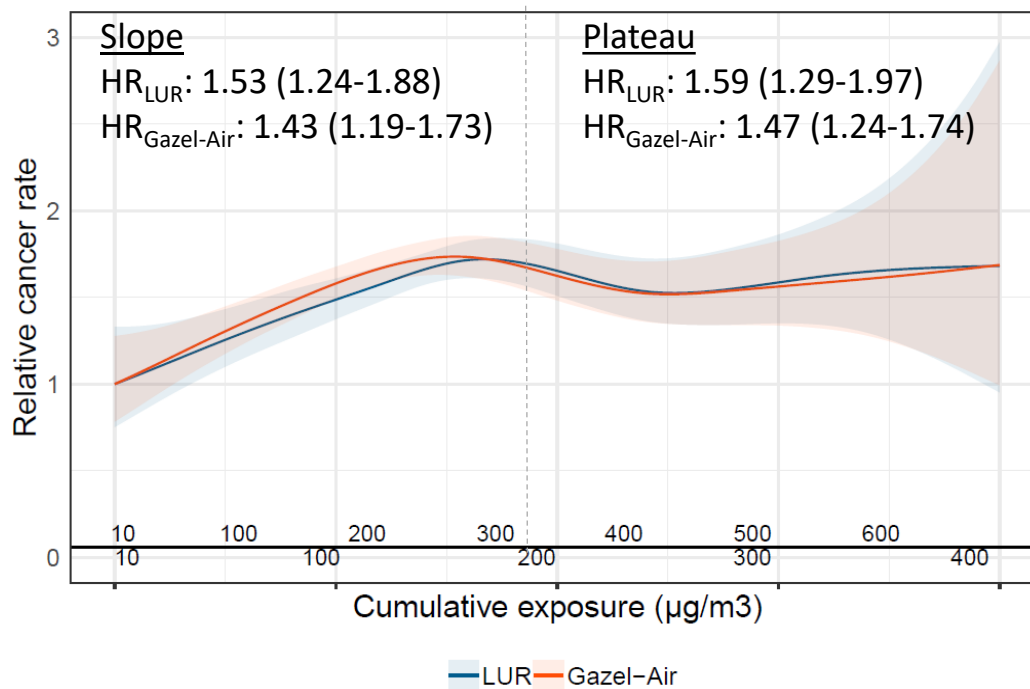


Figure 3: Associations between cumulative PM_{2.5} estimated by the LUR or the Gazel-Air exposure assessments and all-site incident cancer in the Gazel cohort, with 293,188 person-years and 3,711 incident cancer cases. The y-axis represents the Hazard ratios (HR) and confidence intervals with the lowest exposure as the reference for each exposure assessment method using Cox models with a spline function for the exposure, and 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 nine lung carcinogens, age at inclusion and calendar time. The “Slope” HRs on the left represent the slope of the curve below the 65th percentile exposure (vertical dashed line, 315 µg/m³ for the LUR, 185 µg/m³ for Gazel-Air) and are expressed per one IQR increase in cumulative exposure to PM_{2.5}. The “Plateau” HRs on the right represent the plateau above the 65th percentile of exposure, calculated for the 80th percentile of each exposure assessment method (380 µg/m³ for the LUR, 215 µg/m³ for Gazel-Air). IQR= 216 µg/m³ for the LUR, 127 µg/m³ for Gazel-Air. The x-

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

List of tables

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

	Incident cancer during follow-up period	
	No	Yes
	15637	3711
Follow-up time (years)	27.0 [27.0, 27.0]	20.3 [17.0, 23.7]
Age at enrollment (years)	43.5 [41.0, 46.5]	44.5 [42.0, 47.0]
Sex (male)	11147 (71.3)	2956 (79.7)
Smoking status		
Nonsmoker	6859 (44.2)	1425 (38.8)
Former smoker	4442 (28.6)	1075 (29.2)
Current smoker	4210 (27.1)	1177 (32.0)
Smoking - cumulative pack-years*	15.0 [7.5, 25.7]	17.7 [8.5, 29.6]
Passive smoking (yes)	10623 (78.1)	2507 (78.0)
Alcohol use		
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)

Family status (not single)	13722 (89.0)	3288 (90.0)
-----------------------------------	--------------	-------------

Body mass index	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

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

Variable	Category	All years	1990	2000	2008
All combined		0.71	0.70	0.70	0.68
NUTS-1 region	FR1 - Région parisienne	0.71	0.78	0.54	0.58
	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 the nearest major road	<150m	0.73	0.77	0.70	0.73
	150-365m	0.70	0.69	0.67	0.69
	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.

Table 3: main associations between cumulative long-term exposure to PM_{2.5} and first-occurring all-site and lung incident cancer in the Gazel cohort using two exposure models. Hazard Ratios (HR) and their 95% Confidence Interval (CI) correspond to the increased risk of cancer below the 65th percentile of cumulative exposure for all-site cancer and 73rd for lung cancer (“slope”) per one IQR increase of cumulative exposure, and above these percentiles (“plateau”).

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_{2.5}, 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 65th percentile of cumulative exposure and corresponds to 315 µg/m³ and to 185 µg/m³ for the LUR and the Gazel-Air exposure assessments, respectively; p73 is the 73th percentile of cumulative exposure and corresponds to 355 µg/m³ and to 200 µg/m³ for the LUR and the Gazel-Air exposure assessments, respectively. IQR= 216 µg/m³ for the LUR, 127 µg/m³ for Gazel-Air. **: the HR and CI were calculated for the 80th percentile (380 µg/m³ for the LUR, 215 µg/m³ for Gazel-Air).

Table 4: Sensitivity analyses for associations between long-term exposure to PM_{2.5} and first-occurring all-site incident cancer in the Gazel cohort using two exposure models. Hazard Ratios (HR) and their 95% Confidence Interval (CI) correspond to the increased risk of cancer below the 65th percentile of exposure (“slope”) per one IQR increase of cumulative exposure) and above this 65th percentile (“plateau”).

Exposure		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_{2.5}, 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 65th percentile of cumulative exposure and corresponds to 315 µg/m³ and to 185 µg/m³ for the LUR and the Gazel-Air exposure assessments, respectively. IQR= 216 µg/m³ for the LUR, 127 µg/m³ for Gazel-Air.**: the HR and CI were calculated for the 80th percentile (380 µg/m³ for the LUR, 215 µg/m³ for Gazel-Air).

662

663

664

665