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 Sakhvidi[3], Danielle Vienneau[4,5], Kees de Hoogh[4,5], Jie Chen[6], Jean-François Dupuy[7], Valérie Garès[7], Emilie Burte[3], Olivier Bouaziz[8], Alain Le Tertre[9], Vérène Wagner[9], Ole Hertel[10], Jesper Heile Christensen[10], Sergey Zhivin[1], Jack Siemiatycki[2], Marcel Goldberg[1], Marie Zins[1], Bénédicte Jacquemin[3]

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.

Table of Content

Supplementary Methods 3
Lag period 3
Specificities of the Cox model 3
Hazard ratios of the two-piece linear model 3
Supplementary Figures 4
Supplementary Tables 10

Supplementary Methods

Lag period

We implemented a 10-year lag period, since 10 years is the period estimated for a cancer to develop, except hematopoietic cancers with a small frequency in our dataset.

Specificities of the Cox model

We organized the data in a counting process style, with one entry per available year per participant until incidence/censoring. In our dataset, PM2.5 concentrations overall decreased with time, while participants aged and cancer incidence risk increased. For these reasons, using annual exposure variables may lead to spurious associations. Besides, we were interested in the association of cancer incidence and long-term PM2.5 exposure, available at an annual time step. The extended Cox model with time-dependent variables estimates the instantaneous risk with the exposure at the date of the event: using only the annual levels of exposure may not estimate the association between incident cancer and long-term exposure. We addressed these problems by using cumulative exposures for each participant from baseline to incidence or censoring, as previously used (Lequy et al., 2021), and by adjusting for calendar time, dichotomized with a cut-off in 2007.

Hazard ratios of the two-piece linear model

We provided two types of hazard ratios (HR) and their 95% confidence intervals: the first HR will be that provided directly by the model and corresponds to the slope of the first part of the curve (per one IQR increase of cumulative exposure); the second will approximate the value of the plateau using the coefficients of the Cox model to calculate the HR and 95% CI at the 80^{th} percentile of cumulative exposure (380 µg/m^3 for the LUR, 215 µg/m^3 for Gazel-Air).

Supplementary Figures

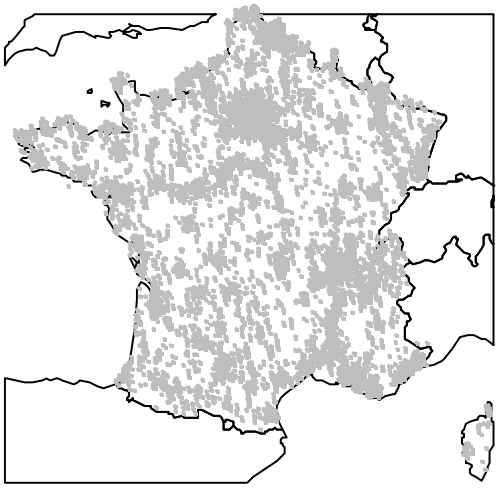

Figure S1: residential places of Gazel participants (gray dots) in mainland France in 1989 (n=20,430).

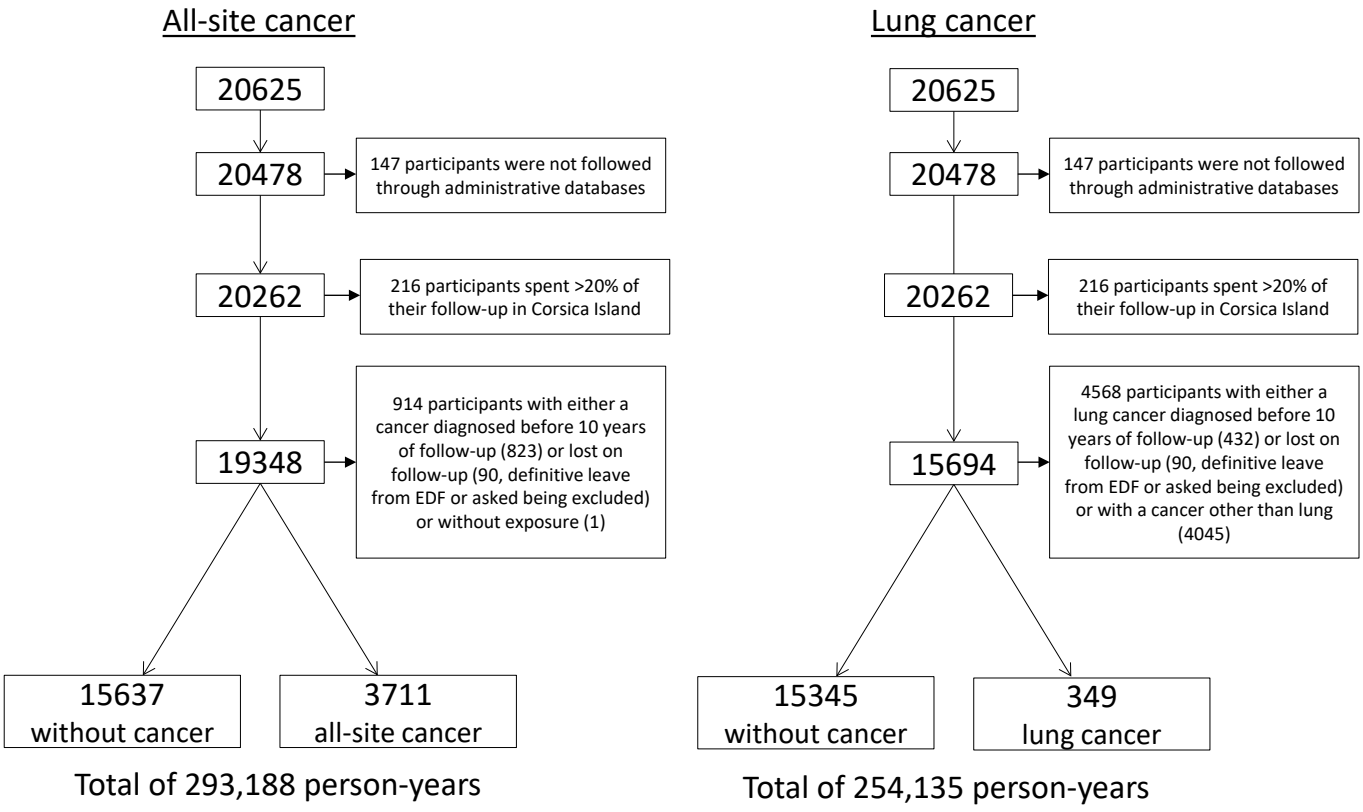

Figure S2: flowchart of the selection of Gazel participants included in the analyses pertaining to all-site and lung cancer.

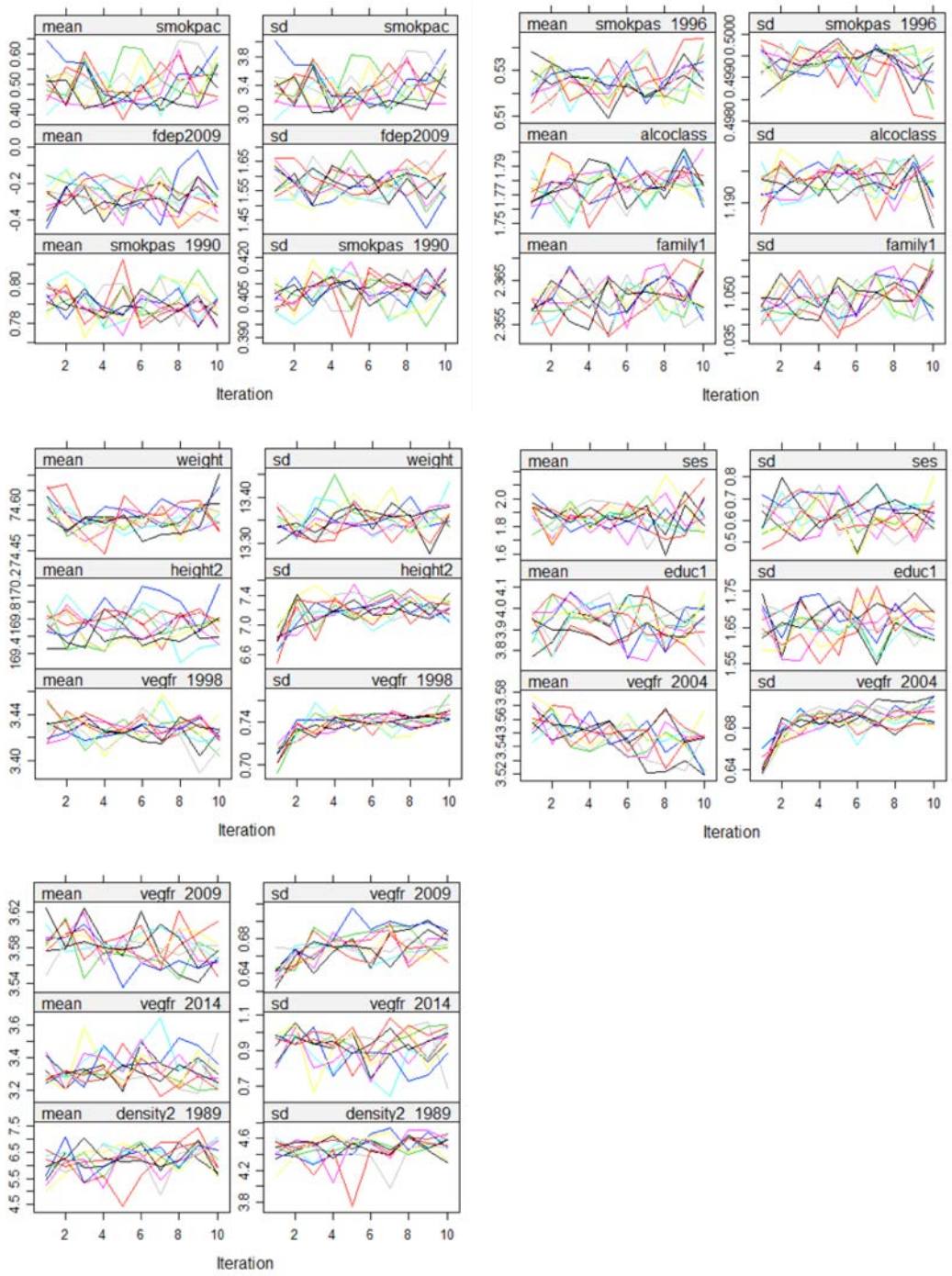

Figure S3: Mean and standard deviation of the synthetic values plotted against iteration number for the imputed data. Smokpac: smoking pack-years; fdep2009: deprivation index (put back as nonimputed values in the final database); smokpas: passive smoking at work or at home; alcoclass: alcohol consumption; family1: marital status; vegfr: fruit and vegetable intake; ses: socioeconomic status; educ1: education; density2: urban classification (put back as nonimputed values in the final database).

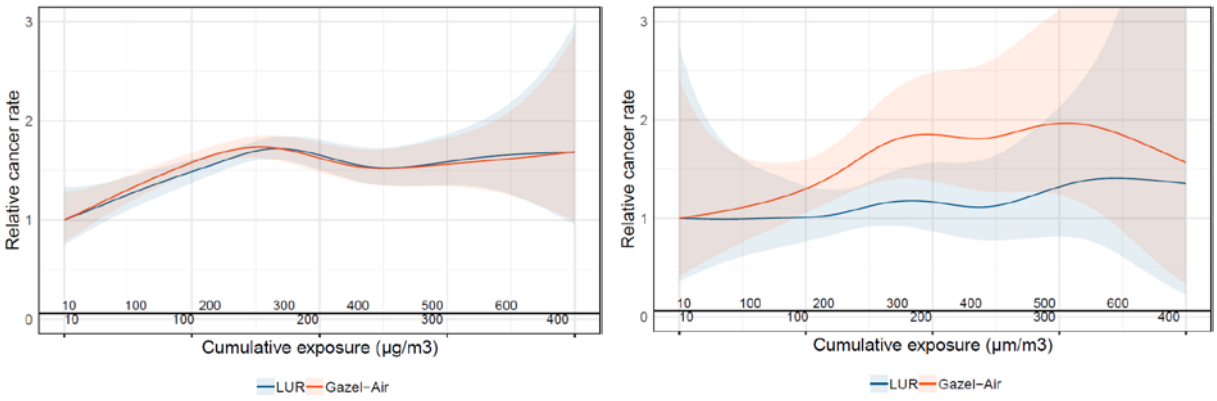

Figure S4: Associations between cumulative PM2.5 and all-site incident cancer (left, with 293,188 person-years and 3,711 incident cancer cases) and lung incident cancer (right, with 254,135 person-years and 349 incident cancer cases) in the Gazel cohort, expressed as hazard ratio (HR) with the lowest exposure as the reference for each exposure assessment method. Cumulative exposure to PM2.5 was estimated by the LUR or the Gazel-Air exposure assessments in separate Cox models with attained age as underlying time-scale and time-dependent variables, adjusted for sex, cumulative smoking pack-years, passive smoking, alcohol use, BMI, education, socioeconomic status, family status, fruit and vegetable consumption, occupational exposure to lung carcinogens, age at inclusion and calendar time. 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.

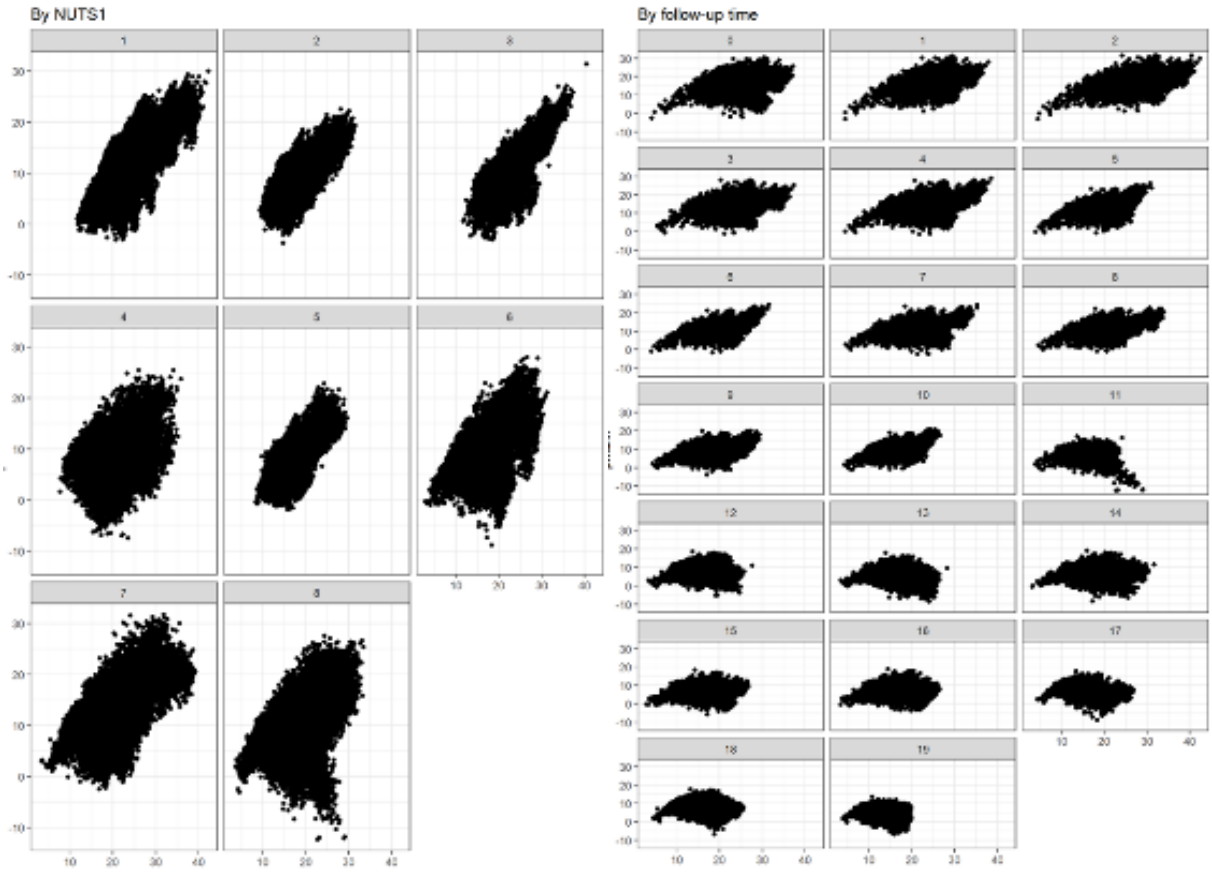

Figure S5: Bland-Altman graphs depicting the differences (y-axis) between the LUR and the Gazel-Air exposure values for Gazel participant in function of the mean exposure (x-axis) between the two exposure assessment methods, whether by European NUTS-1 region (combining all the years) or by year (combining all the regions). From top-left, left panel: NUTS1 from one to eight; from top-left, right panel: follow-up year from 0 (1989) to 19 (2008).

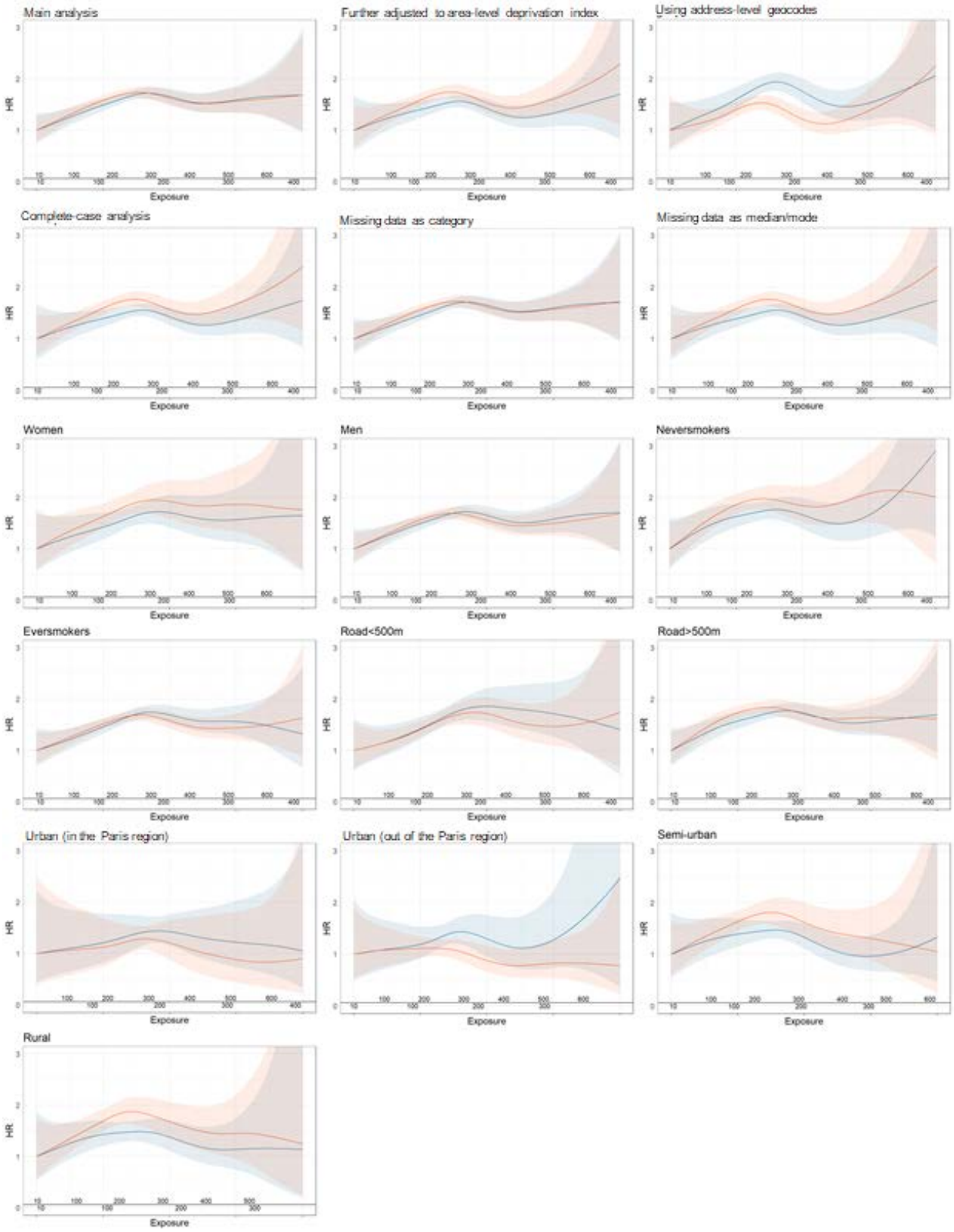

Figure S6: exposure-response relationships between PM2.5 and all-site cancer risk in the Gazel cohort using a spline function for 293,188 person-years (3,711 cases). Main analysis, sensitivity analyses, and effect modification analyses.

Hazard ratios and confidence intervals with the lowest exposure as the reference for each

exposure assessment method were estimated by separate Cox model with attained age as underlying time-scale and time-dependent variables, adjusted for sex, cumulative smoking pack-years, passive smoking, alcohol use, BMI, education, socioeconomic status, family status, vegetable and fruit consumption, occupational exposure to lung carcinogens, age at inclusion and calendar time. Exposures were lagged 10 years. Participants were excluded from the analysis if they were diagnosed with cancer before 1999. 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.

Supplementary Tables

Table S1: associations between long-term exposure to PM2.5 and first-occurring all-site incident cancer in the Gazel cohort using two exposure models using log-transformed cumulative exposure (Hazard Ratios 95% Confidence Interval).

Hazard ratios correspond to the increased risk of cancer for 1-unit increase of the natural log-transformed cumulative exposure to PM2.5.

		LUR		Gazel-Air	
		HR	95% CI	HR	95% CI
Main		1.21	1.07-1.37	1.19	1.07-1.33
Sensitivity analyses					
	Further adjusted for area-level deprivation	1.24	1.10-1.40	1.22	1.09-1.36
	Using address-level geocodes	1.28	0.98-1.66	1.26	1.00-1.58
	Complete-case analysis	1.15	0.93-1.51	1.21	1.02-1.45
	Missing data as category	1.20	1.04-1.38	1.17	1.04-1.32
	Missing data as median/mode	1.22	1.08-1.38	1.21	1.09-1.35
Effect modificiation					
By sex	Women	1.25	0.97-1.61	1.33	1.04-1.69
	Men	1.20	1.04-1.38	1.16	1.02-1.31
By smoking status	Neversmoker	1.24	1.01-1.53	1.28	1.06-1.54
	Eversmoker	1.20	1.03-1.40	1.15	1.00-1.32
By distance to the road:	<500m	1.27	1.01-1.59	1.21	0.99-1.48
	>500m	1.21	1.04-1.41	1.21	1.06-1.38
By urban classification:	Urban (in the Paris region)	1.24	0.84-1.82	1.26	0.83-1.92
	Urban (out of the Paris region)	1.20	0.82-1.74	1.01	0.74-1.40
	Semi-urban	1.17	0.89-1.53	1.21	0.96-1.53
	Rural	1.12	0.85-1.49	1.20	0.94-1.53

Extended Cox model with attained age as time axis, and with time-varying exposure to cumulative exposure to PM2.5, 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).

Table S2: stratified associations between long-term exposure to PM2.5 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 ("slope", per one IQR increase of cumulative exposure) and above ("plateau") the 65th percentile of exposure.

			LUR				Gazel-Air			
	Exposure		<p65* (slope)		>p65 (plateau)		<p65 (slope)		>p65 (plateau)	
Effect modificiation		Cases	HR	95% CI	HR**	CI**	HR	95% CI	HR**	CI**
By sex	Women	755	1.51	0.98-2.30	1.72	0.73-4.07	1.60	1.06-2.41	1.88	0.83-4.04
	Men	2956	1.53	1.21-1.94	1.56	1.18-2.06	1.39	1.13-1.72	1.39	1.12-1.72
By smoking status	Neversmoker	878	1.37	0.96-1.96	1.30	0.72-2.36	1.36	0.98-1.89	1.36	0.84-2.22
	Eversmoker	2833	1.62	1.25-2.11	1.77	1.27-2.46	1.48	1.17-1.86	1.53	1.19-1.98
By distance to the road	<500m	1,244	1.69	1.15-2.48	2.10	1.04-4.26	1.60	1.13-2.26	1.86	1.06-3.26
	>500m	2,467	1.47	1.14-1.89	1.47	1.06-2.06	1.37	1.09-1.72	1.39	1.07-1.80
By urban classification	Urban (in the Parisregion)	423	1.48	0.77-2.84	1.51	0.34-6.80	1.43	0.67-3.06	1.29	0.20-8.49
	Urban (out of the Paris region)	503	1.61	0.85-3.06	1.64	0.25-10.80	1.12	0.63-2.00	0.91	0.20-4.09
	Semi-urban	878	1.27	0.80-2.01	1.01	0.34-3.06	1.49	0.99-2.24	1.44	0.59-3.52
	Rural	722	1.26	0.78-2.03	1.08	0.25-4.61	1.26	0.84-1.89	1.18	0.34-4.09

Extended Cox model with attained age as time axis, and with time-varying exposure to cumulative exposure to PM2.5, 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/m^3 and to 185 µg/m^3 for the LUR and the Gazel-Air exposure assessments, respectively. IQR= 216 µg/m^3 for the LUR, 127 µg/m^3 for Gazel-Air.**: the HR and CI were calculated for the 80th percentile (380 µg/m^3 for the LUR, 215 µg/m^3 for Gazel-Air).

Table S1: associations between all-site incident cancer and all co-variables in the Gazel cohort as Hazard Ratio (HR) and 95% Confidence Intervals (95%CI).

Variable	HR	95% CI
Smoking (for 5 cumulative pack-years)	1.04	1.03-1.05
Passive smoking - no	1 (reference)	
Yes	1.01	0.94-1.09
Alcohol use - light drinker	1 (reference)	
Abstinent	1.04	0.91-1.20
Moderate drinker	1.05	0.96-1.15
Heavy drinker	1.23	1.11-1.37
Unclear pattern	1.02	0.90-1.15
Socio-economic status - low	1 (reference)	
Intermediate	0.93	0.85-1.02
High	0.95	0.84-1.08
Family status - single	1 (reference)	
Not single	0.99	0.89-1.11
Body Mass Index (for 1 unit)	1.01	1.00-1.02
Vegetable & fruit consumption - every day or almost	1 (reference)	
Never or less than once a week	1.72	1.18-2.52
Once or twice a week	1.11	0.95-1.31
More than twice a week, not everyday	1.09	0.99-1.20
Education - 9-11 years	1 (reference)	
12-13 years	0.90	0.78-1.03
14-15 years	1.00	0.86-1.16
Other secondary diploma	0.93	0.81-1.06
Other diploma	0.92	0.74-1.15
Occupational exposure* - none	1 (reference)	
One	0.95	0.84-1.07
Two	1.00	0.89-1.11
Three or more	0.97	0.88-1.06
Age at inclusion (for one year)	1.02	1.00-1.04
Calendar year - before 2007	1 (reference)	
After 2007	1.22	1.07-1.39

*: to nine selected lung carcinogens.
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). Results obtained with 3,711 cases of all-site cancer and 293,188 person-years.