Online Data Supplement

Long-term Fine Particulate Matter Exposure on Lung Cancer Incidence and Mortality in Chinese Nonsmokers

Meng Zhu, PhD[1,2,3†], Yuting Han, PhD[4†], Yuanlin Mou, MD[1†,] Xia Meng, PhD[5†], Chen Ji, MD[1], Xia Zhu, MD[1], Canqing Yu, PhD[4,6,7], Dianjianyi Sun, PhD[4,6,7], Ling Yang, PhD[8,9], Qiufen Sun, PhD[1], Yiping Chen, DPhil[8,9], Huaidong Du, PhD[8,9], Juncheng Dai, PhD[1,2], Zhengming Chen, PhD[9], Zhibin Hu, PhD[1,2], Jun Lv, PhD[4,6,7,10], Guangfu Jin, PhD[1,2,3], Hongxia Ma, PhD[1,2,3], Haidong Kan, PhD[5*], Liming Li, MPH[4,6,7*], Hongbing Shen, PhD[1,2*], on behalf of the China Kadoorie Biobank Collaborative Group

Contents

Supplemental methods
- Modelling strategy of the main analyses
- Definitions and categorizations of covariates
- Polygenic risk score (PRS)
- Population attributable fraction (PAF)

References

Table E1. Summary results of 19 SNPs used for polygenic risk score in the study of Dai et al. (2019)

Table E2. Sensitivity analyses for associations of LCINS and long-term $PM_{2.5}$ exposure, per 10-$\mu g/m^3$

Table E3. Sensitivity analyses for the association of LCINS: excluding incident cases occurred in the first year of follow-up, per interquartile range (IQR) and per 10-$\mu g/m^3$

Table E4. Sensitivity analyses for the association of LCINS: excluding incident cases occurred in the first 3 years of follow-up, per interquartile range (IQR) and per 10-μg/m^3

Table E5. Associations of chronic obstructive pulmonary disease (COPD) and $PM_{2.5}$ exposure, per interquartile range (IQR) and per 10-μg/m^3

Table E6. Associations of LCINS and chronic obstructive pulmonary disease (COPD)

Figure E1. Distribution and trend of average $PM_{2.5}$ concentration and annual trends from 2005 to 2018 across the 10 regions

Figure E2. Diagram of the hypothesized mediation analysis and directed acyclic graphs of the associations of $PM_{2.5}$ and chronic obstructive pulmonary disease (COPD) with LCINS

Figure E3. Subgroup analyses of associations between long-term $PM_{2.5}$ exposure and LCINS according to baseline characteristics

Figure E4. The 5-year absolute risk of participants in the CKB between 2013-2018 and 2004-2013

Figure E5. The relationship between exposure to $PM_{2.5}$ and LCINS, mediated by chronic obstructive pulmonary disease (COPD)

Figure E6. Distribution of polygenic risk score between lung cancer cases or deaths and controls and association of PRS with LCINS

Figure E7. Risk of LCINS and related death according to air pollution (tertiles) and genetic categories

Supplementary Methods

Modelling strategy of the main analyses

In the baseline survey, potential sociodemographic and lifestyle confounders were collected and assessed using the Cox regression model. Confounders that improved the model were retained based on the likelihood ratio test, comparing models with and without each confounder (1). The final model for LCINS incidence risk included the most statistically significant covariates: age, sex, BMI, education, household income, living areas, weekly red meat consumption, weekly fresh fruit intake, second-hand smoke exposure hours per week, fuel types used in cooking at the baseline residence, years exposed to solid fuels, and self-rated health degree. In contrast, physical activities (MET), frequency of preserved vegetable consumption, years of living with smokers, fuel types used in heating at the baseline residence, and family cancer history did not significantly alter the observed associations and were excluded from the final model. Similarly, for the LCINS mortality regression model, we included age, sex, household income, physical activities (MET), weekly red meat consumption, fuel types used in cooking at the baseline residence, fuel types used in heating at the baseline residence, and BMI as covariates.

Definitions and Categorizations of Covariates

Sociodemographic characteristics included age, sex, education, household income, and living areas. Education was classified into four categories: no formal school, primary school, middle school, and high school or above. Household income was categorized into annual income of <10,000, 10,000-19,999, or ≥20,000 Chinese renminbi per year. Areas were classified as urban or rural. Lifestyle factors included the consumption of red meat and fresh fruits, with days per week of consumption assigned according to the midpoint of each frequency category: daily, 4-6 days per week, 1-3 days per week, monthly, or never/rarely.

For second-hand smoke exposure, we asked participants about hours of exposure per week. Participants also reported years of living in their current residence and up to two previous residences. For each residence, we inquired about the frequency of cooking and whether they

heated their house in winter. Participants who cooked monthly or heated their house in winter were asked about the main cooking (e.g., gas, coal, wood, electricity, or other) or heating fuels (i.e., central heating, gas, coal, wood, electricity, or other). Solid fuels included coal and wood, while clean fuels included gas, electricity, and central heating. We categorized fuel types for cooking and heating at the baseline residence as clean fuels, solid fuels, or other fuels, or no monthly cooking or heating in winter. Years of solid fuel use at the past three residences were summed up.

Self-rated health was classified as excellent, good, fair, or poor. A family history of cancer refers to at least one parent or sibling having had any type of cancer. The full electronic questionnaire is available online at: https://www.ckbiobank.org/site/binaries/content/assets/resources/pdf/qs_baseline-final-from10june2004.pdf.

Polygenic risk score (PRS)

The development and validation of PRS-19 have been detailed elsewhere (2). Briefly, using the largest available GWAS datasets of lung cancer in the Chinese population (13,327 cases and 13,328 controls), we evaluated all previously reported SNPs (81 SNPs in 51 loci) associated with lung cancer in the GWAS catalog. The PRS was calculated using an additive genetic model, applying several criteria to filter redundant variants: 1) minor allele frequency less than 0.01% in the Chinese population; 2) SNPs in linkage disequilibrium (LD, $r^2 \geq 0.1$); and 3) *P*-value ≥ 0.00125 (after multiple comparisons at 0.05/40, 40 independent loci remained based on the above two criteria) in Chinese lung cancer GWAS. Ultimately, 19 eligible SNPs were selected to construct PRS-19 (Table E1). The classic PRS method, described in prior studies (3) , was used to generate the lung cancer PRS for Chinese populations as follows:

$$PRS = \sum_{i=1}^{m} \beta_i \times SNP_i$$

Where β_i was per allele natural logarithm odds ratio for lung cancer related SNPs and the ORs were taken from the study of Dai et al (2). SNP_i was recoded as 0, 1, or 2 in terms of the number of risk alleles, and m was 19, namely the total SNPs number in this study. All participants in the CKB were not included in the Chinese lung cancer GWAS.

Population attributable fraction (PAF)

The population attributable fraction (PAF) represents the proportion of incident cases in this population during follow-up can be prevented if the exposure of a risk factor reduced to an ideal level, assuming a causal relationship (4). For binary factors, the PAF was defined as:

$$PAF = \frac{P \times (HR - 1)}{P \times (HR - 1) + 1}$$

where *P* is prevalence and *HR* is the hazard ratio for the risk factor.

As 94.5% of individuals in this study were exposed to $PM_{2.5}$ concentrations higher than the interim target guideline (35 μg/m^3) from China, we used the HR per 10 μg/m^3 $PM_{2.5}$ increase to calculate the PAF for every $PM_{2.5}$ concentration level. The equation was:

$$PAF = \frac{HR^{(c-t)/10} - 1}{HR^{(c-t)/10}}$$

where *HR* is the hazard ratio for the risk factor, *c* is the $PM_{2.5}$ concentrations (μg/m^3) and *t* is the interim target guideline of $PM_{2.5}$ (35 μg/m^3 in China and 10 μg/m^3 in Europe).

References

1. Huang Y, Zhu M, Ji M, Fan J, Xie J, Wei X, Jiang X, Xu J, Chen L, Yin R, Wang Y, Dai J, Jin G, Xu L, Hu Z, Ma H, Shen H. Air Pollution, Genetic Factors, and the Risk of Lung Cancer: A Prospective Study in the UK Biobank. *Am J Respir Crit Care Med* 2021; 204: 817-825.
2. Dai J, Lv J, Zhu M, Wang Y, Qin N, Ma H, He YQ, Zhang R, Tan W, Fan J, Wang T, Zheng H, Sun Q, Wang L, Huang M, Ge Z, Yu C, Guo Y, Wang TM, Wang J, Xu L, Wu W, Chen L, Bian Z, Walters R, Millwood IY, Li XZ, Wang X, Hung RJ, Christiani DC, Chen H, Wang M, Wang C, Jiang Y, Chen K, Chen Z, Jin G, Wu T, Lin D, Hu Z, Amos CI, Wu C, Wei Q, Jia WH, Li L, Shen H. Identification of risk loci and a polygenic risk score for lung cancer: a large-scale prospective cohort study in Chinese populations. *Lancet Respir Med* 2019; 7: 881-891.
3. Choi SW, Mak TS, O'Reilly PF. Tutorial: a guide to performing polygenic risk score analyses. *Nat Protoc* 2020; 15: 2759-2772.
4. Zhu N, Yu C, Guo Y, Bian Z, Han Y, Yang L, Chen Y, Du H, Li H, Liu F, Chen J, Chen Z, Lv J, Li L. Adherence to a healthy lifestyle and all-cause and cause-specific mortality in Chinese adults: a 10-year prospective study of 0.5 million people. *Int J Behav Nutr Phys Act* 2019; 16: 98.

Table E1. Summary results of 19 SNPs used for polygenic risk score in the study of Dai et al. (2019)

No.	Chr.	SNP ID	Position (bp)	Allele A	Allele B	EAF	OR (95%CI)	Beta
1	2	rs3769821	202123430	C	T	0.321	1.08 (1.05-1.11)	0.077
2	2	rs17038564	65496058	G	A	0.161	1.15 (1.10-1.21)	0.14
3	3	rs2293607	169482335	T	C	0.596	1.10 (1.06-1.13)	0.095
4	3	rs11375254	189343242	T	TA	0.525	1.13 (1.10-1.16)	0.122
5	5	rs13167280	1280477	A	G	0.17	1.24 (1.19-1.28)	0.215
6	6	rs1853837	41497035	A	C	0.258	1.12 (1.08-1.15)	0.113
7	6	rs2517873	29875992	A	G	0.145	1.16 (1.11-1.21)	0.148
8	6	rs5879422	117784658	T	TTG	0.51	1.08 (1.05-1.11)	0.077
9	8	rs4236709	32410110	G	A	0.198	1.12 (1.09-1.16)	0.113
10	9	rs35201538	33422488	C	CT	0.676	1.10 (1.06-1.13)	0.095
11	9	rs4573350	124955115	T	C	0.505	1.13 (1.09-1.18)	0.122
12	9	rs10429489	21787521	A	G	0.195	1.11 (1.08-1.15)	0.104
13	10	rs12265047	114487925	G	A	0.243	1.13 (1.09-1.17)	0.122
14	11	rs55768116	118108331	C	A	0.489	1.10 (1.07-1.13)	0.095
15	14	rs1200399	35293185	C	T	0.405	1.11 (1.07-1.15)	0.104
16	15	rs77468143	49376624	T	G	0.767	1.14 (1.10-1.18)	0.131
17	15	rs55781567	78857986	G	C	0.32	1.27 (1.23-1.31)	0.239
18	17	rs200595745	65915289	A	AAATA ATAAT	0.43	1.09 (1.06-1.12)	0.086
19	19	rs56113850	41353107	C	T	0.487	1.09 (1.06-1.12)	0.086

Definition of abbreviations: SNP, single nucleotide polymorphism; EAF, effect allele frequency; OR, odds ratio; CI, confidence interval.

Table E2. Sensitivity analyses for associations of LCINS and long-term $PM_{2.5}$ exposure, per 10-μg/m^3

Sensitivity analyses	No. of Events/ Total No.	Incidence/ 100,000 person-yr	HR (95% CI)
Lung cancer incidence*			
Using 3-year average $PM_{2.5}$ as time-varying variable	2571/255829	84.53	1.15 (1.08, 1.23)
Using 5-year average $PM_{2.5}$ as time-varying variable	2571/255829	84.53	1.19 (1.12, 1.27)
Using average $PM_{2.5}$ from 2005 to 2018 as time-unvarying variable	2571/255829	84.53	1.36 (1.25, 1.48)
Including participants in Henan and Haikou	2961/315819	79.07	1.07 (1.05, 1.11)
Lung cancer mortality†			
Using 3-year average $PM_{2.5}$ as time-varying variable	1416/255829	46.48	1.19 (1.10, 1.30)
Using 5-year average $PM_{2.5}$ as time-varying variable	1416/255829	46.48	1.21 (1.11, 1.32)
Using average $PM_{2.5}$ from 2005 to 2018 as time-unvarying variable	1416/255829	46.48	1.31 (1.18, 1.47)
Including participants in Henan and Haikou	1628/315819	43.41	1.08 (1.03, 1.12)

Definition of abbreviations: HR, hazard ratio; CI, confidence interval; $PM_{2.5}$, fine (≤2.5 μm) particulate matter.

* Adjusted for age, sex, BMI, education, household income, living areas, weekly red meat consumption, weekly fresh fruit intake, second-hand smoke exposure hours per week, fuel types used in cooking at the baseline residence, years exposed to solid fuels, and self-rated health degree.

† Adjusted for age, sex, household income, physical activities (MET), weekly red meat consumption, fuel types used in cooking at the baseline residence, fuel types used in heating at the baseline residence, and BMI.

Table E3. Sensitivity analyses for the association of LCINS: excluding incident cases occurred in the first year of follow-up, per interquartile range (IQR) and per 10-μg/m^3

	$PM_{2.5}$ concentration				
	Quantile 1 (30.2-42.3 μg/m^3)	**Quantile 2 (42.3-50.3 μg/m^3)**	**Quantile 3 (504-54.1 μg/m^3)**	**Quantile 4 (54.2-63.1 μg/m^3)**	**$PM_{2.5}$ per 10 μg/m^3 increased**
Lung cancer incidence					
No. of Events/Total No.	129/31230	564/60725	881/89046	894/74725	2468/255726
Incidence/100,000 person-yr	35.06	79.31	83.21	98.93	81.14
HR (95% CI) *	1.00 (Ref)	1.48 (1.19, 1.84)	1.79 (1.46, 2.19)	1.90 (1.54, 2.35)	1.09 (1.03, 1.16)
Lung cancer mortality					
No. of Events/Total No.	87/31231	336/60741	483/89076	479/74750	1385/255798
Incidence/100,000 person-yr	23.63	47.18	45.54	52.91	45.46
HR (95% CI) †	1.00 (Ref)	1.59 (1.21, 2.08)	1.75 (1.35, 2.27)	1.93 (1.47, 2.52)	1.17 (1.08, 1.26)

Definition of abbreviations: HR, hazard ratio; CI, confidence interval; $PM_{2.5}$, fine (≤2.5 μm) particulate matter.

* Adjusted for age, sex, BMI, education, household income, living areas, weekly red meat consumption, weekly fresh fruit intake, second-hand smoke exposure hours per week, fuel types used in cooking at the baseline residence, years exposed to solid fuels, and self-rated health degree.

† Adjusted for age, sex, household income, physical activities (MET), weekly red meat consumption, fuel types used in cooking at the baseline residence, fuel types used in heating at the baseline residence, and BMI.

Table E4. Sensitivity analyses for the association of LCINS: excluding incident cases occurred in the first 3 years of follow-up, per interquartile range (IQR) and per 10-μg/m^3

	$PM_{2.5}$ concentration				
	Quantile 1 (30.2-42.3 μg/m^3)	**Quantile 2 (42.3-50.3 μg/m^3)**	**Quantile 3 (504-54.1 μg/m^3)**	**Quantile 4 (54.2-63.1 μg/m^3)**	**$PM_{2.5}$ per 10 μg/m^3 increased**
Lung cancer incidence					
No. of Events/Total No.	104/31205	494/60655	766/88931	798/74629	2162/255420
Incidence/100,000 person-yr	28.27	69.48	72.36	88.32	71.10
HR (95% CI) *	1.00 (Ref)	1.65 (1.30, 2.09)	1.93 (1.54, 2.41)	2.11 (1.67, 2.67)	1.14 (1.06, 1.21)
Lung cancer mortality					
No. of Events/Total No.	73/31217	288/60693	421/89014	429/74700	1211/255624
Incidence/100,000 person-yr	19.83	40.44	39.71	47.38	39.76
HR (95% CI) †	1.00 (Ref)	1.70 (1.27, 2.28)	1.88 (1.41, 2.49)	2.13 (1.59, 2.84)	1.23 (1.13, 1.34)

Definition of abbreviations: HR, hazard ratio; CI, confidence interval; $PM_{2.5}$, fine (≤2.5 μm) particulate matter.

* Adjusted for age, sex, BMI, education, household income, living areas, weekly red meat consumption, weekly fresh fruit intake, second-hand smoke exposure hours per week, fuel types used in cooking at the baseline residence, years exposed to solid fuels, and self-rated health degree.

† Adjusted for age, sex, household income, physical activities (MET), weekly red meat consumption, fuel types used in cooking at the baseline residence, fuel types used in heating at the baseline residence, and BMI.

Table E5. Associations of chronic obstructive pulmonary disease (COPD) and $PM_{2.5}$ exposure, per interquartile range (IQR) and per 10-μg/m^3

Groups	$PM_{2.5}$ interval	No. of Events/Total No.	Incidence/1000 persons	OR (95% CI)
$PM_{2.5}$ Q1	30.2-42.3	2006/31235	64.22	Reference
$PM_{2.5}$ Q2	42.3-50.3	5766/60748	94.91	2.42 (2.27-2.57)
$PM_{2.5}$ Q3	50.4-54.1	6546/89086	73.47	1.77 (1.67-1.88)
$PM_{2.5}$ Q4	54.2-63.1	6763/74760	90.46	3.12 (2.92-3.32)
$PM_{2.5}$ 10 μg/m^3 increased	30.2-63.1	21081/255829	82.40	1.61 (1.56-1.66)

Definition of abbreviations: OR, odds ratio; CI, confidence interval; $PM_{2.5}$, fine (≤2.5 μm) particulate matter.

Adjusted for age, sex, BMI, education, household income, living areas, weekly red meat consumption, weekly fresh fruit intake, second-hand smoke exposure hours per week, fuel types used in cooking at the baseline residence, years exposed to solid fuels, and self-rated health degree.

Table E6. Associations of LCINS and chronic obstructive pulmonary disease (COPD).

Prevalent of COPD	No. of Events/Total No.	Incidence/100,000 person-yr	HR (95% CI)	*P* value
LCINS incidence *				
No	2280/234748	81.60	Reference	
Yes	291/21081	117.58	1.24 (1.09-1.40)	5.97×10^{-4}
LCINS mortality †				
No	1252/234748	44.81	Reference	
Yes	164/21081	66.26	1.22 (1.03-1.44)	0.017

Definition of abbreviations: HR, hazard ratio; CI, confidence interval; COPD, chronic obstructive pulmonary disease.

* Adjusted for age, sex, BMI, education, household income, living areas, weekly red meat consumption, weekly fresh fruit intake, second-hand smoke exposure hours per week, fuel types used in cooking at the baseline residence, years exposed to solid fuels, and self-rated health degree.

† Adjusted for age, sex, household income, physical activities (MET), weekly red meat consumption, fuel types used in cooking at the baseline residence, fuel types used in heating at the baseline residence, and BMI.

Figure Legends

Figure E1. Distribution and trend of average $PM_{2.5}$ concentration and annual trends from 2005 to 2018 across the 10 regions

Definition of abbreviations: $PM_{2.5}$, fine (≤2.5 μm) particulate matter.

Figure E2. Diagram of the hypothesized mediation analysis and directed acyclic graphs of the associations of $PM_{2.5}$ and chronic obstructive pulmonary disease (COPD) with LCINS

Definition of abbreviations: $PM_{2.5}$, fine (≤2.5 μm) particulate matter; COPD, chronic obstructive pulmonary disease; BMI, body mass index; MET, metabolic equivalent of task.

The diagram illustrates the hypothesized mediation analysis (A) and directed acyclic graphs showing the associations of $PM_{2.5}$ with LCINS (B), $PM_{2.5}$ with COPD (C), and COPD with LCINS (D).

Figure E3. Subgroup analysis of associations between long-term $PM_{2.5}$ exposure and LCINS according to baseline characteristics

Definition of abbreviations: HR, hazard ratio; CI, confidence interval; $PM_{2.5}$, fine (≤2.5 μm) particulate matter; BMI, body mass index.

The outcomes are (A) LCINS incidence and (B) LCINS-related mortality.

* Adjusted for age, sex, BMI, education, household income, living areas, weekly red meat consumption, weekly fresh fruit intake, second-hand smoke exposure hours per week, fuel types used in cooking at the baseline residence, years exposed to solid fuels, and self-rated health degree.

† Adjusted for age, sex, household income, physical activities (MET), weekly red meat consumption, fuel types used in cooking at the baseline residence, fuel types used in heating at the baseline residence, and BMI.

‡ *P* values for heterogeneity between subgroups were obtained from Cochrane's Q test.

Figure E4. The 5-year absolute risk of participants in the CKB between 2013-2018 and 2004-2013

* The 5-year absolute risk was calculated as the percentage of incident cancer cases occurring

in a given group over a 5-year period. Participants were standardized to 60 years old, and the *P* value was derived from chi-square tests.

Figure E5. The relationship between exposure to $PM_{2.5}$ and LCINS, mediated by chronic obstructive pulmonary disease (COPD)

Definition of abbreviations: $PM_{2.5}$, fine (≤2.5 μm) particulate matter.

The outcomes are (A) LCINS incidence and (B) LCINS-related mortality.

* Adjusted for age, sex, BMI, education, household income, living areas, weekly red meat consumption, weekly fresh fruit intake, second-hand smoke exposure hours per week, fuel types used in cooking at the baseline residence, years exposed to solid fuels, and self-rated health degree.

† Adjusted for age, sex, household income, physical activities (MET), weekly red meat consumption, fuel types used in cooking at the baseline residence, fuel types used in heating at the baseline residence, and BMI.

Figure E6. Distribution of polygenic risk score between lung cancer cases or deaths and controls and association of PRS with LCINS

Definition of abbreviations: CI, confidence interval; PRS, polygenic risk score.

The outcomes are (A/B) LCINS incidence and (C/D) LCINS-related mortality.

* Adjusted for age, sex, BMI, education, household income, living areas, weekly red meat consumption, weekly fresh fruit intake, second-hand smoke exposure hours per week, fuel types used in cooking at the baseline residence, years exposed to solid fuels, and self-rated health degree.

† Adjusted for age, sex, household income, physical activities (MET), weekly red meat consumption, fuel types used in cooking at the baseline residence, fuel types used in heating at the baseline residence, and BMI.

Figure E7. Risk of LCINS and related death according to air pollution (tertiles) and genetic categories

Definition of abbreviations: $PM_{2.5}$, fine (≤2.5 μm) particulate matter; CI, confidence interval;

RERI: relative excess risk due to the interaction; AP, attributable proportion due to the interaction.

The outcomes are (A) LCINS incidence and (B) LCINS-related mortality.

* Estimate: Adjusted for age, sex, BMI, education, household income, living areas, weekly red meat consumption, weekly fresh fruit intake, second-hand smoke exposure hours per week, fuel types used in cooking at the baseline residence, years exposed to solid fuels, and self-rated health degree.

† Adjusted for age, sex, household income, physical activities (MET), weekly red meat consumption, fuel types used in cooking at the baseline residence, fuel types used in heating at the baseline residence, and BMI.

‡ Defined by PRS: low (bottom 30%), intermediate (30%-70%), and high (top 30%).

§ Defined by the distribution of $PM_{2.5}$ concentrations: low (<50.2 $\mu g/m^3$), middle (50.2~54.1 $\mu g/m^3$), and high (≥54.1 $\mu g/m^3$).

‖ To estimate RERI and AP, the lowest pollution category (low $PM_{2.5}$) and the lowest genetic risk (low PRS) groups were the reference categories.

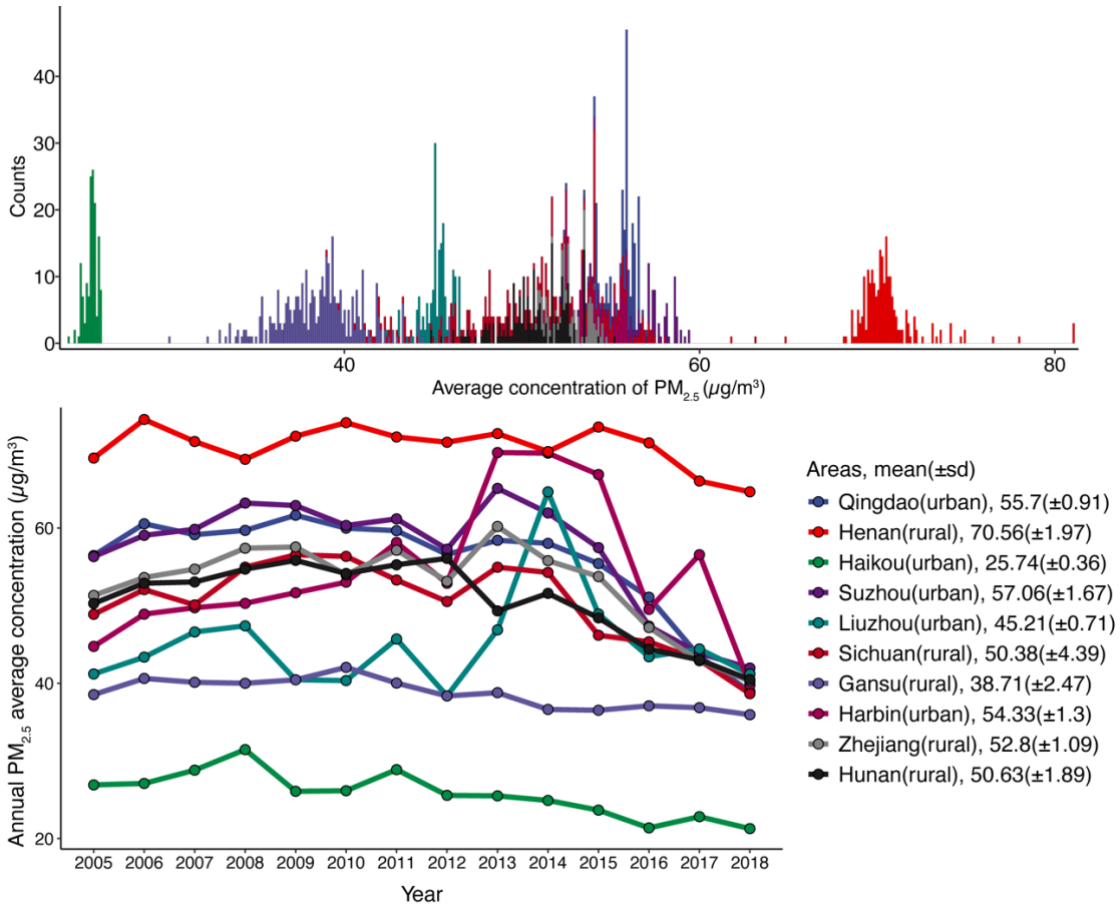

Counts
40
30
20
10
0
40
60
80
Average concentration of PM2.5 (μg/m3)
Annual PM2.5 average concentration (μg/m3)
60
40
20
2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018
Year
Areas, mean(±sd)
Qingdao(urban), 55.7(±0.91)
Henan(rural), 70.56(±1.97)
Haikou(urban), 25.74(±0.36)
Suzhou(urban), 57.06(±1.67)
Liuzhou(urban), 45.21(±0.71)
Sichuan(rural), 50.38(±4.39)
Gansu(rural), 38.71(±2.47)
Harbin(urban), 54.33(±1.3)
Zhejiang(rural), 52.8(±1.09)
Hunan(rural), 50.63(±1.89)

Figure E1

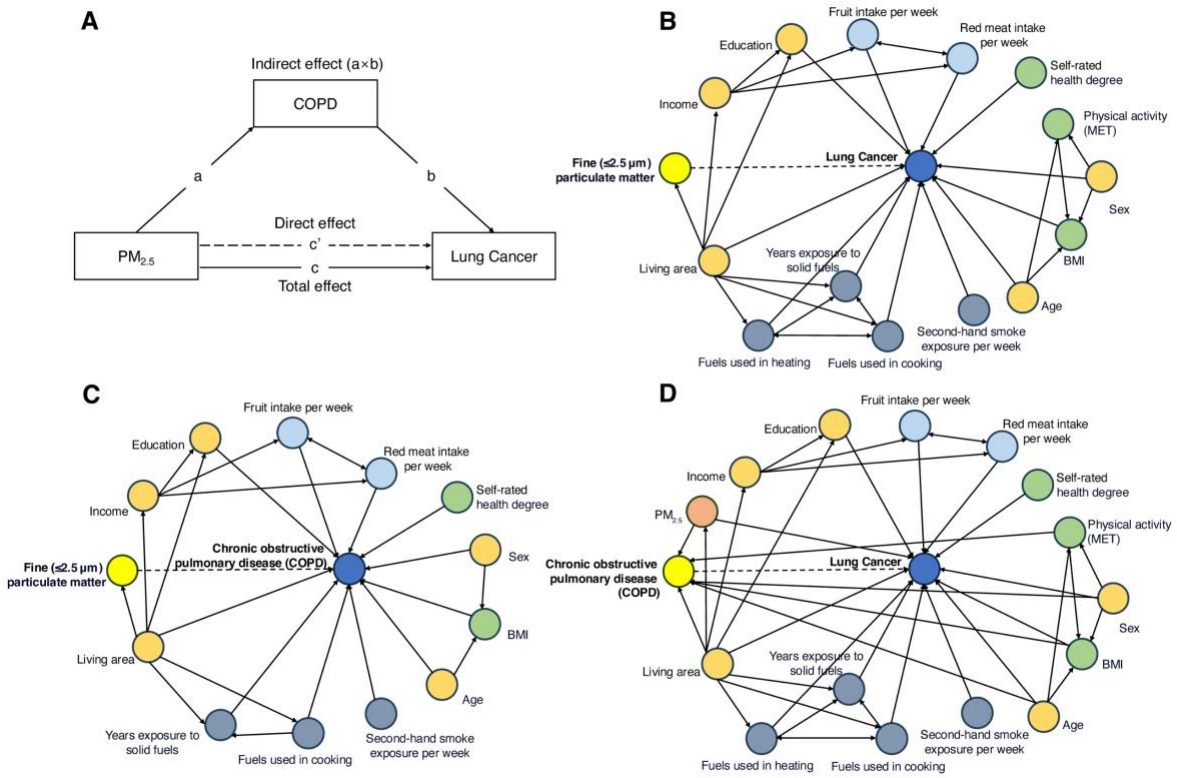

A
Indirect effect (a×b)
COPD
a
b
Direct effect
c'
PM2.5
Lung Cancer
c
Total effect
B
Fruit intake per week
Education
Red meat intake per week
Self-rated health degree
Income
Physical activity (MET)
Fine (≤2.5 μm) particulate matter
Lung Cancer
Sex
Years exposure to solid fuels
Living area
BMI
Age
Second-hand smoke exposure per week
Fuels used in heating
Fuels used in cooking
C
Fruit intake per week
Education
Red meat intake per week
Self-rated health degree
Income
Chronic obstructive pulmonary disease (COPD)
Fine (≤2.5 μm) particulate matter
Sex
BMI
Living area
Age
Years exposure to solid fuels
Fuels used in cooking
Second-hand smoke exposure per week
D
Fruit intake per week
Education
Red meat intake per week
Income
Self-rated health degree
PM2.5
Physical activity (MET)
Chronic obstructive pulmonary disease (COPD)
Lung Cancer
Sex
Living area
Years exposure to solid fuels
BMI
Age
Second-hand smoke exposure per week
Fuels used in heating
Fuels used in cooking

Figure E2

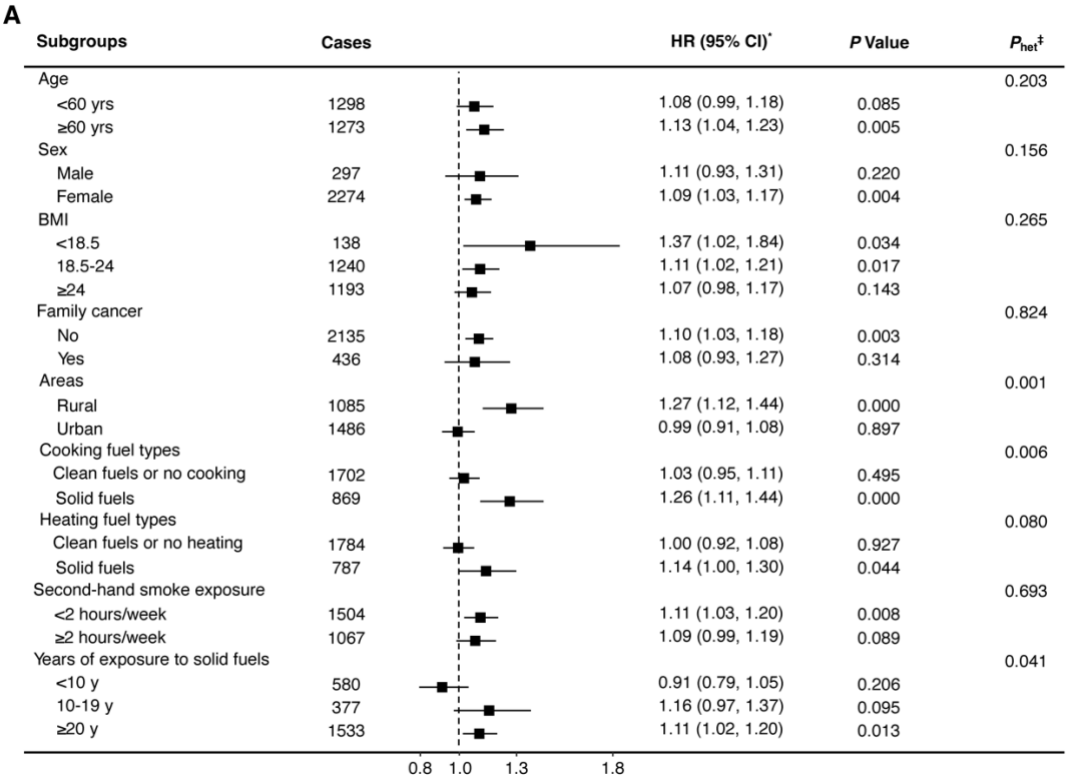

A
Subgroups
Cases
HR (95% CI)*
P Value
Phet‡
Age
0.203
<60 yrs
1298
1.08 (0.99, 1.18)
0.085
≥60 yrs
1273
1.13 (1.04, 1.23)
0.005
Sex
0.156
Male
297
1.11 (0.93, 1.31)
0.220
Female
2274
1.09 (1.03, 1.17)
0.004
BMI
0.265
<18.5
138
1.37 (1.02, 1.84)
0.034
18.5-24
1240
1.11 (1.02, 1.21)
0.017
≥24
1193
1.07 (0.98, 1.17)
0.143
Family cancer
0.824
No
2135
1.10 (1.03, 1.18)
0.003
Yes
436
1.08 (0.93, 1.27)
0.314
Areas
0.001
Rural
1085
1.27 (1.12, 1.44)
0.000
Urban
1486
0.99 (0.91, 1.08)
0.897
Cooking fuel types
0.006
Clean fuels or no cooking
1702
1.03 (0.95, 1.11)
0.495
Solid fuels
869
1.26 (1.11, 1.44)
0.000
Heating fuel types
0.080
Clean fuels or no heating
1784
1.00 (0.92, 1.08)
0.927
Solid fuels
787
1.14 (1.00, 1.30)
0.044
Second-hand smoke exposure
0.693
<2 hours/week
1504
1.11 (1.03, 1.20)
0.008
≥2 hours/week
1067
1.09 (0.99, 1.19)
0.089
Years of exposure to solid fuels
0.041
<10 y
580
0.91 (0.79, 1.05)
0.206
10-19 y
377
1.16 (0.97, 1.37)
0.095
≥20 y
1533
1.11 (1.02, 1.20)
0.013
0.8
1.0
1.3
1.8
B
Subgroups
Cases
HR (95% CI)†
P Value
Phet‡
Age
0.759
<60 yrs
594
1.16 (1.02, 1.31)
0.023
≥60 yrs
822
1.19 (1.07, 1.32)
0.001
Sex
0.237
Male
190
1.18 (0.96, 1.45)
0.090
Female
1226
1.16 (1.06, 1.27)
0.001
BMI
0.272
<18.5
88
1.38 (0.97, 1.95)
0.073
18.5-24
690
1.22 (1.08, 1.36)
0.001
≥24
638
1.09 (0.97, 1.23)
0.140
Family cancer
0.458
No
1195
1.16 (1.06, 1.26)
0.001
Yes
221
1.26 (1.02, 1.55)
0.031
Areas
0.242
Rural
623
1.23 (1.04, 1.45)
0.016
Urban
793
1.09 (0.97, 1.22)
0.150
Cooking fuel types
0.372
Clean fuels or no cooking
882
1.13 (1.02, 1.25)
0.024
Solid fuels
534
1.23 (1.05, 1.45)
0.012
Heating fuel types
0.096
Clean fuels or no heating
960
1.09 (0.98, 1.21)
0.115
Solid fuels
456
1.29 (1.10, 1.51)
0.002
Second-hand smoke exposure
0.458
<2 hours/week
860
1.21 (1.09, 1.34)
0.000
≥2 hours/week
556
1.13 (1.00, 1.29)
0.058
Years of exposure to solid fuels
0.090
<10 y
281
0.95 (0.78, 1.16)
0.628
10-19 y
178
1.33 (1.02, 1.73)
0.035
≥20 y
913
1.17 (1.06, 1.30)
0.003
0.8
1.0
1.3
1.8

Figure E3

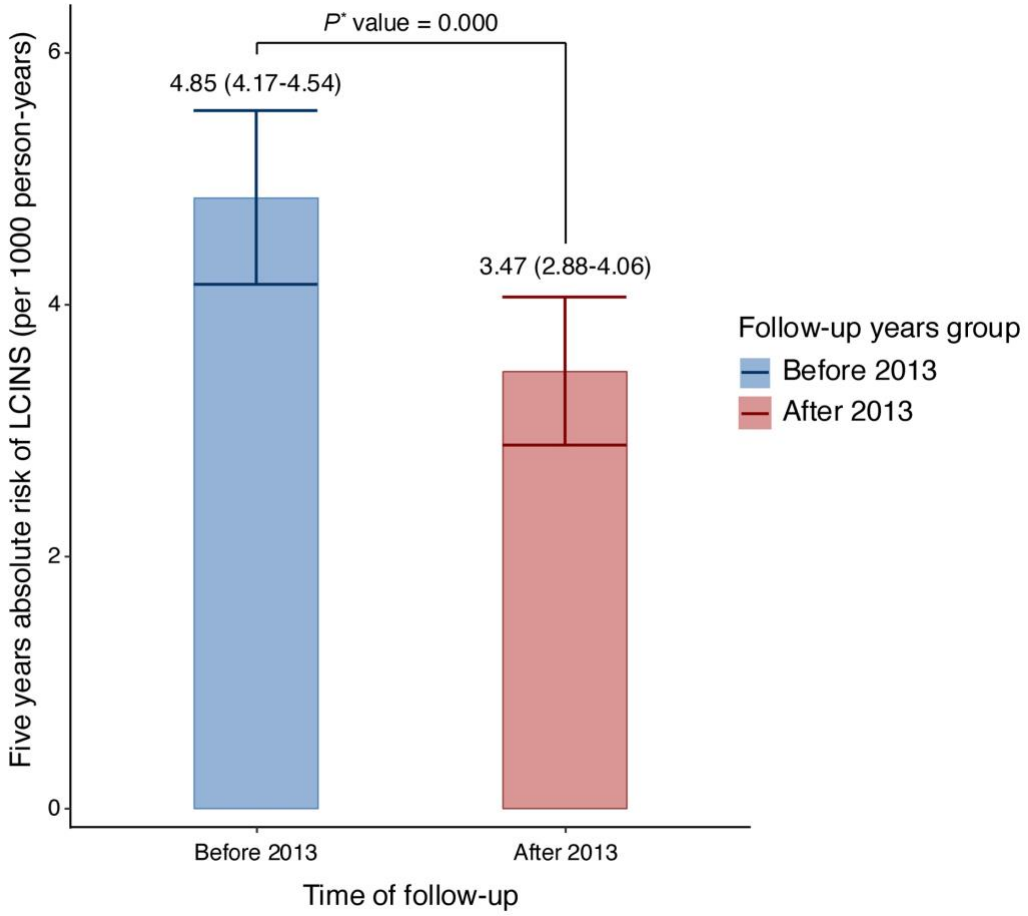

P* value = 0.000
4.85 (4.17-4.54)
3.47 (2.88-4.06)
Five years absolute risk of LCINS (per 1000 person-years)
6
4
2
0
Before 2013
After 2013
Time of follow-up
Follow-up years group
Before 2013
After 2013

Figure E4

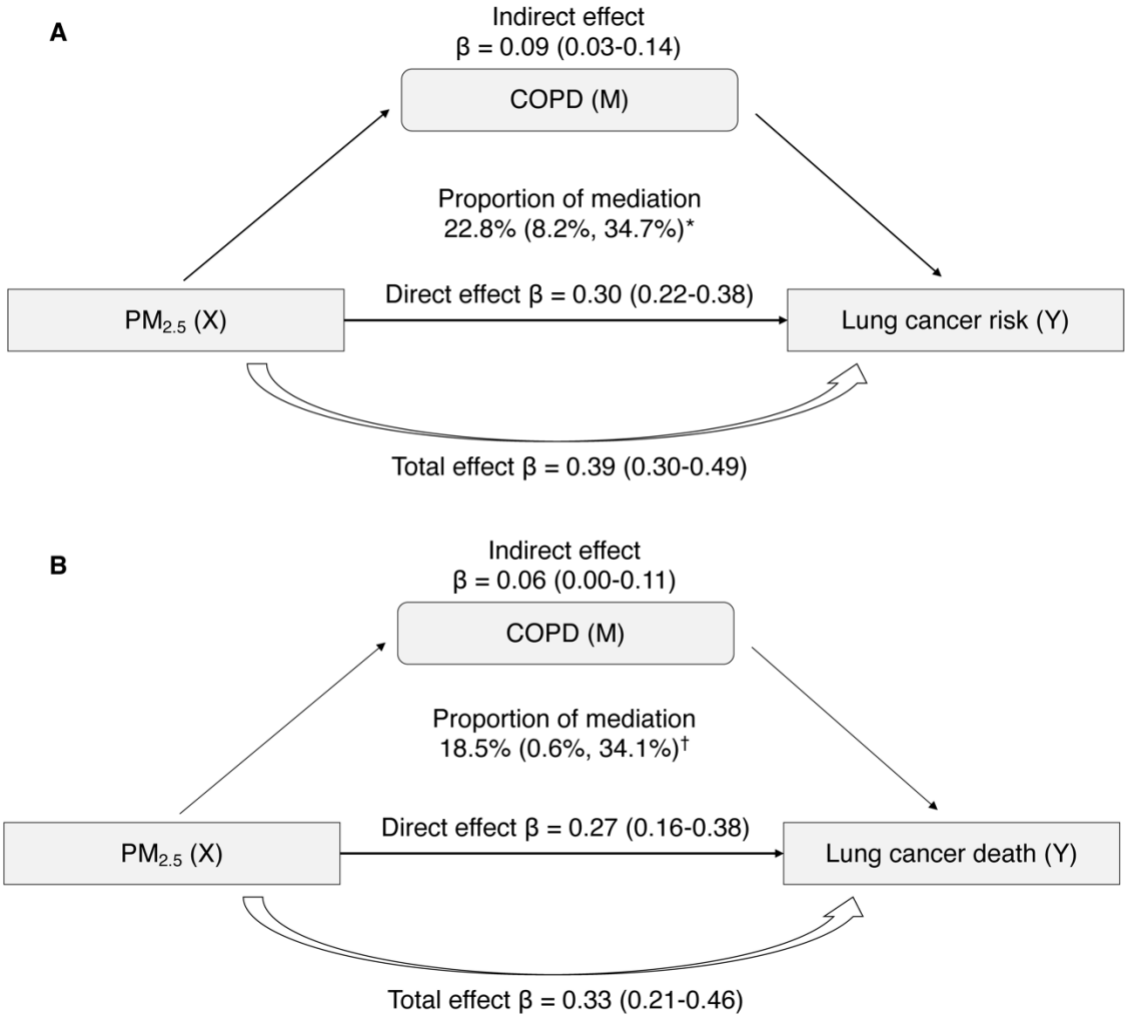

A
Indirect effect
β = 0.09 (0.03-0.14)
COPD (M)
Proportion of mediation
22.8% (8.2%, 34.7%)*
PM2.5 (X)
Direct effect β = 0.30 (0.22-0.38)
Lung cancer risk (Y)
Total effect β = 0.39 (0.30-0.49)
B
Indirect effect
β = 0.06 (0.00-0.11)
COPD (M)
Proportion of mediation
18.5% (0.6%, 34.1%)†
PM2.5 (X)
Direct effect β = 0.27 (0.16-0.38)
Lung cancer death (Y)
Total effect β = 0.33 (0.21-0.46)

Figure E5

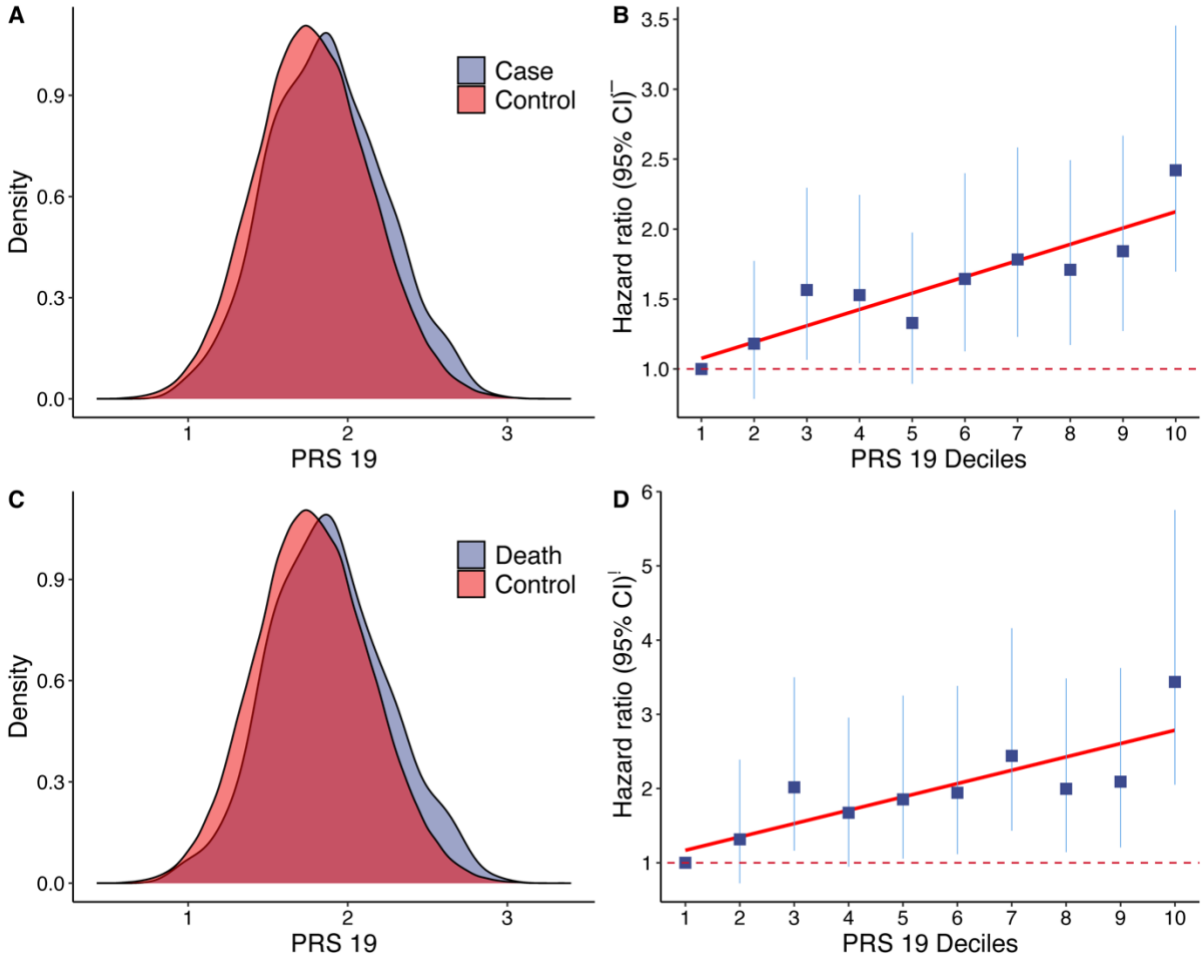

A
Case
Control
Density
PRS 19
B
Hazard ratio (95% CI)
PRS 19 Deciles
C
Death
Control
Density
PRS 19
D
Hazard ratio (95% CI)
PRS 19 Deciles

Figure E6

A

Risk group	No. of Events/ Total No.	Incidence/ 100,000 person-yr	Hazard Ratio (95% CI)*	*P* Value	RERI[II] (95% CI)	AP[II] (95% CI)
Low Genetic Risk‡						
Low $PM_{2.5}$ pollution§	36/4569	68.33	Reference		Reference	Reference
Middle $PM_{2.5}$ pollution	46/4913	80.25	1.09 (0.70-1.70)	0.697		
High $PM_{2.5}$ pollution	59/5394	90.02	1.03 (0.67-1.58)	0.886		
Intermediate Genetic Risk						
Low $PM_{2.5}$ pollution	60/6430	80.46	1.09 (0.72-1.64)	0.693		
Middle $PM_{2.5}$ pollution	77/6734	97.83	1.33 (0.89-1.99)	0.159	0.16 (-0.53-0.67)	0.12 (-0.36-0.53)
High $PM_{2.5}$ pollution	105/6671	129.56	1.50 (1.01-2.23)	0.043	0.38 (-0.22-0.86)	0.25 (-0.14-0.61)
High Genetic Risk						
Low $PM_{2.5}$ pollution	60/5534	93.16	1.22 (0.80-1.84)	0.356		
Middle $PM_{2.5}$ pollution	80/4932	139.33	1.94 (1.30-2.89)	0.001	0.63 (-0.02-1.25)	0.33 (-0.00-0.62)
High $PM_{2.5}$ pollution	86/4410	161.20	1.92 (1.28-2.88)	0.001	**0.67 (0.03-1.28)**	**0.35 (0.01-0.64)**

B

Risk group	No. of Events/ Total No.	Incidence/ 100,000 person-yr	Hazard Ratio (95% CI)†	*P* Value	RERI (95% CI)	AP (95% CI)
Low Genetic Risk						
Low $PM_{2.5}$ pollution	19/4569	36.02	Reference		Reference	Reference
Middle $PM_{2.5}$ pollution	16/4913	27.87	0.81 (0.41-1.58)	0.528		
High $PM_{2.5}$ pollution	35/5394	53.32	1.40 (0.78-2.51)	0.256		
Intermediate Genetic Risk						
Low $PM_{2.5}$ pollution	31/6430	41.50	1.09 (0.62-1.93)	0.768		
Middle $PM_{2.5}$ pollution	38/6734	48.18	1.39 (0.79-2.42)	0.250	0.49 (-0.47-1.21)	0.35 (-0.30-0.88)
High $PM_{2.5}$ pollution	59/6671	72.64	1.93 (1.12-3.31)	0.018	0.43 (-0.60-1.23)	0.22 (-0.27-0.63)
High Genetic Risk						
Low $PM_{2.5}$ pollution	32/5534	49.60	1.31 (0.74-2.32)	0.352		
Middle $PM_{2.5}$ pollution	41/4932	71.20	2.05 (1.18-3.55)	0.011	0.93 (-0.02-1.95)	0.45 (-0.01-0.82)
High $PM_{2.5}$ pollution	46/4410	85.96	2.37 (1.36-4.13)	0.002	0.65 (-0.43-1.70)	0.27 (-0.18-0.63)

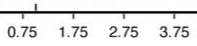


Figure E7