Supplementary Appendix

Associations between long-term air pollution and kidney diseases in the US Medicare population

Whanhee Lee[1], Xiao Wu[2], Seulkee Heo[1], Kelvin C. Fong[1], Ji-Young Son[1], M. Benjamin Sabath[3], Danielle Braun[2,4], Jae Yoon Park[5,6], Yong Chul Kim[7], Jung Pyo Lee[7,8], Joel Schwartz[9], Ho Kim[10,11], Francesca Dominici[2], Michelle L. Bell[1]

Correspondence to: Whanhee Lee, PhD. Yale School of the Environment, Yale University, (Address) 195 Prospect Street, New Haven, CT 06511, USA. Telephone: (1) 203 432-9869. Fax: (1) 203 466-9158. E-mail: whanhee.lee@yale.edu

[1]Yale School of the Environment, Yale University, New Haven, CT, USA.

[2]Department of Biostatistics, Harvard T H Chan School of Public Health, Boston, MA, USA.

[3]Faculty of Arts and Sciences Research Computing Department, Harvard University, Boston, MA, USA.

[4]Department of Data Science, Dana-Farber Cancer Institute, Boston, MA, USA.

[5]Department of Internal Medicine, Dongguk University Ilsan Hospital, Goyang, Republic of Korea.

[6]Department of Internal Medicine, Dongguk University College of Medicine, Goyang, Republic of Korea.

[7]Department of Internal Medicine, Seoul National University Hospital, Seoul, Seoul, Republic of Korea.

[8]Department of Internal Medicine, Seoul National University Boramae Medical Center, Seoul, Republic of Korea.

[9]Department of Environmental Health, Harvard T H Chan School of Public Health, Boston, MA, USA.

[10]Department of Public Health Science, Graduate School of Public Health, Seoul National University, Seoul, Republic of Korea.

[11]Institute for Sustainable Development, Graduate School of Public Health, Seoul National University, Seoul, Republic of Korea.

Lee W, *et al. BMJMED* 2022; 1:e000009. doi: 10.1136/bmjmed-2021-000009

A. STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	page 1 (Title)
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	page 3 (Abstract)
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	page 4 (Introduction)
Objectives	3	State specific objectives, including any prespecified hypotheses	page 4 (Introduction)
Methods			
Study design	4	Present key elements of study design early in the paper	page 4-5 (Methods – Study design and participants)
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	page 4-5 (Methods – Study design and participants)
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	page 4-5 (Methods – Study design and participants)
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case	NA.
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	page 4-5 (Methods)
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	page 4-6 (Methods)
Bias	9	Describe any efforts to address potential sources of bias	page 6 (Sensitivity analysis)
Study size	10	Explain how the study size was arrived at	page 4-5 (Methods – Study design and participants)
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were	page 4-6 (Methods)

		chosen and why	
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	page 5 (Methods – Statistical Analysis)
		(*b*) Describe any methods used to examine subgroups and interactions	page 6 (Methods – Subgroup analysis)
		(*c*) Explain how missing data were addressed	page 5 (Methods – Statistical Analysis)
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy	NA.
		(*e*) Describe any sensitivity analyses	page 6 (Methods – Sensitivity analysis)

Continued on next page

Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	page 6 (Results)
		(b) Give reasons for non-participation at each stage	NA.
		(c) Consider use of a flow diagram	-
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	page 6-7 (Results), Figure 1-2, Table 1
		(b) Indicate number of participants with missing data for each variable of interest	NA.
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	page 6 (Results), Table 2
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	page 6 (Results), Table 2, Figure 1
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	-
		Cross-sectional study—Report numbers of outcome events or summary measures	-
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	page 6-7 (Results), Table 2-3
		(*b*) Report category boundaries when continuous variables were categorized	page 5 (Statistical Analysis) Table 1, 3
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	-
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	page 6-7 (Results), Table 3, Table S3-4
Discussion			
Key results	18	Summarise key results with reference to study objectives	page 7 (Discussion)
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	page 7 (Discussion)
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	page 7 (Discussion)
Generalisability	21	Discuss the generalisability (external validity) of the study results	page 7 (Discussion)
Other information			

Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	page 12 (Funding)

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

B. Details on air pollution data

The predictions for $PM_{2.5}$ and NO_2 were estimated by hybrid ensemble models incorporating random forest, gradient boosting, and neural network. Multiple predictor variables including monitoring data, satellite data, meteorological conditions, land-use variables, and chemical transport model simulation values were included in these models. The technical details of the prediction models have been previously published with excellent performance.[1-3]

$PM_{2.5}$. We collected daily $PM_{2.5}$ predictions at a lkm^2 spatial resolution across the 48 contiguous US states and Washington DC from a well-validated ensemble model.[1] The model included more than 100 predictors from monitoring data, satellite-based aerosol optical depth (AOD) data, chemical transport model outputs, meteorological data, aerosol index, and land-use data. The model was calibrated with daily $PM_{2.5}$ concentrations measured at 2156 monitors operated by the US Environmental Protection Agency's Air Quality System database and IMPROVE monitoring network.[4] For AOD data, the algorithm called MAIAC that retrieves AOD with a spatial resolution of $1km^2$ from the Moderate Resolution Imaging Spectroradiometer (MODIS) was used. The GEOS-Chem chemical transport model, a global 3-dimensional chemical transport model, was used to simulate ground-level $PM_{2.5}$ concentrations. Meteorological variables were collected from NCEP North American Regional Reanalysis data, which includes daily estimated meteorological variables (at 0.3º grid cells), such as air temperature, accumulated total precipitation, downward shortwave radiation flux, wind speed, and humidity. Absorbing aerosol index (AAI) was also considered in the prediction model to consider absorbing aerosols, such as organic carbon and soil dust. In addition, land-use data were used to consider elevation, road density, emission inventory, population density, % of urban, and residential greenness. R^2 between cross-validated predicted $PM_{2.5}$ and monitored $PM_{2.5}$ was calculated to quantify model performance, and 10-fold cross-validated R^2 of 0.89 for annual $PM_{2.5}$ predictions across the US.

NO_2. We collected daily NO2 predictions at a lkm^2 spatial resolution across the 48 contiguous US states and Washington DC from a well-validated ensemble model.[2] Total 912 No2 monitoring sites operated by the US Environmental Protection Agency were included in this prediction model. First, 16 meteorological variables from the National Centers for Environmental Prediction (NCEP) and National Center for Atmospheric Research (NCAR), such as surface air temperature, accumulated total precipitation, specific humidity at 2m, and medium cloud area fraction, were used with spatial resolution of approximately 32 km. NO_2 column density (from the Aura satellite) and chemical transport models (the global-scale GEOS-Chem and the regional-scale Community Multiscale Air Quality model) were used to estimate surface-level NO_2 as a predictor variable in the modeling. In addition, seven categories of land-cover variables (land-cover types, truck traffic, road density, restaurant density, elevation, normalized difference vegetation index, and nighttime light) were considered as predictor variables in the prediction model. Finally, other ancillary variables, such as variables related to aerosol concentration and aerosol type, cloud coverage, surface albedo/reflectance, were also considered in this model. R^2 between cross-validated predicted NO_2 and monitored NO_2 was calculated to quantify model performance, and 10-fold cross-validated R^2 of 0.84 for annual NO_2 predictions across the US.

C. Details on neighborhood-level indicators

We adjusted for a total of 12 neighborhood-level indicators in the main model to consider potential confounding: eight ZIP code-level indicators, two county-level indicators, average temperature for each ZIP code, and indicator variables indicating geographical regions.

ZIP code-level indicators: Eight indicators available at ZIP Code Tabulation Areas (ZCTA) level were derived from the 2000 U.S. Census, the 2010 U.S. Census, and the American Community Survey (ACS) from 2005-2016. If indicators were missing for a year, we linearly interpolated or extrapolated their values using available data. The ZCTA indicators included the % of the population below the poverty level, population density (persons per km^2), median home value (US $), % of the population that is Black, % of the population that is Hispanic, % of the population with other race (not Black or White), median household income (US $), % of homes with owner-occupied housing and % of the population without a high school education. These ZCTA data were matched to ZIP code.

County-level indicators: Two county-level indicators (average body mass index (BMI) and % of the population that had ever smoked) were collected from the Behavioral Risk Factor Surveillance System (BRFSS) for the period of 2000-2016. These county-level indicators were matched to ZIP code if the ZIP code centroids fell within the country boundary.

ZIP code-level average temperature: We collected annual average air temperature (2m) across the continental US (2000-2016), which was provided from the North American Regional Reanalysis (NARR) with grids that were approximately 32km*32km, and assigned the annual average temperature for each ZIP code based on the nearest grid cell for each ZIP code centroid.

Region indicator: We used indicator variables for regions in the US (Northeast, Southeast, Midwest, Southwest, and West).

D. Supplementary Tables

Table S1. Region-specific association (HR and 95% CI) between air pollution and the first hospital admission due to total renal system disease or chronic kidney disease (CKD) in Medicare Part A FFS 2000-2016. Hazard ratio: $PM_{2.5}$ (per 5 $\mu g/m^3$) and NO_2 (per 10 ppb).

	$PM_{2.5}$	NO_2
Total renal system disease		
Midwest	1.080 (1.066, 1.094)	1.113 (1.085, 1.141)
Northeast	1.070 (1.057, 1.083)	0.998 (0.992, 1.004)
Southeast	1.049 (1.037, 1.062)	1.038 (1.031, 1.046)
Southwest	1.132 (1.113, 1.151)	1.001 (0.991, 1.010)
West	0.990 (0.980, 0.999)	1.055 (1.048, 1.061)
Chronic kidney disease (CKD)		
Midwest	1.129 (1.105, 1.154)	1.093 (1.046, 1.141)
Northeast	1.069 (1.046, 1.093)	0.978 (0.968, 0.988)
Southeast	1.043 (1.023, 1.063)	1.030 (1.016, 1.043)
Southwest	1.193 (1.161, 1.226)	0.992 (0.978, 1.006)
West	0.979 (0.962, 0.996)	1.029 (1.019, 1.040)

Table S2. Sensitivity analysis of the association (HR and 95% CI) between air pollution and risk of first hospital admissions for total renal system disease or chronic kidney disease (CKD) based on: selection of confounders, inclusion of prevalent cases, exposure time window, and potential outcome misclassification in Medicare Part A FFS 2000-2016. Hazard ratio: $PM_{2.5}$ (per 5 $\mu g/m^3$) and NO_2 (per 10 ppb).

	$PM_{2.5}$	**NO_2**
Total renal system disease		
Main model	1.076 (1.071, 1.081)	1.040 (1.036, 1.043)
Excluding neighborhood-level confounders	1.105 (1.099, 1.110)	1.002 (0.998, 1.006)
Excluding region indicator variables	1.089 (1.097, 1.112)	1.037 (1.033, 1.040)
Primary diagnosis code only	1.104 (1.097, 1.112)	1.062 (1.047, 1.076)
Exposure with 1-year lag period	1.074 (1.069, 1.079)	1.041 (1.038, 1.045)
Excluding potential prevalent cases (exclusion of the first 2 years of follow-up)	1.083 (1.077, 1.088)	1.044 (1.040, 1.048)
Chronic kidney disease (CKD)		
Main model	1.106 (1.097, 1.115)	1.013 (1.008, 1.019)
Excluding neighborhood-level confounders	1.145 (1.136, 1.154)	0.953 (0.948, 0.958)
Excluding region indicator variables	1.128 (1.119, 1.138)	1.017 (1.012, 1.022)
Primary diagnosis code only	0.987 (0.963, 1.012)	0.940 (0.895, 0.987)
Exposure with 1-year lag period	1.103 (1.094, 1.112)	1.013 (1.008, 1.018)
Excluding potential prevalent cases (exclusion of the first 2 years of follow-up)	1.116 (1.106, 1.125)	1.015 (1.009, 1.020)

References for the Appendix

1. Di Q, Kloog I, Koutrakis P, et al. Assessing PM2. 5 exposures with high spatiotemporal resolution across the continental United States. *Environmental science technology* 2016;50(9):4712-21.
2. Di Q, Amini H, Shi L, et al. Assessing NO2 concentration and model uncertainty with high spatiotemporal resolution across the contiguous United States using ensemble model averaging. *Environmental science technology* 2019;54(3):1372-84.
3. Wei Y, Yazdi MD, Di Q, et al. Emulating causal dose-response relations between air pollutants and mortality in the Medicare population. *Environmental Health* 2021;20(1):1-10.
4. Shi L, Wu X, Yazdi MD, et al. Long-term effects of PM2· 5 on neurological disorders in the American Medicare population: a longitudinal cohort study. 2020