HDSI | Harvard Data Science Initiative

HARVARD T.H. CHAN SCHOOL OF PUBLIC HEALTH
Powerful ideas for a healthier world

Low-Level Air Pollution Exposure & Risk of Mortality in Older Americans

Francesca Dominici, PhD

Clarence Gamble Professor of Biostatistics, Population Health and Data Science

Harvard T.H. Chan School of Public Health

Co-Director of the Harvard Data Science Initiative

Danielle Braun, PhD

Senior Research Scientist

Harvard T.H. Chan School of Public Health

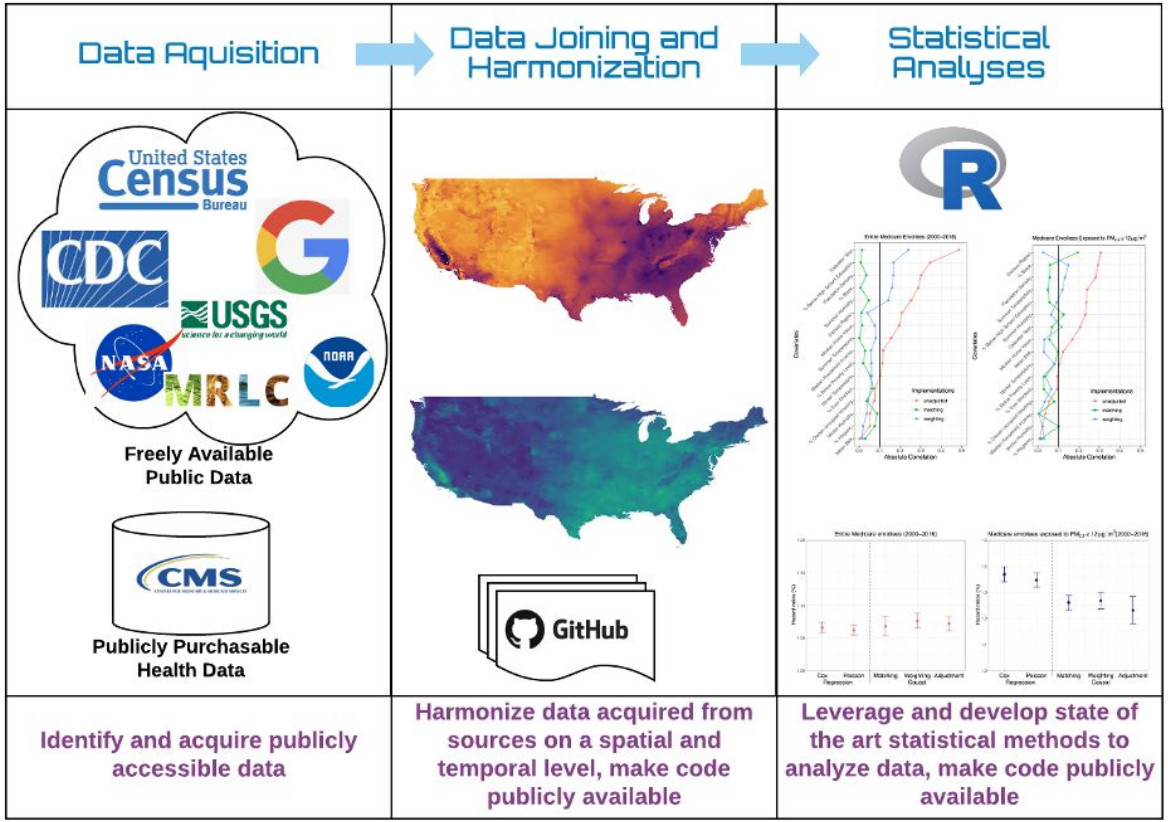
Dana-Farber Cancer Institute

Overall Goals

Our study was designed to advance **four** critical areas:

1) Predicting short- and long-term exposures to ambient fine particulate matter ($PM_{2.5}$), nitrogen dioxide (NO_2), and ozone (O_3) at high spatial resolution (1 km x 1 km) for the continental US during the period 2000–2016 and linking these predictions to health data.
2) Developing new causal inference methods for estimating the exposure-response curve and adjusting for measured confounders.
3) Applying these methods to claims data from Medicare and Medicaid enrollees to estimate health effects associated with short- and long-term exposure to low levels of ambient air pollution.
4) Developing pipelines for reproducible research.

Description of Research Data Platform



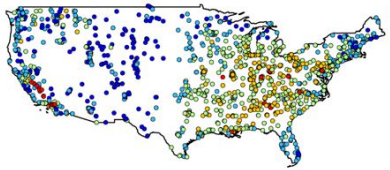
(detailed list and software codes are available at https://github.com/NSAPH/National-Casual-Analysis)

Air Pollution Data

Joel Schwartz

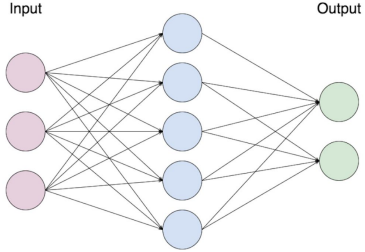
Satellite Imaging

$PM_{2.5}$ Monitor Data

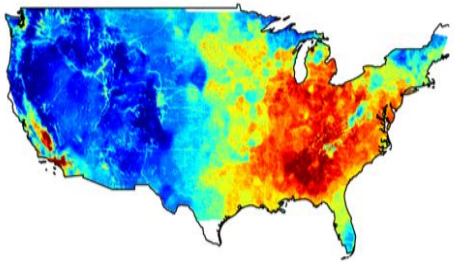


Weather data

Land use data



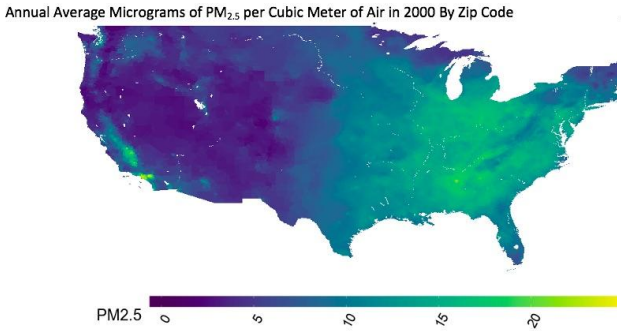
Daily 1km x 1km Estimates

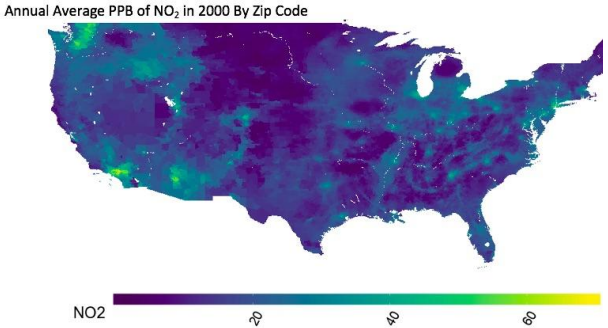


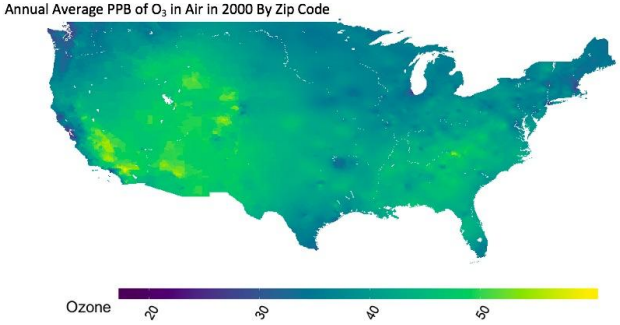
Di Q et al. 2019. An ensemble-based model of PM2.5 concentration across the contiguous United States with high spatiotemporal resolution. Environ Int 130:104909, 10.1016/j.envint.2019.104909
Requia WJ et al. 2020. An ensemble learning approach for estimating high spatiotemporal resolution of ground-level ozone in the contiguous United States. Environmental science & technology, 10.1021/acs.est.0c01791
Di Q et al. 2020. Assessing NO2 concentration and model uncertainty with high spatiotemporal resolution across the contiguous United States using ensemble model averaging. Environmental science & technology, 10.1021/acs.est.9b03358

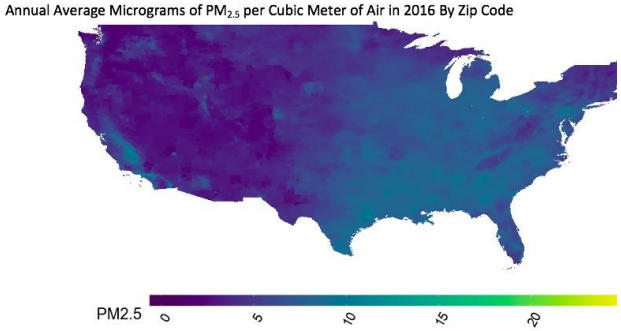
Publicly Available:
https://beta.sedac.ciesin.columbia.edu/data/collection/aqdh

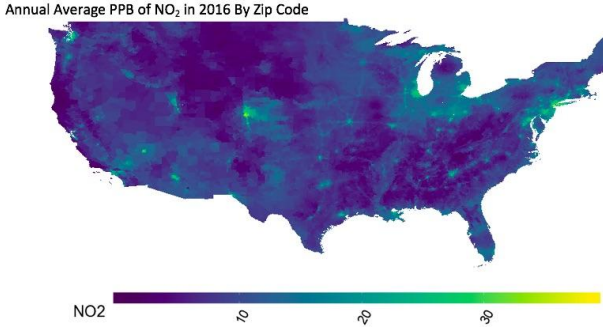
Air Pollution Data

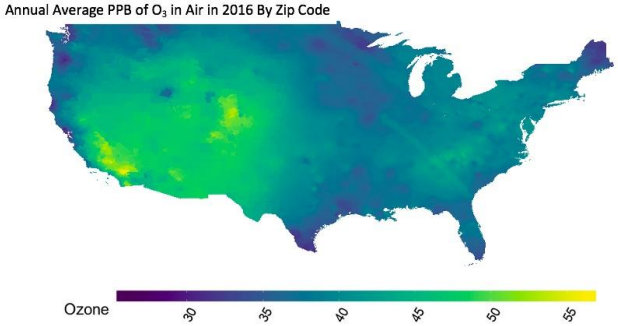












Publicly Available:
https://beta.sedac.ciesin.columbia.edu/data/collection/aqdh

Data Sources

Source	Dataset	Website
NOAA	Re-analysis meteorological data	http://www.noaa.gov/
NASA	MAIAC AOD data	https://www.nasa.gov/
	Surface reflectance data	
	NDVI data	
	OMI Aerosol Index Data GEOS-Chem simulation outputs	http://acmg.seas.harvard.edu/geos/
U.S. Geological Survey	Global terrain elevation data	https://lta.cr.usgs.gov/
Census Bureau	Road density, population count and area	https://www.census.gov/
MRLC	National Land Cover Dataset	https://www.mrlc.gov/
EPA	AQS monitoring data ($PM_{2.5}$ and O_3)	https://www.epa.gov/aqs
CMS	Medicare denominator files	https://www.cms.gov/
	Medicare Current Beneficiary Survey	
CDC	BMI, smoking rate	https://www.cdc.gov/

(detailed list and software codes are available at https://github.com/NSAPH/National-Casual-Analysis)

Potential Confounders

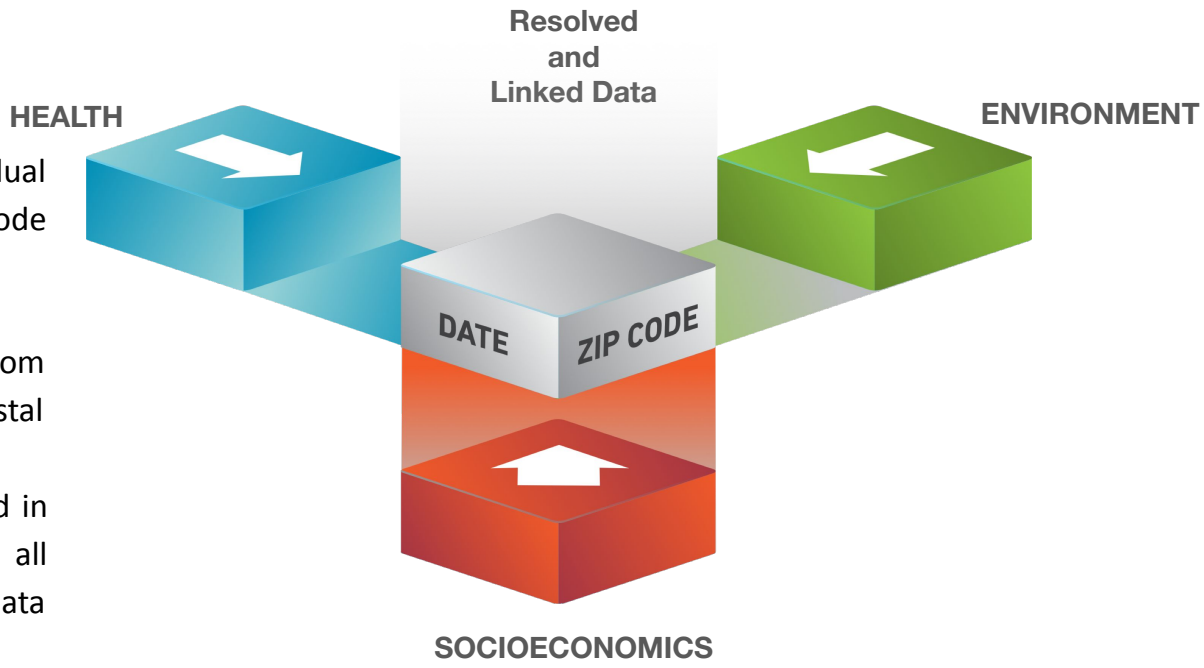
- Socioeconomic (SES) indicators from the 2000 and 2010 Census and the 2005–2016 American Community Surveys (ACS)
- County-level information from the Centers for Disease Control and Prevention's Behavioral Risk Factor Surveillance System (BRFSS).
- Zip code-level meteorological variables using area-weighted aggregations based on daily temperature and humidity data on 4 km^2 gridded rasters from Gridmet via Google Earth Engine (Abatzoglou 2013; Gorelick et al. 2017).

Health Data (Medicare)

- All Medicare participants (n=68,503,979) in the continental United States from 2000 to 2016
- Outcomes: all-cause mortality and cause specific hospitalization
- Individual level information: date of death, age of entry, year of entry, sex, race, whether eligible for Medicaid (proxy for SES)
- Zip code of residence and other covariates

Harmonizing and Integrating Heterogeneous Sources of Data

- Health data were available at individual level, residence is known at postal zip code level.
- Exposure is assigned at zip code level
- Potential confounders are mapped from zip code tabulation areas (ZCTA) to postal zip codes.
- The total number of zip codes included in our main analysis with information on all outcome, exposure, and confounder data was 31,337.



Di et al. 2017

The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 JUNE 29, 2017 VOL. 376 NO. 26

Air Pollution and Mortality in the Medicare Population

Qian Di, M.S., Yan Wang, M.S., Antonella Zanobetti, Ph.D., Yun Wang, Ph.D., Petros Koutrakis, Ph.D., Christine Choirat, Ph.D., Francesca Dominici, Ph.D., and Joel D. Schwartz, Ph.D.

Table 1. Cohort Characteristics and Ecologic and Meteorologic Variables.

Characteristic or Variable	Entire Cohort	Ozone Concentration		$PM_{2.5}$ Concentration	
		≥50 ppb*	<50 ppb	≥12 $\mu g/m^3$	<12 $\mu g/m^3$
Population					
Persons (no.)	60,925,443	14,405,094	46,520,349	28,145,493	32,779,950
Deaths (no.)	22,567,924	5,097,796	17,470,128	10,659,036	11,908,888
Total person-yr†	460,310,521	106,478,685	353,831,836	212,628,154	247,682,367
Median yr of follow-up	7	7	7	7	7
Average air-pollutant concentrations‡					
Ozone (ppb)	46.3	52.8	44.4	48.0	45.3
$PM_{2.5}$ ($\mu g/m^3$)	11.0	10.9	11.0	13.3	9.6

Di et al. 2017

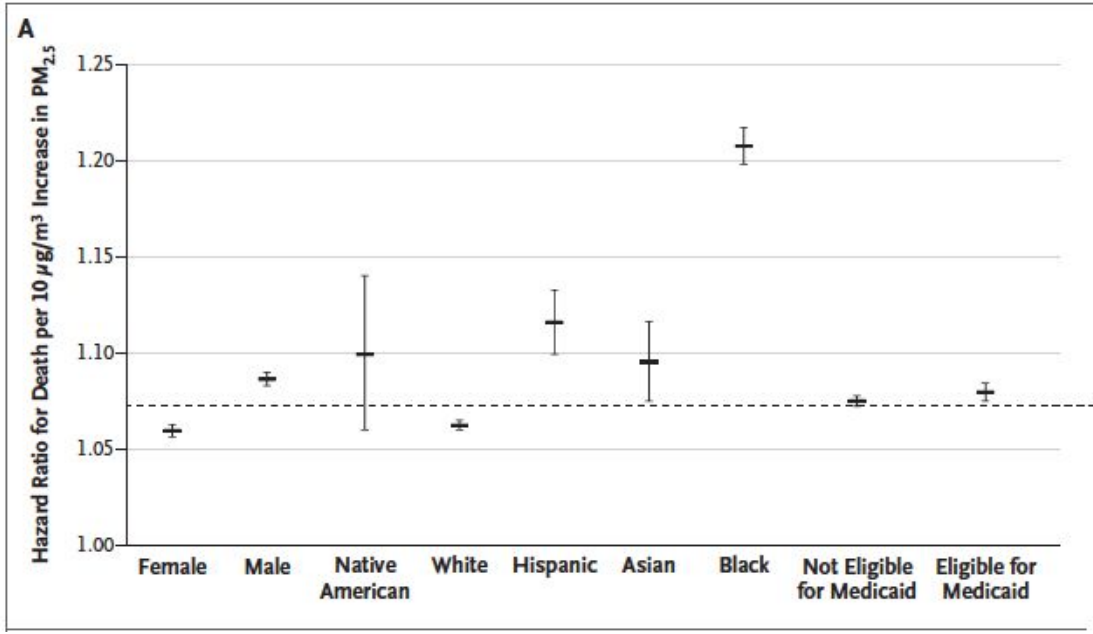


Figure 2. Risk of Death Associated with an Increase of 10 μg per Cubic Meter in $PM_{2.5}$ Concentrations and an Increase of 10 ppb in Ozone Exposure, According to Study Subgroups.

Hazard ratios and 95% confidence intervals are shown for an increase of 10 μg per cubic meter in $PM_{2.5}$ and an increase of 10 parts per billion (ppb) in ozone. Subgroup analyses were conducted by first restricting the population (e.g., considering only male enrollees). The same two-pollutant analysis (the main analysis) was then applied to each subgroup. Numeric results are presented in Tables S3 and S4 in the Supplementary Appendix. Dashed lines indicate the estimated hazard ratio for the overall population.

- A 10 units increase in $PM_{2.5}$ is associated with a 7.3% increase in all cause mortality among 60 million older American (evidence of a link is even stronger at levels of $PM_{2.5}$ below the NAAQS)
- African American have a risk of death from $PM_{2.5}$ exposure that is three times higher than the national average

Di, et al. Air pollution and mortality in the Medicare population. New England Journal of Medicine. 2017 Jun 29;376(26):2513-22.

Wu et al. 2020

Science Advances

RESEARCH ARTICLES

Cite as: X. Wu *et al.*, *Sci. Adv*
10.1126/sciadv.aba5692 (2020).

Evaluating the impact of long-term exposure to fine particulate matter on mortality among the elderly

X. Wu,[1]† D. Braun,[1,2]† J. Schwartz,[3] M. A. Kioumourtzoglou,[4] F. Dominici[1]*

Statistical Approaches for Estimating Health Effects

- Di et al use only **traditional approaches**.
- Wu et al 2020, extends the analysis to 2016 using:

Five models to estimate health effects between long-term exposure to $PM_{2.5}$ and all-cause mortality among the elderly.

- **Traditional** approaches
  1. Cox Proportional Hazard Approach
  2. Poisson Regression
- **Causal inference** approaches using generalized propensity scores (GPS)
  1. Matching approach
  2. Weighting approach
  3. Adjustment approach

We evaluate the 95% confidence intervals (CIs) for all models by m-out-n subsampling blocked bootstrap to account for spatial correlation.

Generalized Propensity Score (GPS)

GPS estimation: modeled the conditional density of exposure (i.e., zip code-level annual average $PM_{2.5}$) on the 14 zip code- or county-level time-varying covariates, as well as a dummy region variable and dummy calendar year variable, by using gradient boosting machine with normal residuals (Chen and Guestrin 2016; Zhu et al. 2015).

$PM_{2.5}$~ area-level risk factors + meteorological variables + dummy year + dummy region + ε, where $\varepsilon \sim N(0, \sigma^2)$.

Area-level risk factors:

1) Two county-level variables: average BMI and smoking rate;
2) Eight zip code-level census variables: proportion of Hispanic residents, proportion of Black residents, median household income, median home value, proportion of residents in poverty, proportion of residents with a high school diploma, population density, and proportion of residents that own their house;
3) Four zip code-level meteorological variables: the summer (June-September) and winter (December-February) averages of maximum daily temperatures and relative humidity.

GPS implementation: following GPS estimation we match/weigh/adjust by GPS

Outcome analysis: post-matching/weighting fit a Poisson regression model

Causal Inference Approaches

- Can quantify and visualize how closely we are able to approximate a randomized study.
  - Visualizing whether the measured confounders are balanced across exposed and non-exposed groups (Austin 2019; Imai and Ratkovic 2014).
  - Assessing the sensitivity of results to unmeasured confounding bias (Rosenbaum 2002).
- This framework allows us to assess how confident we can make statements about causality using observational data under a set of explicit assumptions necessary for causal inference.

Results: Cohort Characteristics

Table 1. Characteristics for the study cohorts. Note that mean (SD) is presented for continuous variables.

Variables	Entire Medicare enrollees	Medicare enrollees exposed to $PM_{2.5} \leq 12\ \mu g/m^3$
Number of individuals	68,503,979	38,366,800
Number of deaths	27,106,639	10,124,409
Total person years	573,370,257	259,469,768
Median years of follow-up	8.0	8.0
Individual-level characteristics		
Age at entry (years)		
65–74 (%)	80.6	88.1
75–84 (%)	14.9	9.0
85–94 (%)	4.1	2.6
95 or above (%)	0.4	0.2
Mean (SD)	69.2 (6.7)	67.6 (5.6)
Sex		
Female (%)	55.5	53.8
Male (%)	44.5	46.2
Race		
White (%)	83.9	84.7
Black (%)	9.1	7.3
Asian (%)	1.8	1.8
Hispanic (%)	2.0	2.2
North American Native (%)	0.3	0.4
Medicaid eligibility		
Eligible (%)	11.7	10.9

Wu X, et al. Evaluating the impact of long-term exposure to fine particulate matter on mortality among the elderly. Science advances. 2020 Jul 17;6(29):eaba5692.

Results: Covariate Balance

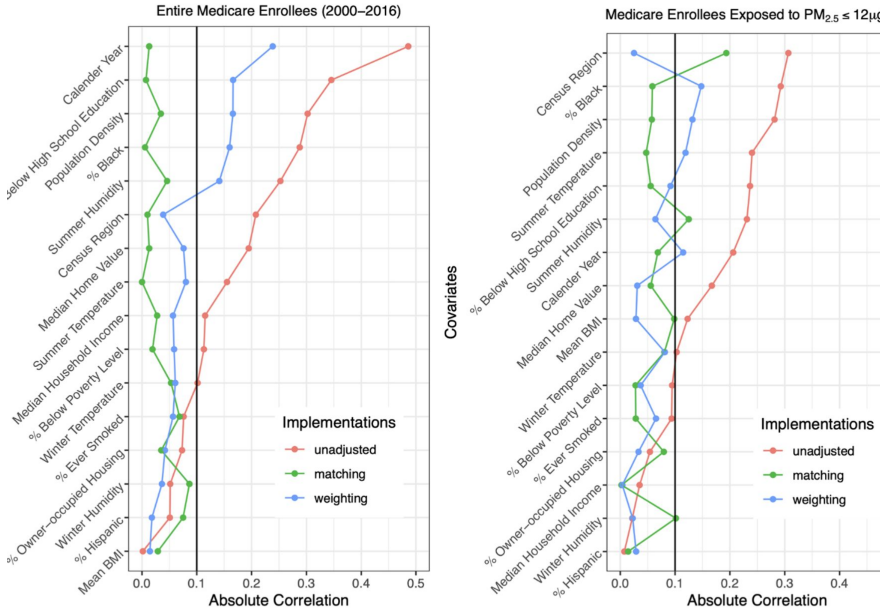


Fig. 2. Mean absolute correlation (AC) for Unadjusted, Weighted, and Matched Populations. Mean AC was smaller than 0.1 using causal inference GPS methods (matching and weighting). AC values <0.1 indicate good covariate balance, strengthening the interpretability and validity of our analyses as providing evidence of causality.

The causal inference framework lends itself to the evaluation of covariate balance for measured confounders. An absolute correlation (AC), with values <0.1 indicating high quality recovering randomized experiments.

Wu X, et al. Evaluating the impact of long-term exposure to fine particulate matter on mortality among the elderly. Science advances. 2020 Jul 17;6(29):eaba5692.

Results

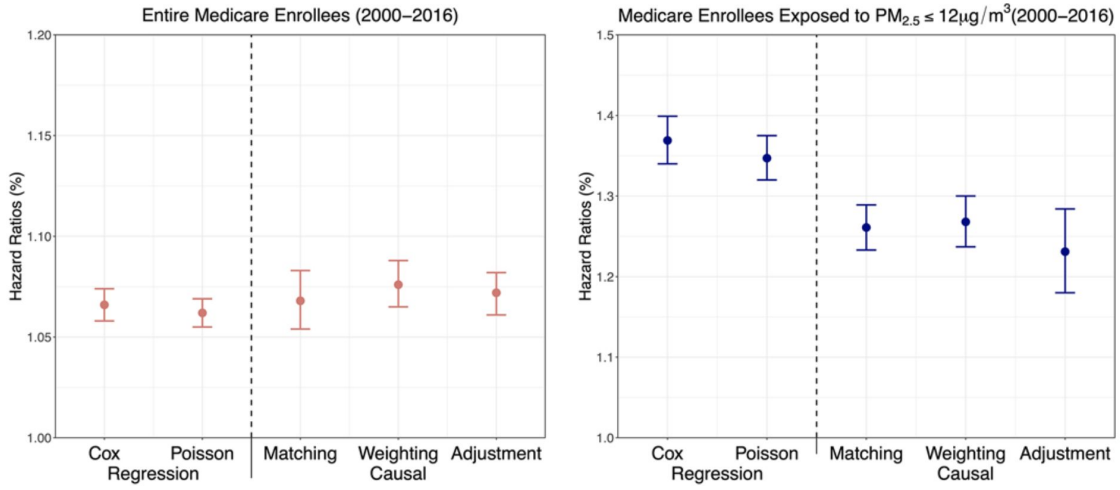


Fig. 3. Hazard Ratios (HR) and 95% Confidence Intervals (CIs). The estimated HRs were obtained under five different statistical approaches (two traditional approaches and three causal inference approaches). HRs were adjusted by 10 potential confounders, four meteorological variables, geographic region, and year.

Using five distinct statistical approaches, we found that a decrease of 10 μg/m^3 $PM_{2.5}$ leads to a statistically significant 6%–7% decrease in mortality risk.

Based on these models, lowering the air quality standard to 10 μg/m^3 would save 143,257 lives (95% confidence interval 115,581–170,645) in one decade

Wu X, et al. Evaluating the impact of long-term exposure to fine particulate matter on mortality among the elderly. Science advances. 2020 Jul 17;6(29):eaba5692.

Strengths of Causal Inference Methods

- Separate the design stage from the outcome analysis,
  - → mimicking a randomized experiment under a set of explicit identification assumptions
  - → increasing the objectiveness of causal analysis.
- Guide researchers to explicitly state all assumptions,
  - → using sensitivity analysis tools to understand how likely the identification assumptions are held (e.g., covariate balance, E-value, etc.).
- More robust to model misspecification compared to traditional regression approaches.

Limitations of Causal Inference Methods

- Often require increased computational resources due to the complexity of algorithms.
- Some methods require steeper learning curves for new researchers due to the logic complexity and are often less familiar to many researchers.
- Methods based on GPS are still affected by unmeasured confounding bias.
- Propagation of exposure error in health effects analyses under a causal inference framework is challenging.

  $\rightarrow$ error in the exposure also affects the propensity score (see Wu et al. 2019)

Methods for Estimating Non-linear Exposure Response Functions

- Applied the proposed GPS matching method to estimate the effect of long-term exposures to $PM_{2.5}$, NO_2, and O_3 on all-cause mortality
- To estimate the **Exposure Response Curve (ERC)**
  - Used a Kernel smoothing approach, a non-parametric approach.
  - Defined the baseline rate as the estimated hazard rate corresponding to an exposure level equal to the 1st percentile of the distribution of that pollutant.
  - To avoid extrapolation at the support boundaries, exclude the highest 1% and lowest 1% of pollutants exposures when plotting the ERC curves.
  - We evaluate the 95% confidence intervals (CIs) for all models by m-out-n subsampling blocked bootstrap to account for spatial correlation.

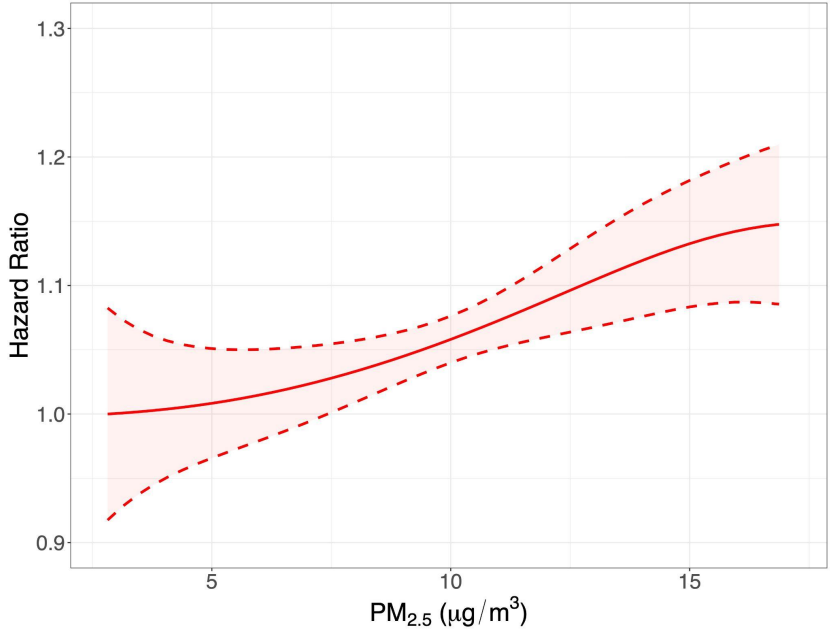
Single vs. Multi-Pollutant ERC

For each of the three pollutants we present **two** ERC:

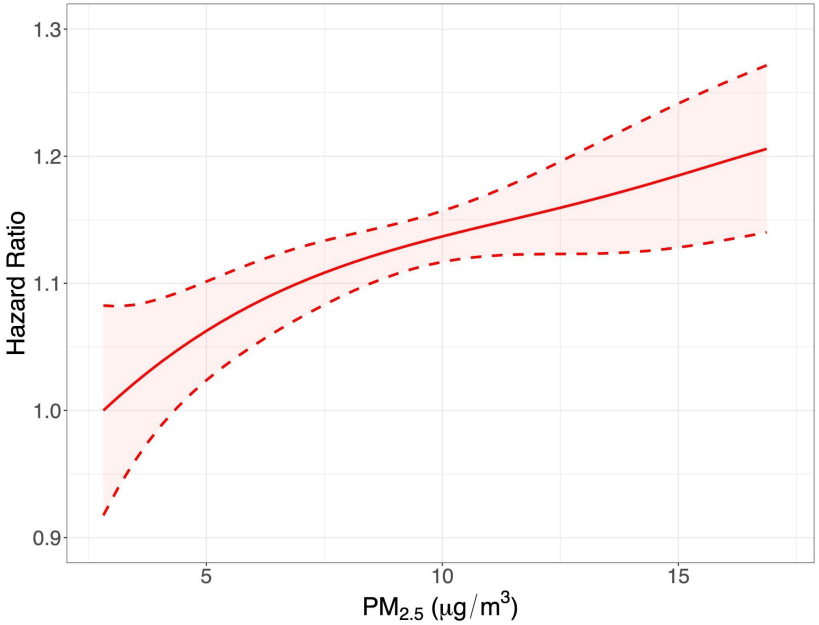
1. Multiple pollutant models adjusting for other two pollutants as potential confounders.
   - **GPS model:** includes the 14 zip code- or county-level time-varying covariates and in addition the **two other pollutants.**
2. Single pollutant models without adjusting for other pollutants.
   - **GPS model:** includes only the 14 zip code- or county-level time-varying covariates.

Results for Non-linear Exposure Response Function; $PM_{2.5}$

multiple pollutant models adjusting for other two pollutants as potential confounders



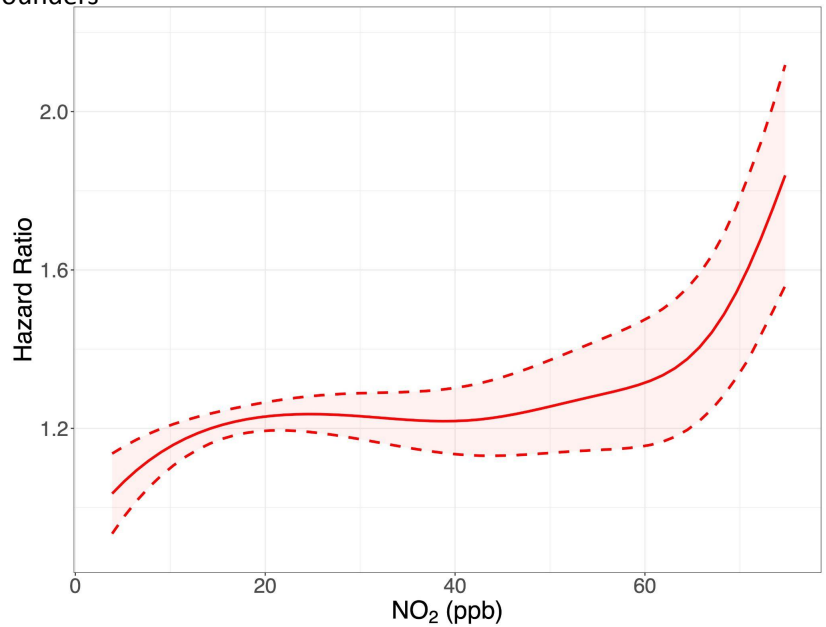
single pollutant models without adjusting for other pollutants



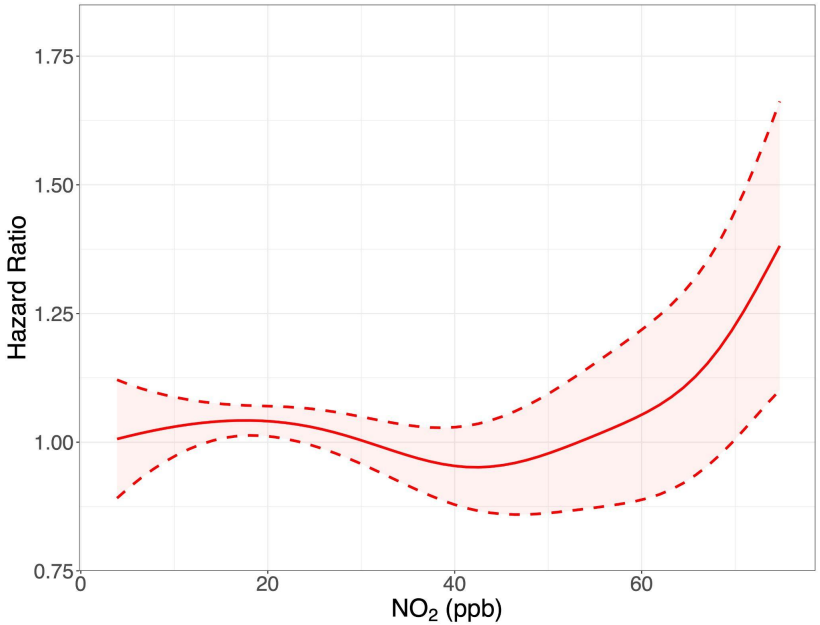
- Evidence of a harmful causal relationship between mortality and long-term $PM_{2.5}$ exposures adjusted for NO_2 and O_3 across the range of annual average between 2.77 and 17.16 (included > 98% of observations).
- The curve is almost linear at low exposure levels → aggravated harmful effects at exposure levels even below the national standards.

Results for Non-linear Exposure Response Function; NO_2

multiple pollutant models adjusting for other two pollutants as potential confounders



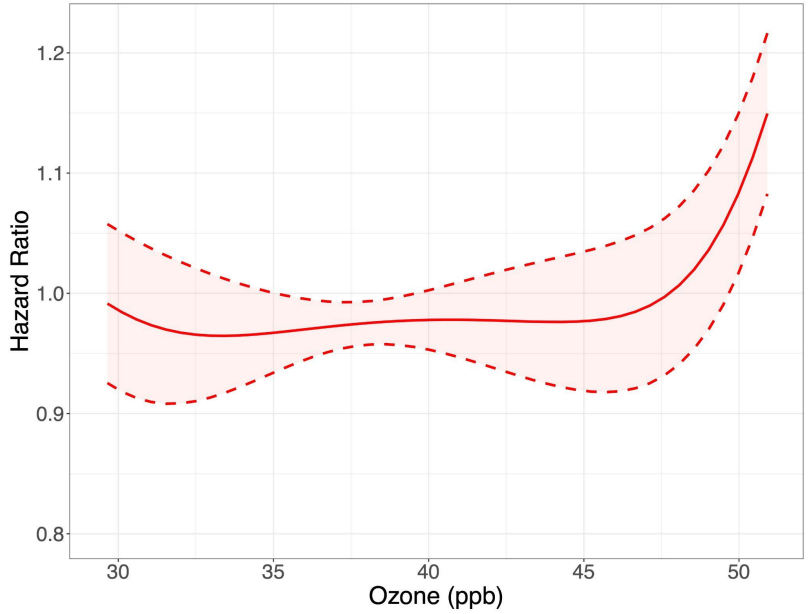
single pollutant models without adjusting for other pollutants.



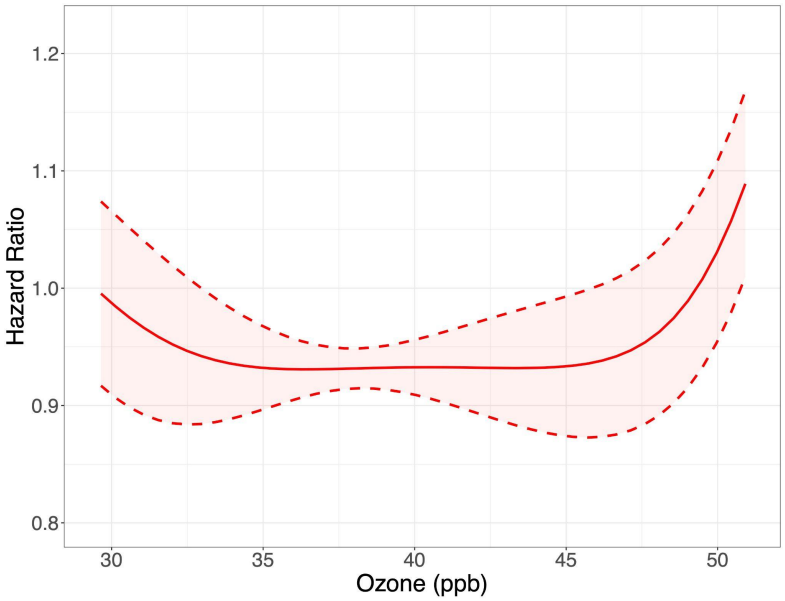
- Harmful impact of long-term NO_2 exposures to mortality adjusted for $PM_{2.5}$ and O_3 across the range of annual average between 3.4 and 80 ppb (included > 98% of observations).
- Within low levels, the causal impacts of NO_2 exposures on all-cause mortality is non-linear with statistical uncertainty.

Results for Non-linear Exposure Response Function; Ozone

multiple pollutant models adjusting for other two pollutants as potential confounders



single pollutant models without adjusting for other pollutants.



- The exposure-response curves of long-term O_3 exposures on all-cause mortality adjusted for $PM_{2.5}$ and NO_2 is almost flat below 45 ppb, which shows no statistically significant effect.
- We observed an increased hazard when the O_3 exposures are higher than 45 ppb.

Results for Multi-Pollutant Analysis (Constant HR)

Pollutants	Models	GPS Matching	Poisson
$PM_{2.5}$	Adjusted for NO_2 and O_3	1.036 (1.023, 1.065)	1.056 (1.049, 1.063)
	Unadjusted for NO_2 and O_3	1.063 (1.050, 1.077)	1.067 (1.060, 1.075)
	Adjusted for NO_2 only	1.044 (1.031, 1.057)	1.055 (1.048, 1.062)
NO_2	Adjusted for $PM_{2.5}$ and O_3	0.997 (0.992, 1.001)	1.009 (1.006, 1.012)
	Unadjusted for $PM_{2.5}$ and O_3	0.996 (0.992, 1.001)	1.017 (1.014, 1.020)
O_3	Adjusted for $PM_{2.5}$ and NO_2	1.004 (0.995, 1.012)	0.994 (0.990, 0.998)
	Unadjusted for $PM_{2.5}$ and NO_2	1.007 (0.999, 1.015)	0.996 (0.992, 1.000)

The HRs, per 10 $\mu g/m^3$ increase in $PM_{2.5}$, relate three air pollutants to all-cause mortality among Medicare enrollees (2000–2016). These estimated HRs are obtained using both the GPS matching method and multivariate Poisson regression method under the assumption of a constant linear HR.

Results for Multi-Pollutant Analysis

Overall, adjusting for the other two pollutants:

- slightly attenuated the causal effects of $PM_{2.5}$
- slightly elevated the causal effects of NO_2
- results for O_3 remained almost unchanged

Pipelines for Reproducible Research

- Study relied entirely on publicly available data.
- Relied instead on *privacy-protected but publicly available* Medicare health data including almost 97% of the US population older than 65 years over the years 2000–2016.
- Made the software code and workflows available in open, trusted digital repositories.
- Reproducibility instructions and open-source software are hosted on GitHub and are publicly available, https://github.com/NSAPH/National-Casual-Analysis
- Developed statistical R package CausalGPS, available on CRAN,

  2903 downloads:

  https://github.com/fasrc/CausalGPS

Dev status

R-CMD-check passing

CRAN 0.2.7 – 2022-02-07

downloads 2903

codecov 94%

launch binder

Overall Study Strengths

1. The massive and representative study population
2. The numerous sensitivity analyses
3. The transparent assessment of covariate balance that indicates the quality of causal inference for recovering randomized experiments
4. This work relies on publicly available data, and we provide code that allows for reproducibility of our analyses.

→ Conditional on the required assumptions for causal inference, collectively our results indicate that long-term $PM_{2.5}$ exposure is likely to be causally related to mortality

Overall Study Assumptions

1) **No unmeasured confounders**
2) **Model misspecification**
3) **Measurement error**
4) **Spatial correlation**
5) **Hybrid study design**

Overall Study Assumptions

1) **No unmeasured confounders**
   - We account for individual- and area-level potential confounders.
   - To mitigate unmeasured confounding bias, we assessed the results' sensitivity by including year as a surrogate for some unmeasured confounders that might have covaried over time with $PM_{2.5}$ and mortality and, thus, confounded their association.
   - Also conducted further sensitivity analyses to unmeasured confounding by calculating the E-value → our results are **robust** to unmeasured confounding bias.

Overall Study Assumptions

2) Model Misspecification

- The causal inference approaches require the estimation of the GPS.
- Assuming all causal inference assumptions hold, these approaches are more robust to outcome model misspecification, and allow for the transparent evaluation of covariate balance.
- However, it is important to note that if the models are accurately specified and all assumptions are met, the traditional approaches have the potential to inform causal relationships as well.

Overall Study Assumptions

3) Measurement error

- Air pollution exposures were estimated from prediction models, which, while very good, are not perfect.
- The $PM_{2.5}$ exposure prediction model developed by Di et al. indicated excellent model performance, with a 10-fold cross-validated R^2 of 0.89 for annual $PM_{2.5}$ predictions.
- Di et al. assessed the robustness of the results to the exposure predictions by repeating the analysis based on $PM_{2.5}$ exposure data obtained from 1928 EPA ambient monitors.
  - While this subset does not represent the entire population, analysis based on nearest monitoring site led to a HR estimate that was only slightly lower than the one obtained using the exposure prediction model (i.e., 1.061, 95% CI [1.059 to 1.063] vs. 1.073, 95% CI [1.071 to 1.075]).

Overall Study Assumptions

3) Measurement error (continued)

- How to propagate exposure error under a causal inference framework for a continuous exposure under a causal inference framework is still an area of active research.
  - Wu et al. proposed a regression calibration approach for GPS analysis under categorical exposures.
  - The proposed approach was applied in the context of long-term $PM_{2.5}$ exposure and mortality using Medicare data in the Northeastern US. When accounting for exposure error, there was a higher and still statistically significant association between exposure to $PM_{2.5}$ and mortality, although with larger CIs.
  - Currently working on extending this to continuous setting.
- Potential measurement error in covariates is also important to account for and is subject of future research.

Josey KP, et al. Estimating a Causal Exposure Response Function with a Continuous Error-Prone Exposure: A Study of Fine Particulate Matter and All-Cause Mortality. 2021, arXiv preprint https://arxiv.org/abs/2109.15264

Ren B, et al. Bayesian modeling for exposure response curve via gaussian processes: Causal effects of exposure to air pollution on health outcomes. 2021, arXiv preprint https://arxiv.org/abs/2105.03454

Overall Study Assumptions

4) Spatial Correlation

- The model parameterization assumes that zip code-specific information is spatially independent, given covariates.
- Since we adjusted for numerous zip code-level predictors of mortality, including SES and meteorological variables, this assumption likely holds.
- Any remaining spatial dependence is partially accounted for by our bootstrapping procedure
    - By randomly sampling zip codes for each bootstrap replicate, we were able to break down spatial dependence given covariates. Therefore, it is unlikely that our results are impacted by spatial correlation.
- We adjusted for potential spatial confounding that is not captured by zip code-level observed covariates by including a dummy region variable.

Overall Study Assumptions

5) Hybrid study design

- Medicare claims are available at individual level, and they include information on age, sex, race, eligibility to Medicaid (a proxy for low income).
- To increase confidence in our results, we conducted a study by Makar et al. where we linked Medicare claims data to data from the Medicare Current Beneficiary Survey (MCBS) at the individual level.
  - MCBS provides information on an extensive list of individual-level behavioral risk factors (over 100 potential measured confounders at the individual level).
  - We found that the estimated hazard ratios remain unchanged.

Conclusions

- Our work provides comprehensive evidence on the association between exposure to $PM_{2.5}$, NO_2, and O_3 and various health outcomes.
- We observe a causal link between long-term exposure to $PM_{2.5}$ and mortality, even at $PM_{2.5}$ levels below 12 $\mu g/m^3$ and mortality among Medicare enrollees (65 years of age or older) (Wu et al. 2020).
  - This work relies on newly developed causal inference methods for continuous exposures (Wu et al. 2018b).
  - Developed statistical software: over **2900** downloads!
- In the multi-pollutant analyses adjusting for the other two pollutants:
  - slightly attenuated the causal effects of $PM_{2.5}$
  - slightly elevated the causal effects of NO_2
  - results for O_3 remained almost unchanged
- Our studies are based on publicly available data sources, and we have made all code publicly available → maximizes reproducibility and transparency.

Team Members

Francesca Dominici

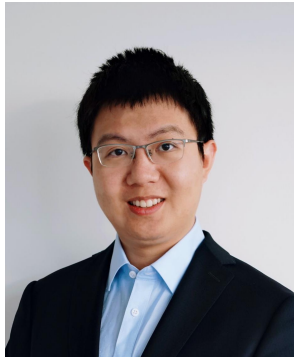
Antonella Zanobetti

Joel Schwartz

Danielle Braun

Marianthi Kioumourtzoglou

Rachel Nethery

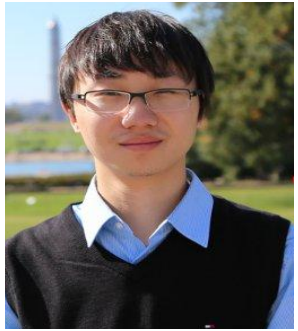


Xiao Wu

Naeem Khoshnevis

Leila Kamareddine

Mahdieh Danesh Yazdi



Yaguang Wei

Acknowledgements

We would like to thank the Health Effects Institute for their support of our work, without which we would not have been able to advance the methods or evidence base at a scope and scale that is essential to understanding the health impacts of long-term exposure to ambient pollution.

We would also like to thank Harvard University for providing a stellar research computing environment. The computations in these articles were run on the Odyssey cluster, which is supported by the Harvard Faculty of Arts and Sciences' Division of Science, Research Computing Group, and on the Research Computing Environment, which is supported by the Institute for Quantitative Social Science in the Faculty of Arts and Sciences, both at Harvard University.

References

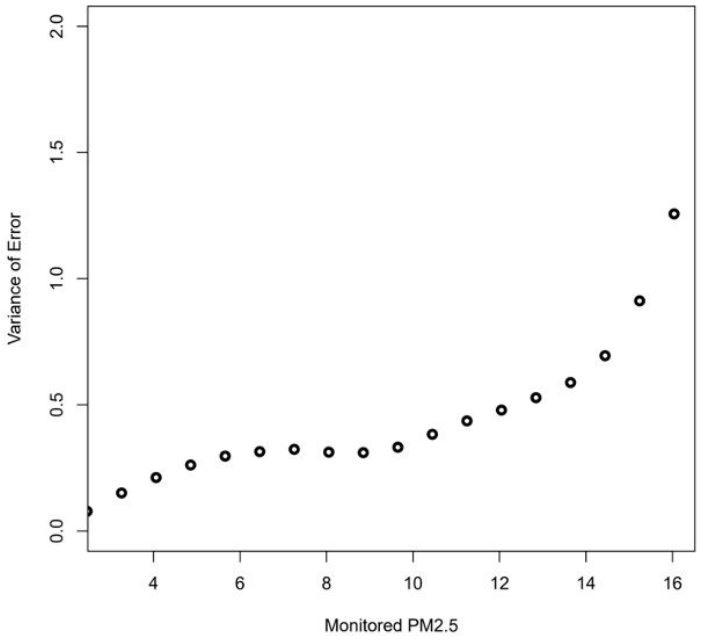
- Di Q, Wang Y, Zanobetti A, Wang Y, Koutrakis P, Dominici F, Schwartz J. (2017). Air Pollution and Mortality in the Medicare Population. *New England Journal of Medicine*, 376:2513-2522, June 29, 2017. PMID: 29020593 DOI: 10.1056/NEJMoa1702747 PMCID: 5766848
- Wu X, Braun D, Schwartz J, Kioumourtzoglou M, Dominici F, 2020 Evaluating the Impact of Long-term Exposure to Fine Particulate Matter on Mortality Among the Elderly, Science Advances, *Science Advances* 26 Jun 2020: eaba5692 DOI: 10.1126/sciadv.aba5692
- Wu X, Braun D, Kioumourtzoglou MA, Choirat C, Di Q, Dominici F. Causal inference in the context of an error prone exposure: air pollution and mortality. The annals of applied statistics. 2019 Mar;13(1):520.
- Wu X, Mealli F, Kioumourtzoglou MA, Dominici F, Braun D. Matching on generalized propensity scores with continuous exposures. arXiv preprint arXiv:1812.06575. 2018 Dec 17. https://arxiv.org/abs/1812.06575
- Josey KP, et al. Estimating a Causal Exposure Response Function with a Continuous Error-Prone Exposure: A Study of Fine Particulate Matter and All-Cause Mortality. 2021, arXiv preprint https://arxiv.org/abs/2109.15264
- Ren B, et al. Bayesian modeling for exposure response curve via gaussian processes: Causal effects of exposure to air pollution on health outcomes. 2021 arXiv preprint https://arxiv.org/abs/2105.03454
- Di Q et al. 2019. An ensemble-based model of $PM_{2.5}$ concentration across the contiguous United States with high spatiotemporal resolution. Environ Int 130:104909, 10.1016/j.envint.2019.104909
- Requia WJ et al. 2020. An ensemble learning approach for estimating high spatiotemporal resolution of ground-level ozone in the contiguous United States. Environmental science & technology, 10.1021/acs.est.0c01791
- Di Q et al. 2020. Assessing NO2 concentration and model uncertainty with high spatiotemporal resolution across the contiguous United States using ensemble model averaging. Environmental science & technology, 10.1021/acs.est.9b03358

Questions

Additional Slides

Uncertainty of Exposure Models at Low Levels

- Found that the exposure error is smaller, not larger, at concentrations below 12 μg/m^3



Study Population

- More than 68.5 million Medicare enrollees with $\geq$ 65 years old 2000-2016, who reside in 31, 414 zip codes.
- Medicare claims data is an open cohort, including demographic information such as age, sex, race/ethnicity, date of death, and residential zip-code.
- A unique patient ID is assigned to each person to allow for tracking over time.

Table 1. Characteristics for the study cohorts

Variables	Entire Medicare Enrollees	Medicare Enrollees Exposed to $PM_{2.5}$ ≤12 μg/m^3
Number of individuals	68,503,979	38,366,800
Number of deaths	27,106,639	10,124,409
Total person-years	573,370,257	259,469,768
Median years of follow-up	8.0	8.0