The Long-run Effect of Air Pollution on Survival

Tatyana Deryugina
Gies College of Business
and NBER

Julian Reif
Gies College of Business
and NBER

NBER Summer Institute
July 27, 2023

How bad is air pollution for adult health?

- Air pollution harms health in both the short and long run

- But, the magnitude of the effect remains uncertain
  - Observational estimates are prone to bias
  - Quasi-experimental studies focus on short-run effects

- Identifying the long-run effect of chronic exposure is hard
  - Limited data on long-run outcomes
  - Variation in long-run exposure hard to find

How do we address these challenges?

1. Use variation in wind direction as instrument for daily pollution
   - Trace out mortality patterns up to one month following acute exposure
   - Limited to short-run effects of acute exposure

2. Integrate empirical estimates into dynamic production model of health
   - Can be internally validated using quasi-experimental estimates

Treatment exposure	Short-run outcomes	Long-run outcomes
Acute	Empirical estimates	Model
Chronic	-	Model

How do we address these challenges?

1. Use variation in wind direction as instrument for daily pollution
   - Trace out mortality patterns up to one month following acute exposure
   - Limited to short-run effects of acute exposure

2. Integrate empirical estimates into dynamic production model of health
   - Can be internally validated using quasi-experimental estimates

Treatment exposure	Short-run outcomes	Long-run outcomes
Acute	Empirical estimates	Model
Chronic	-	Model

Research questions

- Setting: United States population, 1972–1988
- Pollutant: sulfur dioxide (SO_2)

1. What is the short-run causal effect of acute (one-day) exposure to SO_2?
   - Instrumental variables research design
   - Main outcome: monthly (28-day) mortality

2. What is the long-run effect of chronic exposure to SO_2?
   - Production model of health from Lleras-Muney and Moreau (2022)
   - Main outcome: life expectancy

Main results

- A 1-day, 10% increase in SO_2 increases same-day mortality by 0.3 percent

- In the month following exposure:
    - Cumulative effect for cancer deaths falls to zero ("mortality displacement")
    - Cumulative effect for other diseases more than triples ("accelerated aging")
    - On net, cumulative mortality more than doubles

- Benefit of reducing lifetime SO_2 exposure by 10% is 1.2 years of extra life
    - 90% of benefits occur after age 50

Contributions to the literature

- Framework for estimating long-run survival effects of chronic exposure
  - Model calculations differ from IV extrapolation
  - Approach is similar in spirit to Athey, Chetty, and Imbens (2020)

- Health effects of air pollution (Chay and Greenstone 2003; Currie and Neidell 2005; Schlenker and Walker 2016; Hollingsworth and Rudik 2021; Alexander and Schwandt 2022; Heo, Ito, and Kotamarthi 2023)
  - We are the largest quasi-experimental study (17 years, 18 million deaths)
  - We focus on mortality dynamics

Background and Data

EPA regulates six air pollutants

- Carbon monoxide (CO)
- Ozone (O_3)
- Nitrogen dioxide (NO_2)
- Lead
- Particulate matter (PM)
- Sulfur dioxide (SO_2)

- We focus on SO_2, which is well-measured during our 1972–1988 time period
  - Regulated at the daily and annual levels

SO_2 has immediate and delayed effects

- Direct exposure to SO_2 impairs respiratory function
- SO_2 leads to formation of sulfates, a component of PM 2.5 (fine particulates)
  - Acute exposure to PM 2.5 causes premature death
- Chronic exposure to air pollution associated with "accelerated aging"
  - Risk factors for cardiovascular disease (eg, coronary artery calcification)
  - Initiation and promotion of lung cancer

Daily environmental data

- Data on SO_2 obtained from EPA site monitors
  - Not available for all counties $\rightarrow$ limiting factor in the final size of our sample

- Temperature and precipitation obtained from Schlenker and Roberts (2009)

- Wind direction and wind speed obtained from Japan Meteorological Agency

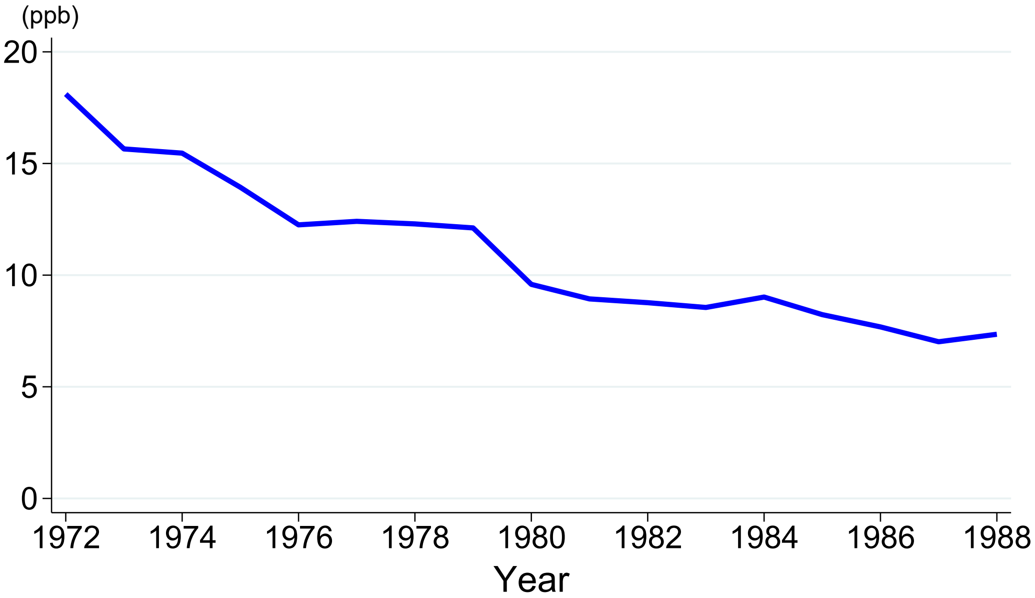
- All data are aggregated to the county-day level

Daily mortality data

- National Vital Statistics, 1972–1988
  - Exact date of death
  - County of occurrence
  - Cause of death
  - Age, sex, and race of decedent

- Merge with environmental data at the county-day level
  - Main specification includes 2.03 million county-day observations

SO_2 levels are declining during our sample period



Summary statistics

	(1)	(2)	(3)
	Mean	Std. Dev.	Observations
A. Pollution outcomes			
SO_2, ppb	8.96	12.62	2,032,338
NO_2, ppb	21.25	15.60	792,784
CO, ppm	1.64	1.37	848,067
Ozone, ppb	25.53	13.69	669,261
TSP, μg/m^3	63.11	40.19	628,932
B. One-day mortality rate outcomes			
All-cause mortality, deaths per million	24.70	24.32	2,032,338
Cardiovascular	12.21	16.04	2,032,338
Cancer	5.15	9.16	2,032,338
Other	5.45	10.02	2,032,338
External	1.89	7.99	2,032,338

Summary statistics

	(1)	(2)	(3)
	Mean	Std. Dev.	Observations
A. Pollution outcomes			
SO_2, ppb	8.96	12.62	2,032,338
NO_2, ppb	21.25	15.60	792,784
CO, ppm	1.64	1.37	848,067
Ozone, ppb	25.53	13.69	669,261
TSP, $\mu g/m^3$	63.11	40.19	628,932
B. One-day mortality rate outcomes			
All-cause mortality, deaths per million	24.70	24.32	2,032,338
Cardiovascular	12.21	16.04	2,032,338
Cancer	5.15	9.16	2,032,338
Other	5.45	10.02	2,032,338
External	1.89	7.99	2,032,338

Empirical Analysis

Short-run effects of acute exposure

Empirical strategy: instrumental variables (2SLS)

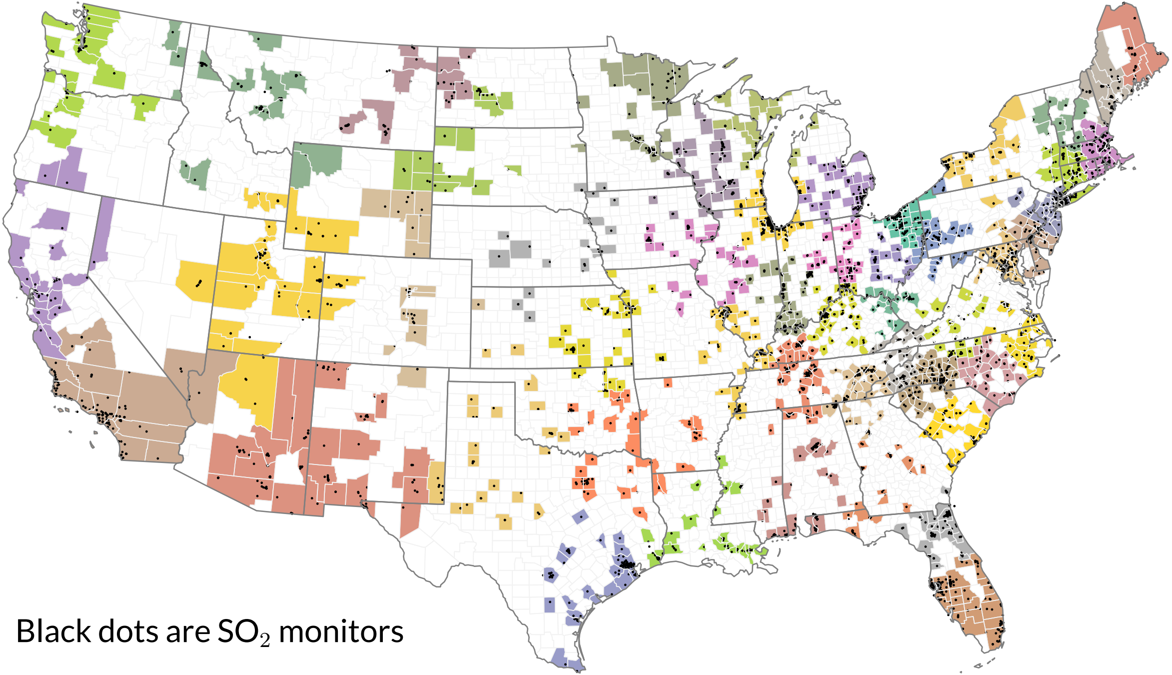
- Wind carries pollutants over long distances

- Key insight: no need to isolate the pollution source! (Deryugina et al. 2019)
  - Maximizes the size of our estimation sample

- Identifying assumption:
  - Wind direction unrelated to health except through pollution

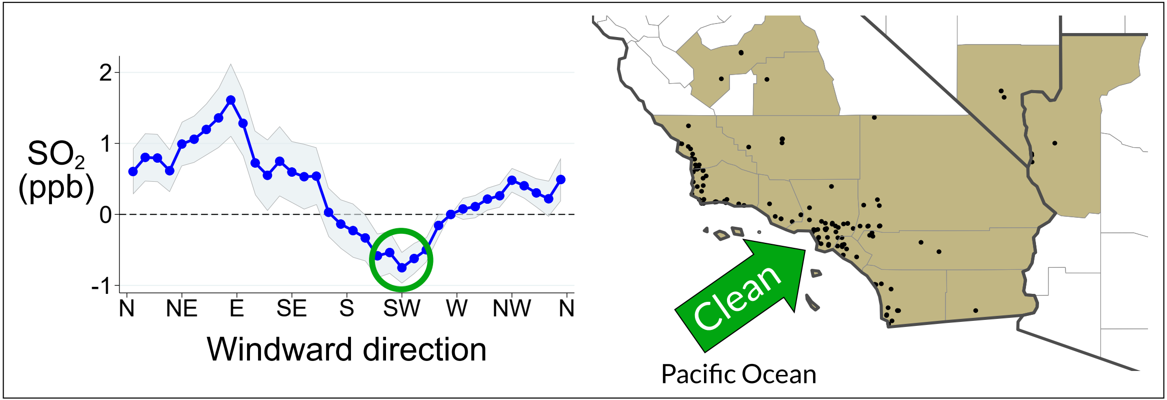
How do we construct our instruments?

- Use clustering algorithm to assign pollution monitors to 50 regional groups
- First stage is group-specific relationship between wind direction and pollution
- Allow pollution transport patterns to vary across groups
  - Wind blowing from west has different effect in California than in Massachusetts



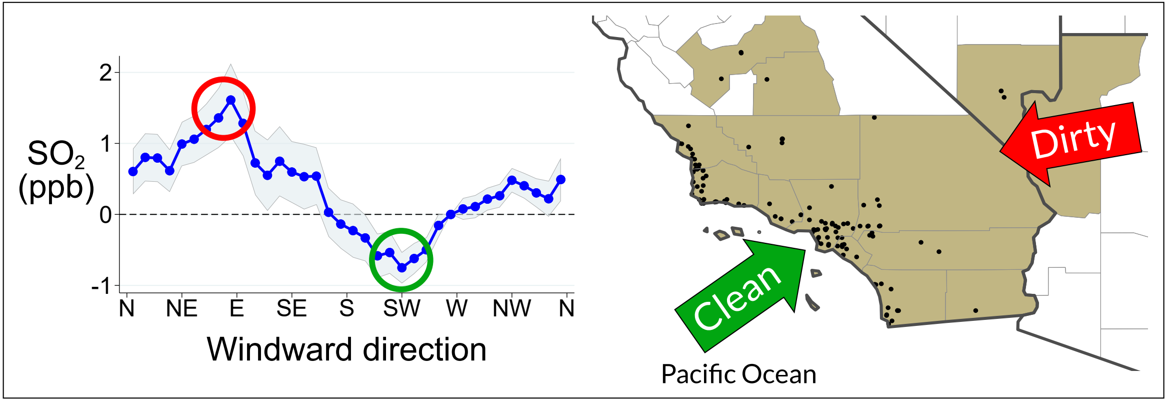
Black dots are SO_2 monitors

Wind direction and SO_2 in Southern California area



Blue shading depicts 95% confidence intervals
Black dots on map are SO_2 monitors

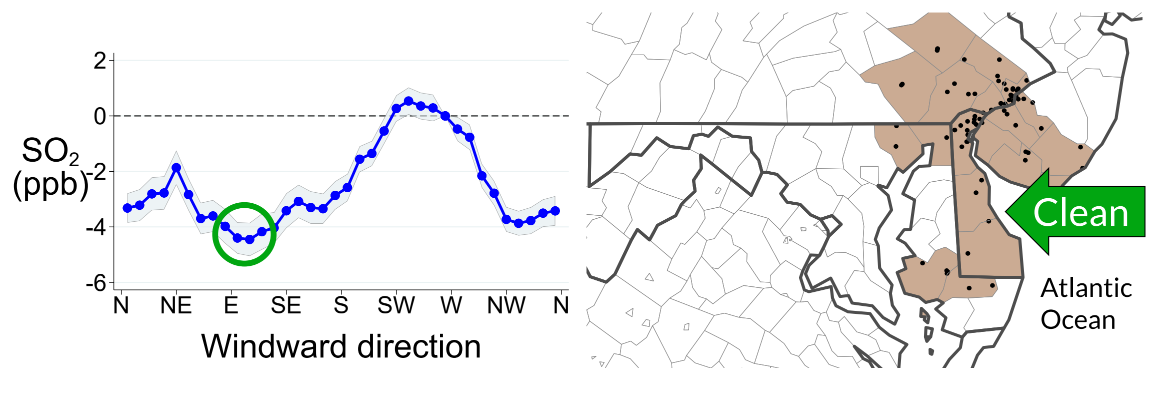
Wind direction and SO_2 in Southern California area



Blue shading depicts 95% confidence intervals

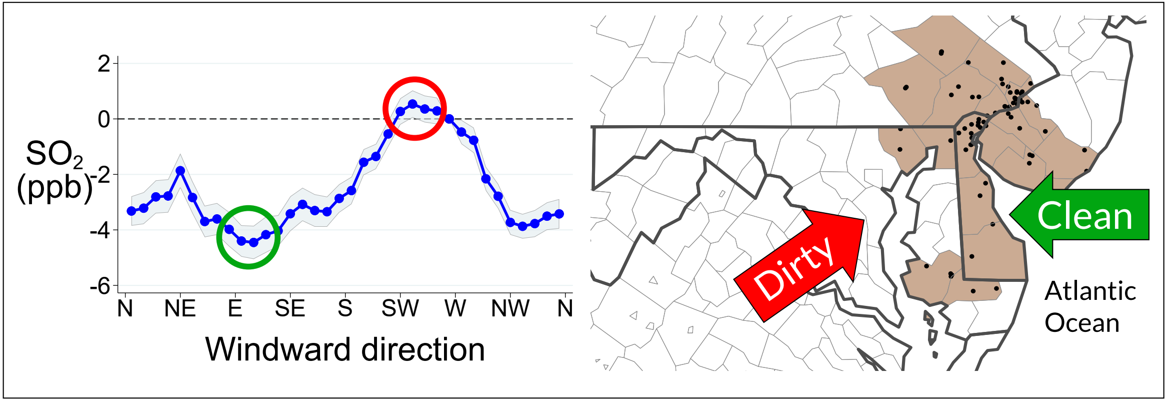
Black dots on map are SO_2 monitors

Wind direction and SO_2 in Greater Philadelphia area



Blue shading depicts 95% confidence intervals
Black dots on map are SO_2 monitors

Wind direction and SO_2 in Greater Philadelphia area



Blue shading depicts 95% confidence intervals
Black dots on map are SO_2 monitors

First stage: excluded instrument is wind direction

$$\text{SO2}_{cd} = \sum_{g=1}^{50} \boldsymbol{f^g(\theta_{cd})} + X_{cd}^{k\,\prime}\delta + \alpha_{cm} + \alpha_{my} + \varepsilon_{cd}$$

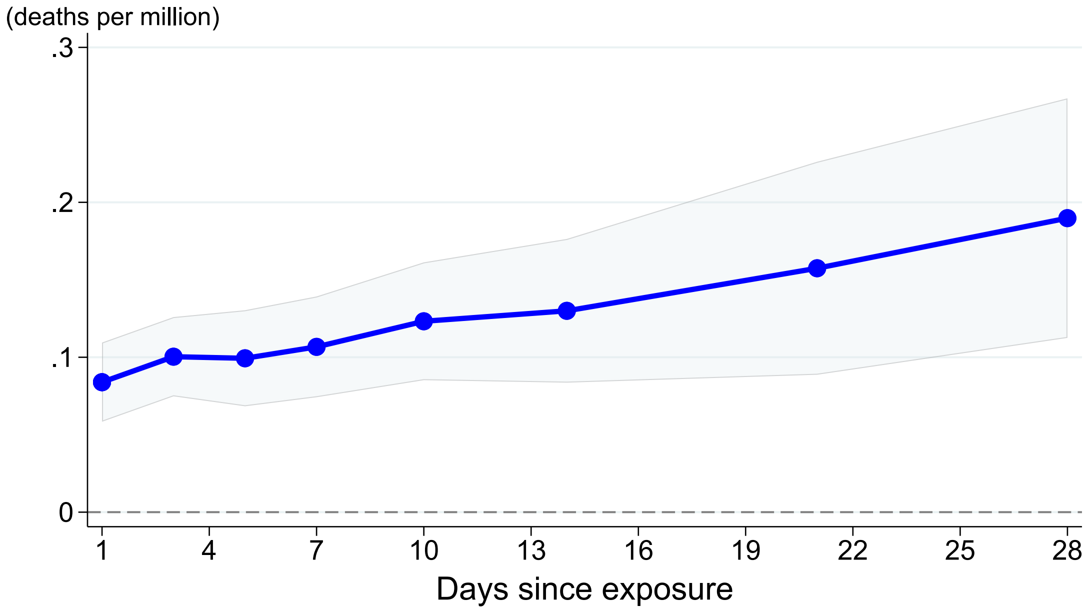
- Dependent variable is level of SO_2 in county c on day d
- Effect of wind direction, θ_{cd}, varies across 50 geographic groups, g
- Consider two functional forms for $\boldsymbol{f^g(\theta_{cd})}$
  - Non-parametric 10-degree bins (1750 instruments)
  - Parametric sin function (100 instruments, preferred specification) ▸ Example

Second-stage regression

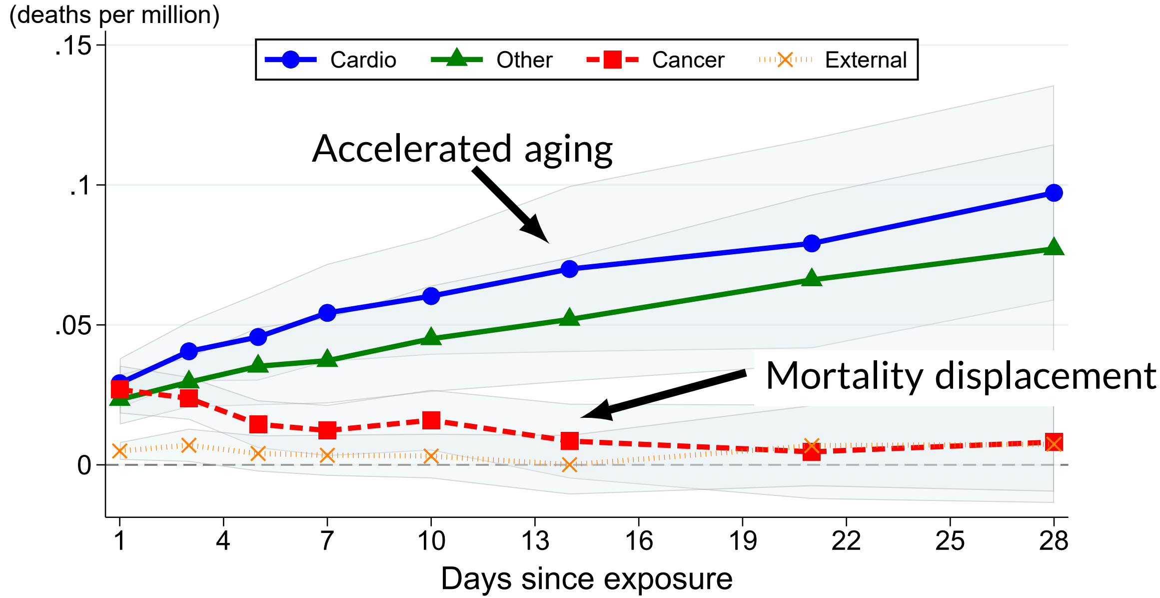
$$Y_{cd}^{k} = \beta^{k} \widehat{\text{SO2}}_{cd} + X_{cd}^{k\,\prime}\delta + \alpha_{cm} + \alpha_{my} + \varepsilon_{cd}$$

- Estimate effect of 1-day exposure on k-day mortality rate (up to $k = 28$)
- Control for county-by-month (α_{cm}) and month-by-year (α_{my}) fixed effects
- Flexibly control for max temperature, precipitation, and wind speed
- Cluster standard errors at the county level, weight by county population

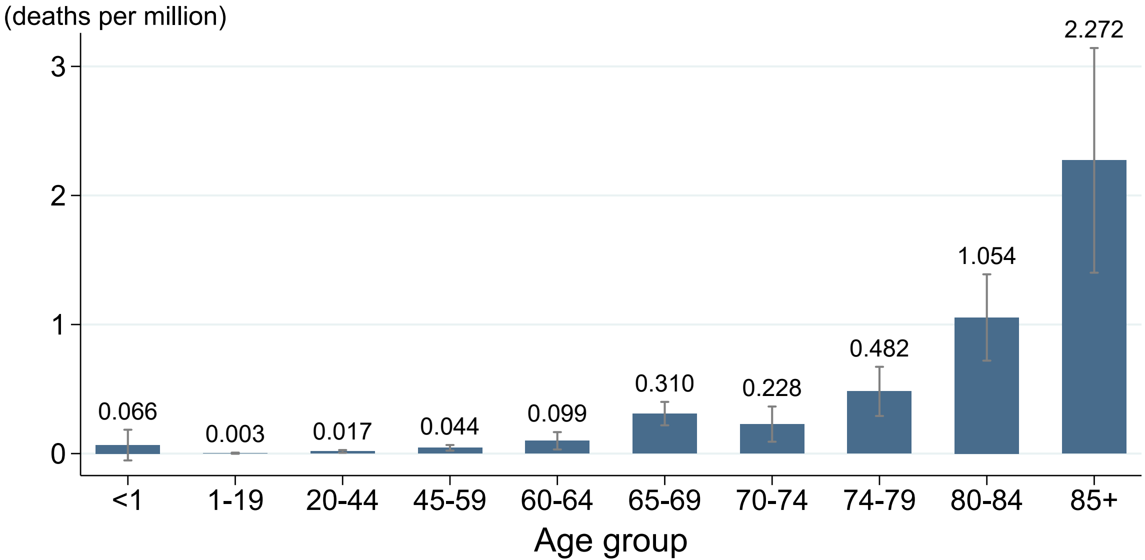
Cumulative mortality effect grows over time



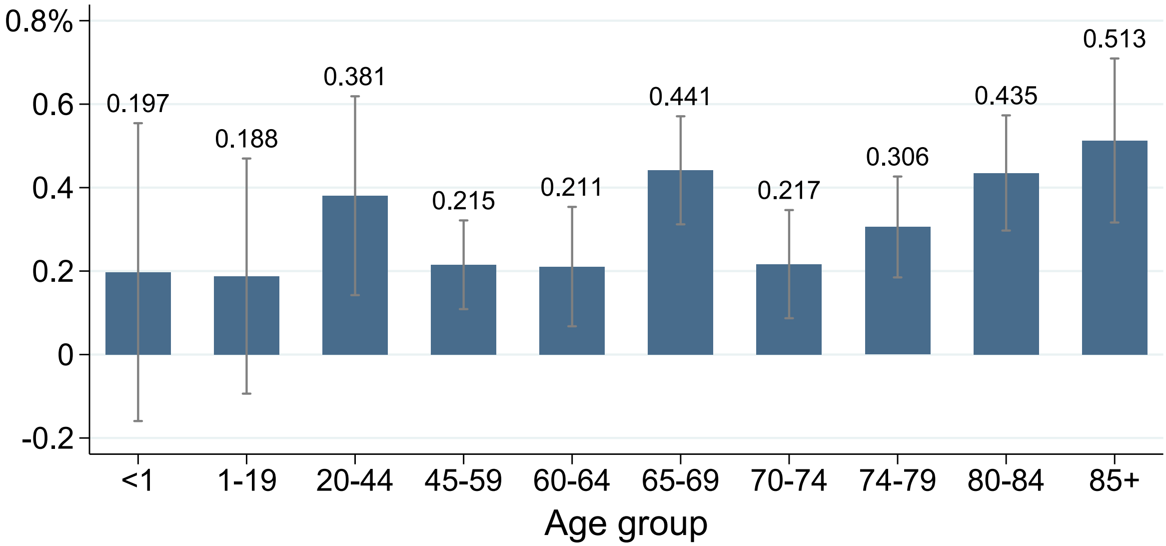
Divergent patterns by cause of death



1-day mortality by age group (deaths per million)



1-day mortality by age group (relative effect)



Alternative specifications and robustness checks

- Accounting for other air pollutants ▸ Table
- Sensitivity check: alternative weather controls ▸ Table
- Falsification test: SO_2 on day t has no effect on mortality on day $t-1$ ▸ Table
- Placebo test: random wind direction produce weak first stage ($F \leq 2$) ▸ Table

Long-run Survival

Model: Lleras-Muney and Moreau (2022)

Health capital for individual i at age t:

$$H_{it} = H_{i,t-1} - \underbrace{\delta t^{\alpha}}_{\text{depreciation}} + I + \varepsilon_{it}$$

where:

$$H_{i0} = H_{i0}^{*} \sim N(\mu_H, 1)$$
$$\varepsilon_{it} \sim N(0, \sigma_{\varepsilon}^{2})$$

Model: Lleras-Muney and Moreau (2022)

$$H_{it} = H_{i,t-1} - \delta t^{\alpha} + I + \varepsilon_{it}$$

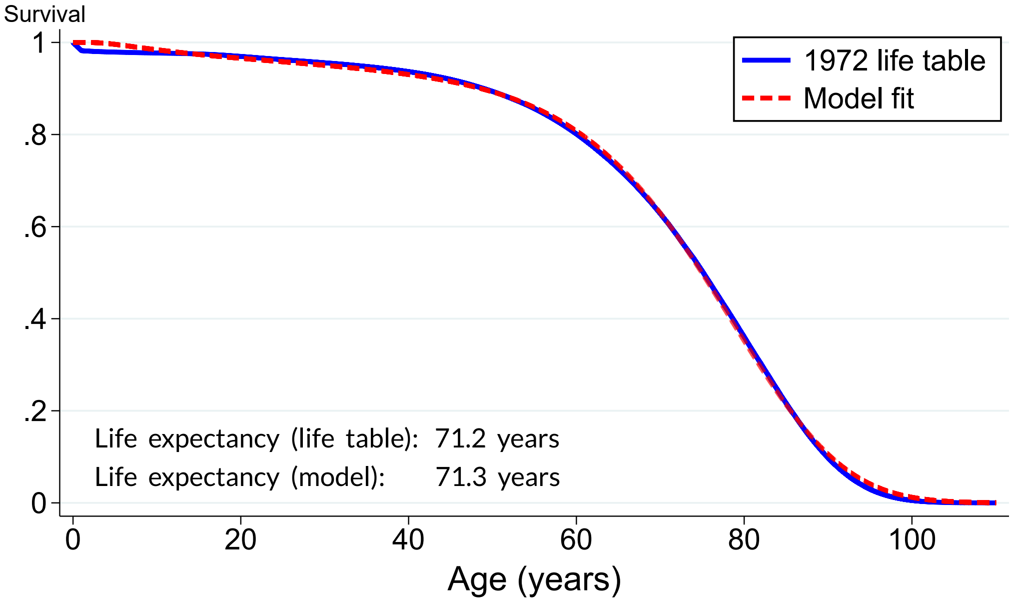
- Death occurs when health capital falls below threshold $\underline{H} = 0$:

$$D_{i0} = 1\left[H_{i0} < \underline{H}\right],$$
$$D_{it} = 1\left[H_{it} < \underline{H} \middle| D_{i,t-1} = 0\right],\ t > 0$$

- Simulate model for N agents $\rightarrow$ survival curve

- Model captures a variety of real-world mortality dynamics
  - Mortality displacement
  - Accelerated aging

Calibrate baseline parameters using 1972 period life table



Key structural assumption for incorporating IV estimates

- Effect of pollution on model parameters depends only on current exposure
  - Effect on parameters is same for old and young
  - Effect on parameters is independent of exposure history

- Thus, we can calibrate the effect of exposure using any age group

- Testable implication: calibration from one age predicts survival for other ages

Calibrate using 1-day IV estimates

$$H_{it} = H_{i,t-1} - \delta t^{\alpha} + I + \varepsilon_{it}$$
$$D_{it} = 1\left[H_{it} < \underline{\text{H}} \,\middle|\, D_{i,t-1} = 0\right],\ t > 0$$

Acute exposure affects mortality through two channels:

1. Raises depreciation for 1 day, $\delta \rightarrow \widetilde{\delta}$
   - accelerated aging effect
   - calibrate using 1-day non-cancer IV estimate

2. Raises death threshold for 1 day, $\underline{\text{H}} \rightarrow \widetilde{\underline{\text{H}}}$
   - mortality displacement
   - calibrate using 1-day cancer IV estimate

Calibration steps for age group a

1. Solve for $\underline{\widetilde{\mathrm{H}}}_a$ such that 1-day mortality increases by $\widehat{\beta}^1_{a,cancer}$

2. Solve for $\widetilde{\delta}_a$ such that 1-day mortality effect of $\{\underline{\widetilde{\mathrm{H}}}_a, \widetilde{\delta}_a\}$ equals $\widehat{\beta}^1_{a,all}$

Do calibration for older age groups only (65 and over)

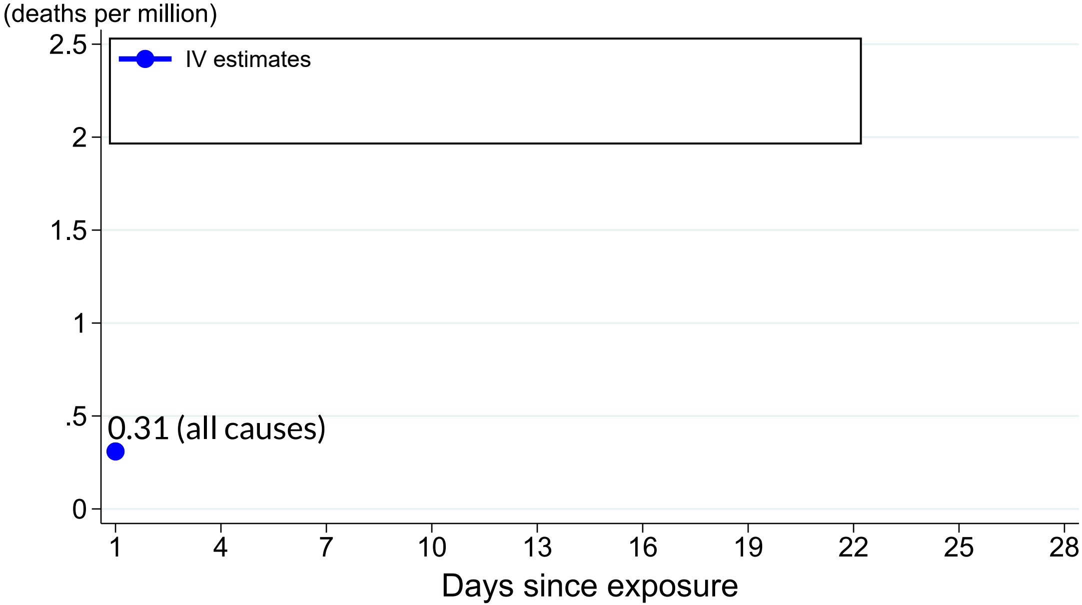
Any pair $\{\underline{\widetilde{\mathrm{H}}}_a, \widetilde{\delta}_a\}$ can be used for predictions
$\rightarrow$ Preferred estimate uses average of all older age groups

Example: ages 65–69

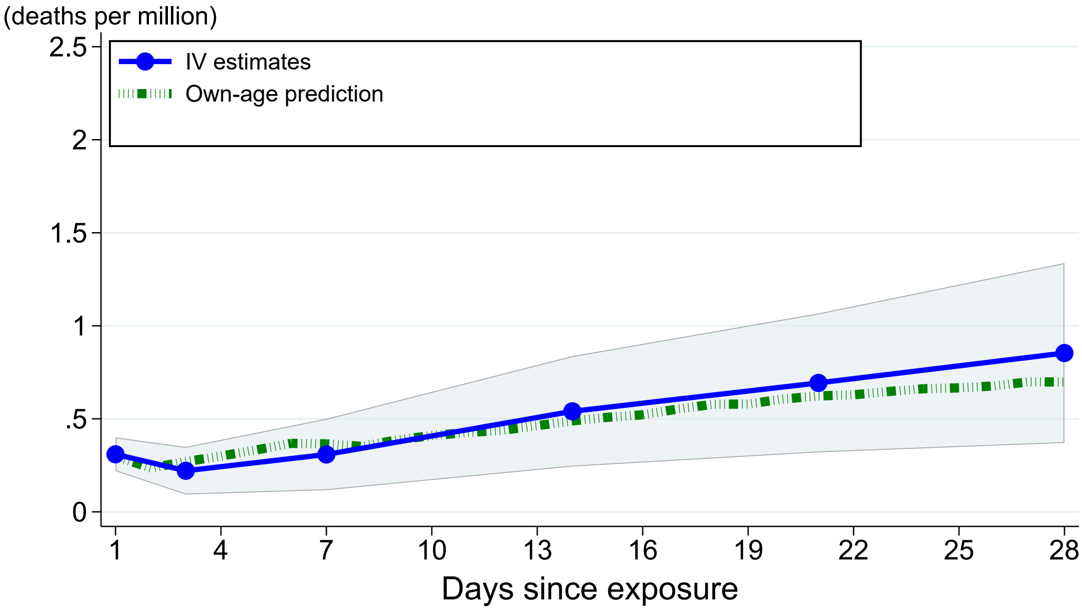
	(1)	(2)
Age group	All causes	Cancer-related causes
65–69	0.31**	0.17**
	(0.046)	(0.028)
70–74	0.23**	0.14**
	(0.070)	(0.034)
75–79	0.48**	0.13**
	(0.097)	(0.040)
80–84	1.1**	0.18**
	(0.17)	(0.065)
85+	2.3**	0.17*
	(0.44)	(0.084)

Notes: Dependent variable is deaths per million on the day of exposure.

Example: ages 65–69



Example: ages 65–69

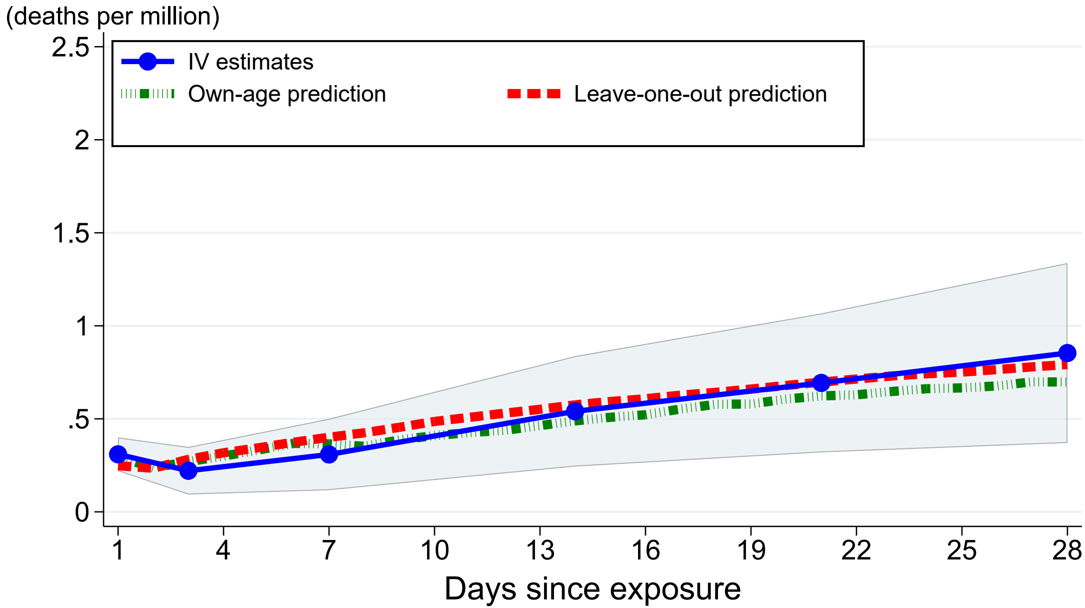


“Leave-one-out” validation: calibrate using other ages

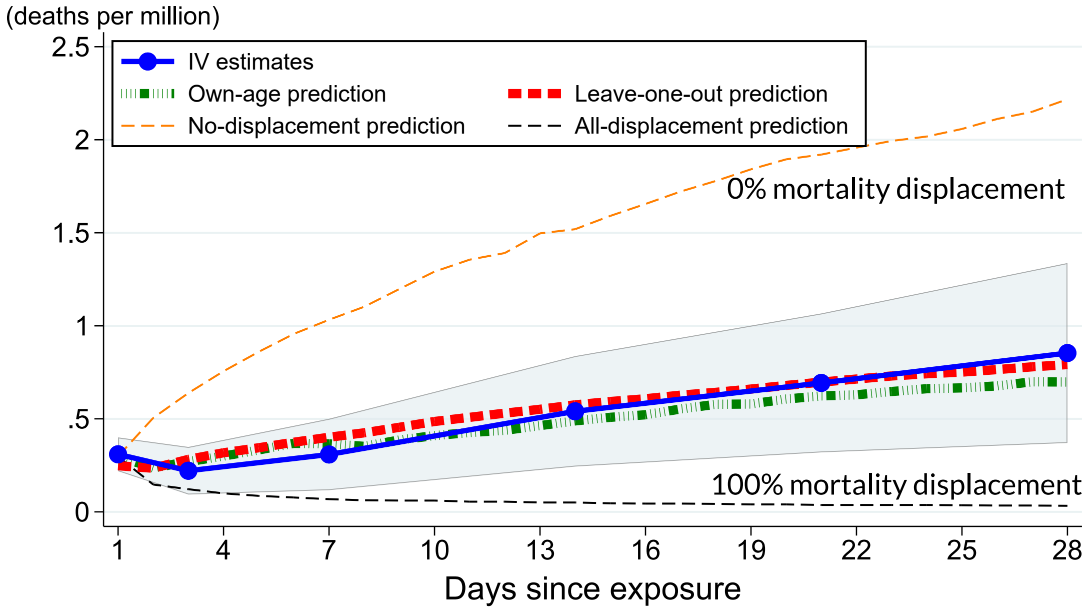
	(1)	(2)
Age group	All causes	Cancer-related causes
65–69	0.31**	0.17**
	(0.046)	(0.028)
70–74	0.23**	0.14**
	(0.070)	(0.034)
75–79	0.48**	0.13**
	(0.097)	(0.040)
80–84	1.1**	0.18**
	(0.17)	(0.065)
85+	2.3**	0.17*
	(0.44)	(0.084)

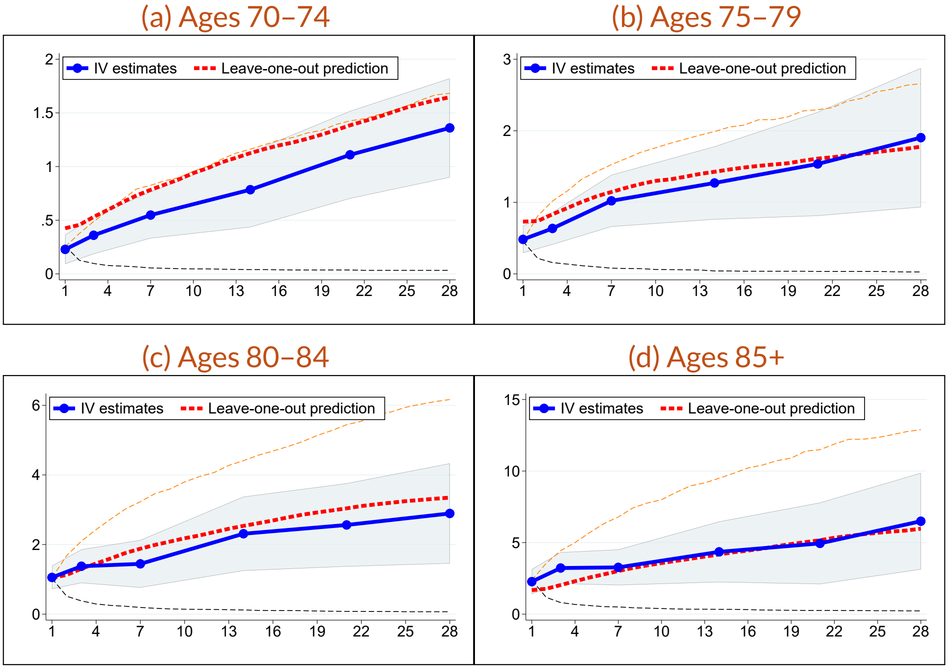
Notes: Dependent variable is deaths per million on the day of exposure.

Example: ages 65–69

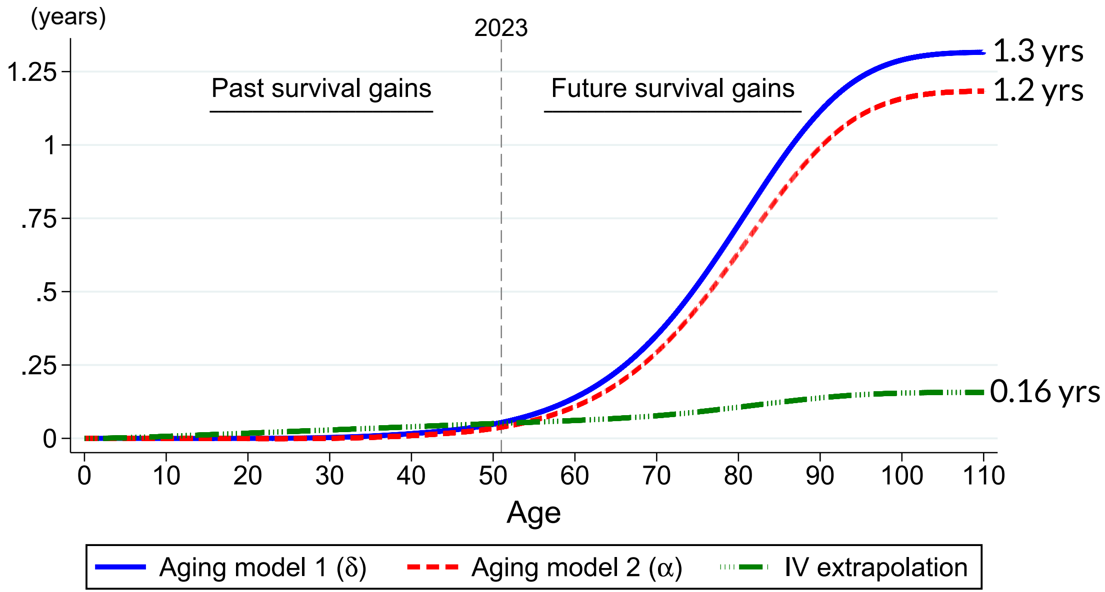


Example: ages 65–69




(a) Ages 70–74
IV estimates
Leave-one-out prediction
0
.5
1
1.5
2
1
4
7
10
13
16
19
22
25
28
(b) Ages 75–79
IV estimates
Leave-one-out prediction
0
1
2
3
1
4
7
10
13
16
19
22
25
28
(c) Ages 80–84
IV estimates
Leave-one-out prediction
0
2
4
6
1
4
7
10
13
16
19
22
25
28
(d) Ages 85+
IV estimates
Leave-one-out prediction
0
5
10
15
1
4
7
10
13
16
19
22
25
28

Survival benefit of 1-unit reduction in chronic exposure



Interpreting long-run survival estimates

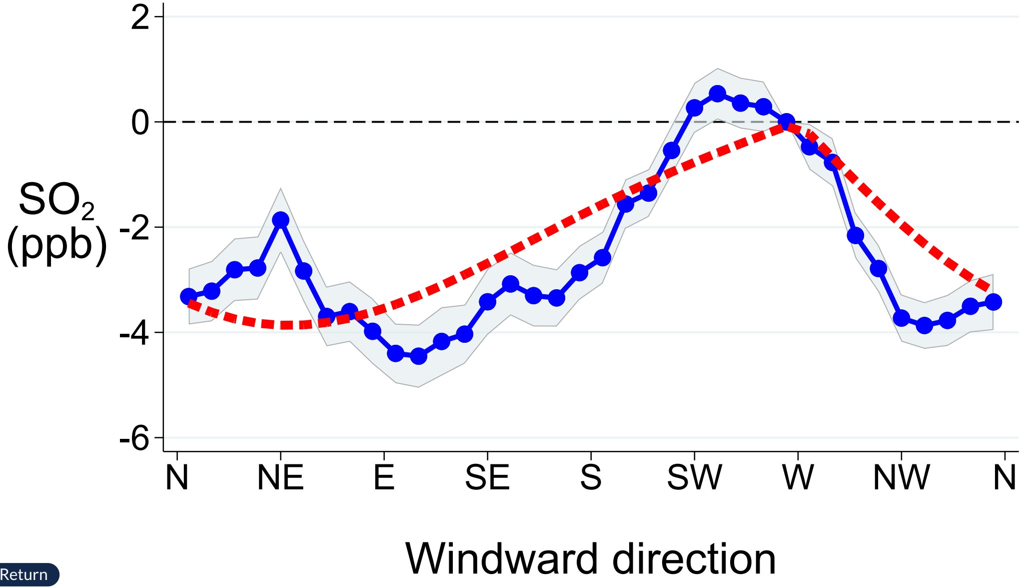
- Uncertainty in IV estimates produces uncertainty in long-run estimates
    - 5th and 95th percentiles from bootstrap yield range of $[0.3, 2.2]$ years
- SO_2 estimates may also include effects from particulate matter
- Survival model holds behavior fixed
    - We interpret estimates as gross benefits (Graff Zivin and Neidell 2012; Currie et al. 2014)

Conclusion

- Air pollution causes mortality displacement and accelerating aging
- Permanent, 10% reduction in exposure improves life expectancy by 1.2 yrs
  - 7 times larger than extrapolation of short-run estimate
  - Benefits concentrated in ages 50+

The End

First stage: parametric sin fit for Greater Philadelphia area



◂ Return

Sensitivity check: alternative weather controls

	(1)	(2)	(3)	(4)
SO_2, parts per billion	0.098**	0.084**	0.084**	0.085**
	(0.014)	(0.013)	(0.013)	(0.012)
First-stage F-statistic	32	42	68	33
Mean outcome	25	25	25	25
Sample size	2,032,340	2,032,338	2,032,272	2,031,752
Weather controls				
Baseline weather variables		X	X	X
Minimum temperature variables			X	X
More granular bins				X

Notes: Dependent variable is 1-day mortality (deaths per million).

Return

IV estimates: accounting for multiple air pollutants (1/2)

	(1)	(2)	(3)	(4)	(5)	(6)
SO_2, ppb	0.084**	0.060**	0.065**	0.066**	0.059**	0.064**
	(0.012)	(0.013)	(0.014)	(0.012)	(0.012)	(0.014)
TSP, μg/m^3		0.012**	0.014**	0.014**	0.013**	0.015**
		(0.0036)	(0.0037)	(0.0033)	(0.0040)	(0.0035)
NO_2, ppb			−0.014			0.0023
			(0.013)			(0.017)
Ozone, ppb				−0.044*		−0.046*
				(0.021)		(0.022)
CO, ppm					−0.20	−0.24
					(0.17)	(0.20)
First-stage F-statistic	81	21	17	11	20	10
Mean outcome	27	27	27	27	27	27
Sample size	78,946	78,946	78,946	78,946	78,946	78,946

Notes: Dependent variable is 1-day mortality (deaths per million).

IV estimates: accounting for multiple air pollutants (2/2)

	(1)	(2)
SO_2, ppb	0.079**	0.035*
	(0.014)	(0.015)
TSP, μg/m^3		0.019**
		(0.0045)
First-stage F-statistic	96	50
Mean outcome	25	25
Sample size	627,304	627,304

Notes: Dependent variable is 1-day mortality (deaths per million). A */** indicates significance at the 5%/1% level. "TSP" is total suspended particulates.

Return

Placebo and falsification tests

	(1)	(2)	(3)	(4)
SO_2, ppb	–0.079	0.18	–0.041	
	(0.062)	(0.23)	(0.49)	
SO_2 on day $t+1$, ppb				–0.0036
				(0.0048)
Outcome window, days	1	7	28	1
First-stage F-statistic	2.0	1.9	1.9	28
Mean outcome	25	173	691	25
Sample size	2,023,456	2,023,435	2,023,369	2,031,165
Placebo test	X	X	X	
Falsification test				X

Notes: Dependent variable is number of deaths per million people over a window of 1, 7, or 28 days.

Return