Geographic and Socioeconomic Heterogeneity in the Benefits of Reducing Air Pollution in the United States

TATYANA DERYUGINA,
NOLAN MILLER,
DAVID MOLITOR,
JULIAN REIF

Air pollution is costly for health, even in the modern United States

- Acute pollution exposure is harmful to health even in low-pollution areas (e.g., Ward 2015; Knittel et al. 2016; Schlenker and Walker 2016; Deryugina et al. 2019)
  - There may be substantial social benefits to further reducing US air pollution

- But, additional emissions reductions may require increasingly costly measures

- Crucial to understand where such reductions would be most beneficial

How should pollution reduction efforts be targeted?

- Typically: Based on current pollution levels
  - e.g., target areas with high fine particulate matter ($PM_{2.5}$) concentrations
  - But what about based on where individuals vulnerable to $PM_{2.5}$ reside?

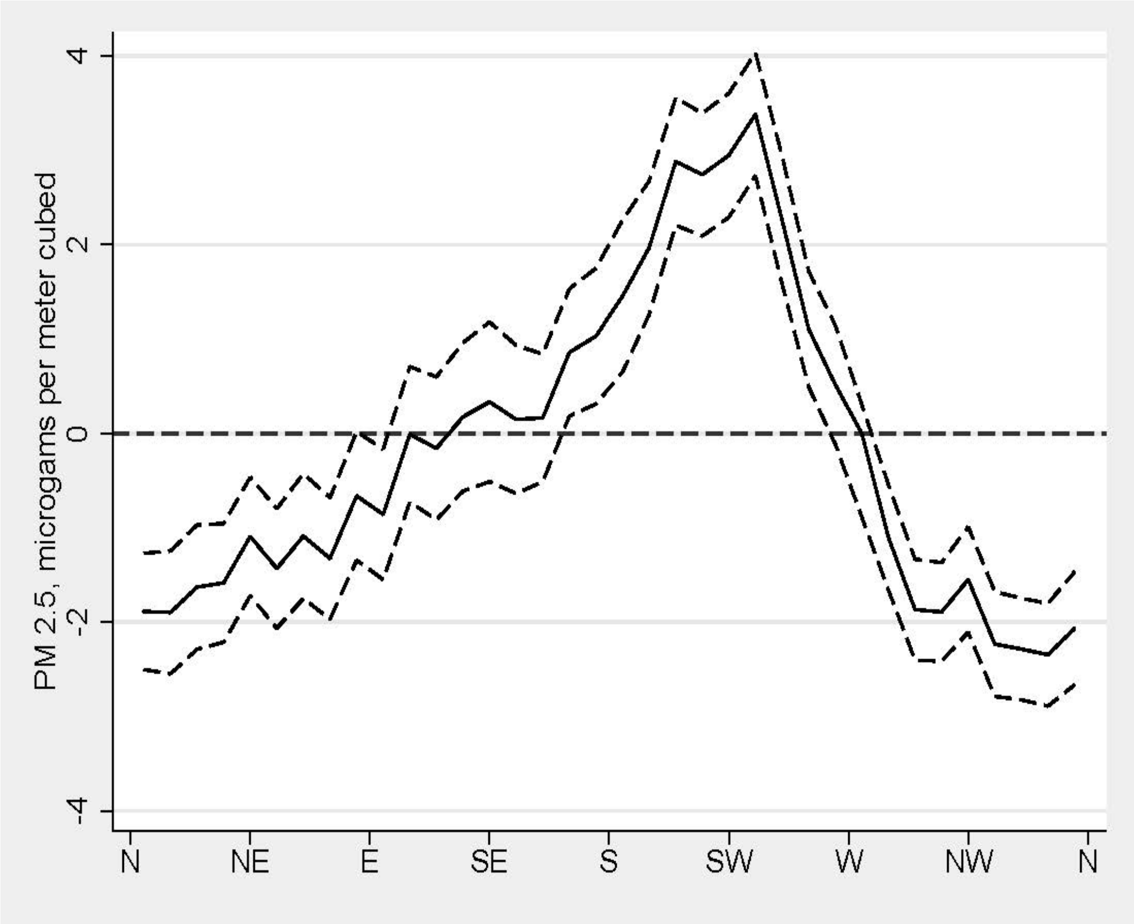
- <u>Q1: How closely related are $PM_{2.5}$ and vulnerability</u>?
  - A1: in our sample of elderly Medicare beneficiaries, they are (slightly) negatively related

- <u>Q2: What kind of areas have more vulnerable beneficiaries</u>?
  - A2: counties that are lower-income, less urban, hotter, and have worse health behaviors

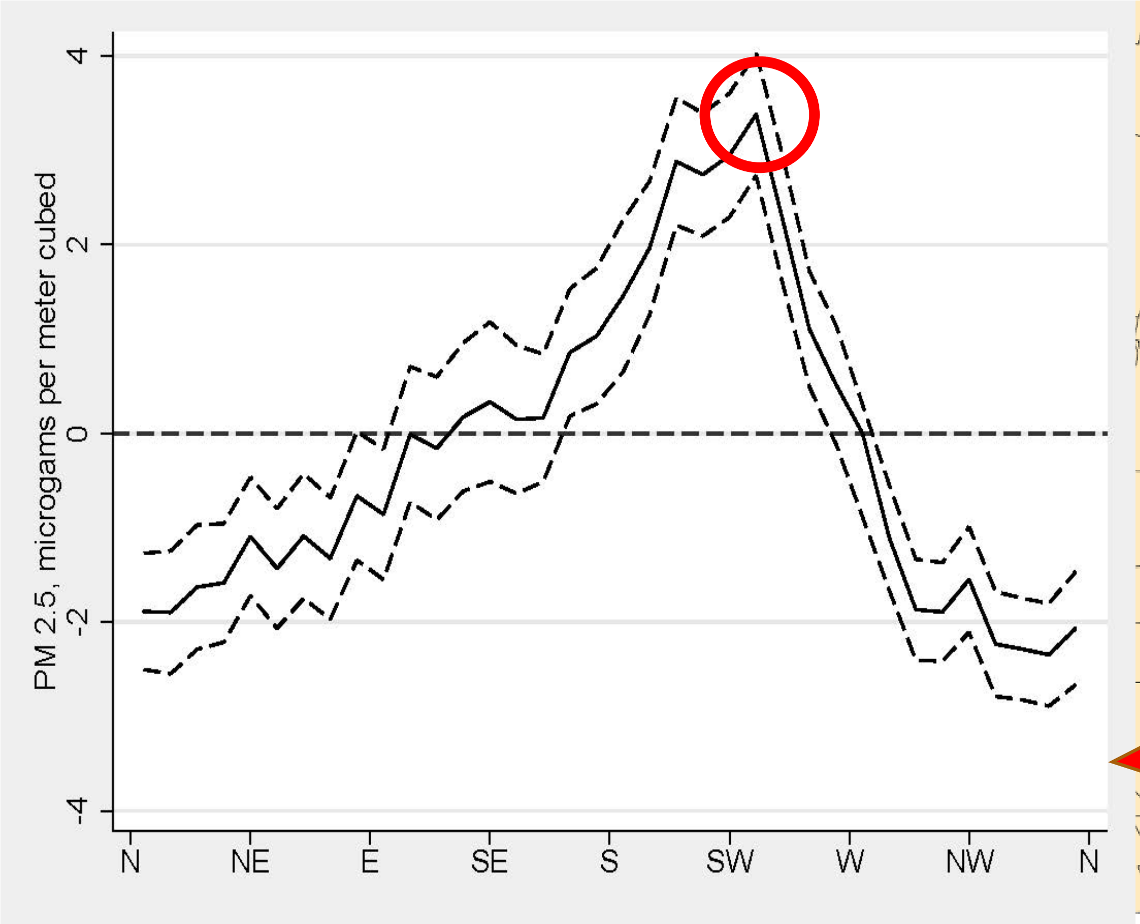
We build on Deryugina, Heutel, Miller, Molitor, and Reif (2019) [DHMMR]

1. Use daily variation in local wind direction to isolate fluctuations in $PM_{2.5}$ that are as good as random
   - Addresses concern about confounding factors (e.g., traffic) and alleviates measurement error issues
2. Classify county-days as “high-pollution” or “low-pollution” based on wind direction and how that wind direction affects pollution

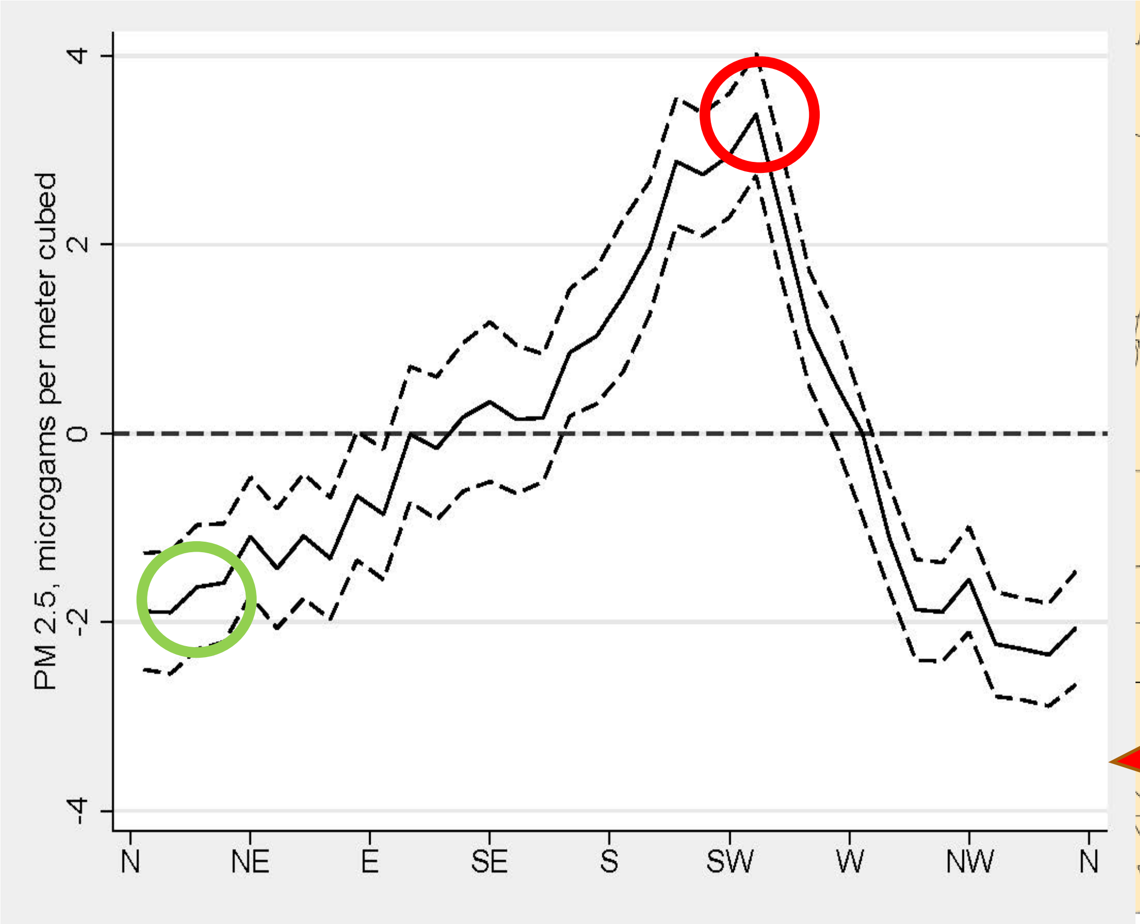
Wind direction and $PM_{2.5}$ in the Boston area

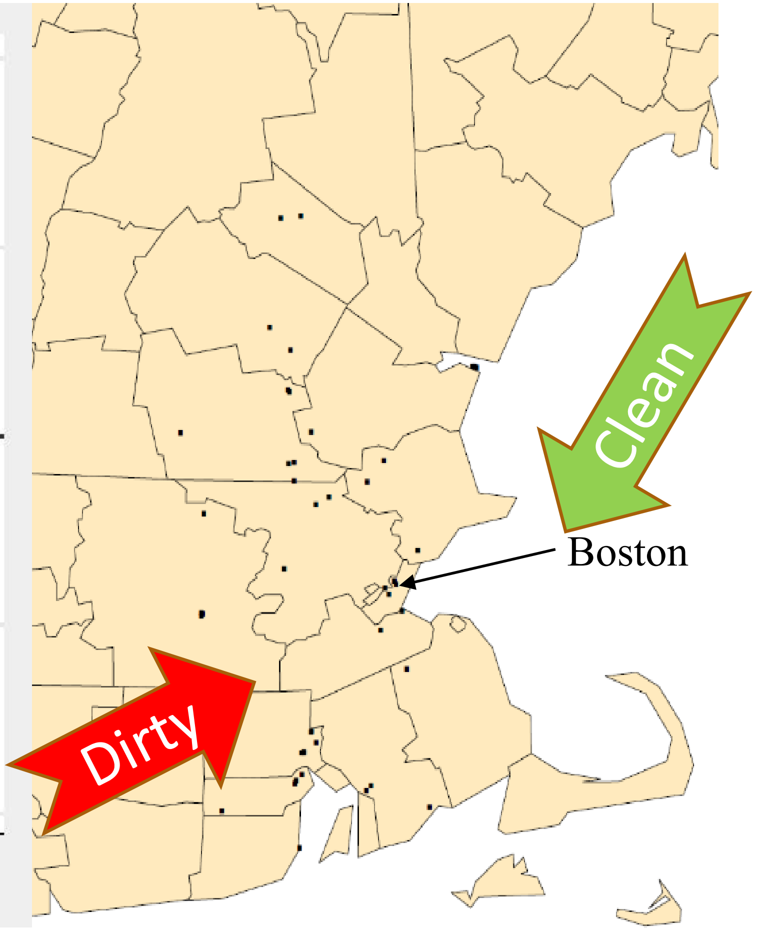


Wind direction and $PM_{2.5}$ in the Boston area



Wind direction and $PM_{2.5}$ in the Boston area

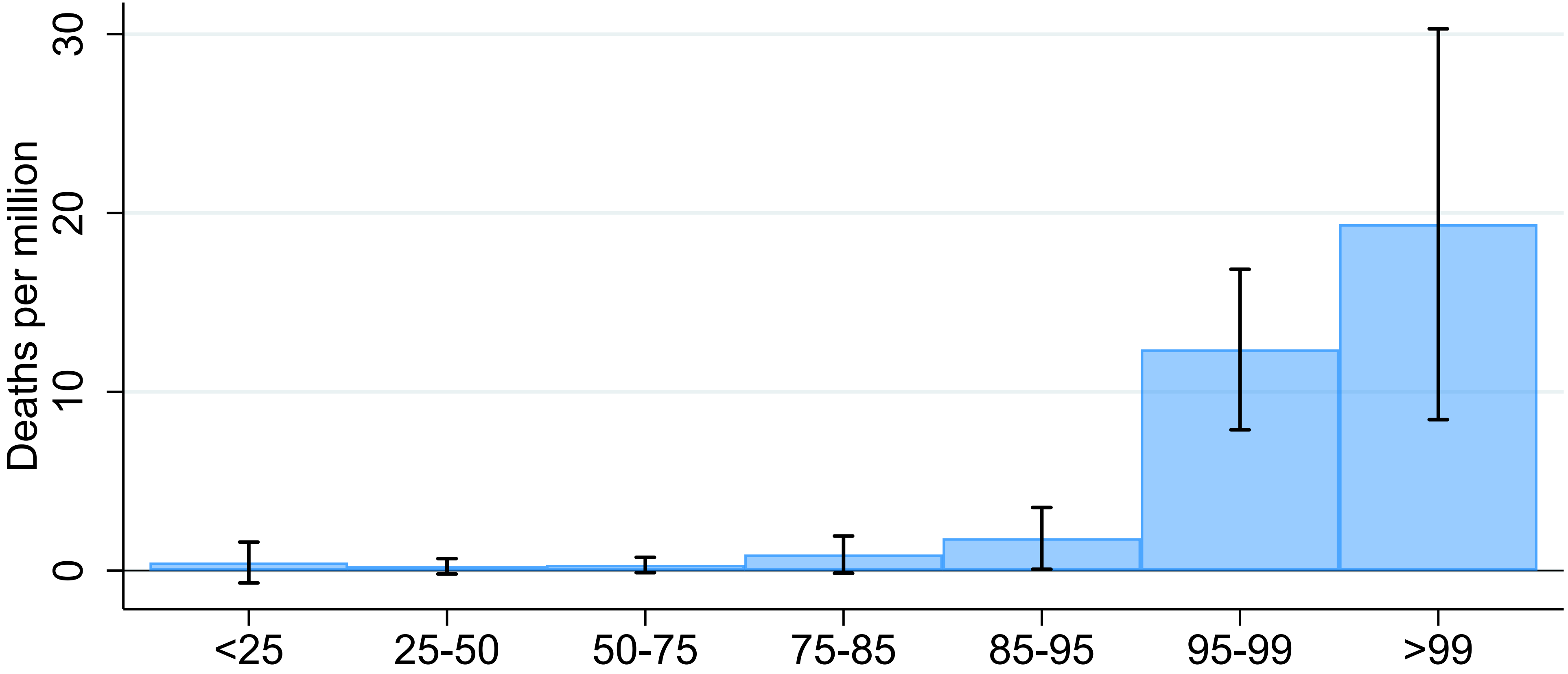




Predicting vulnerability

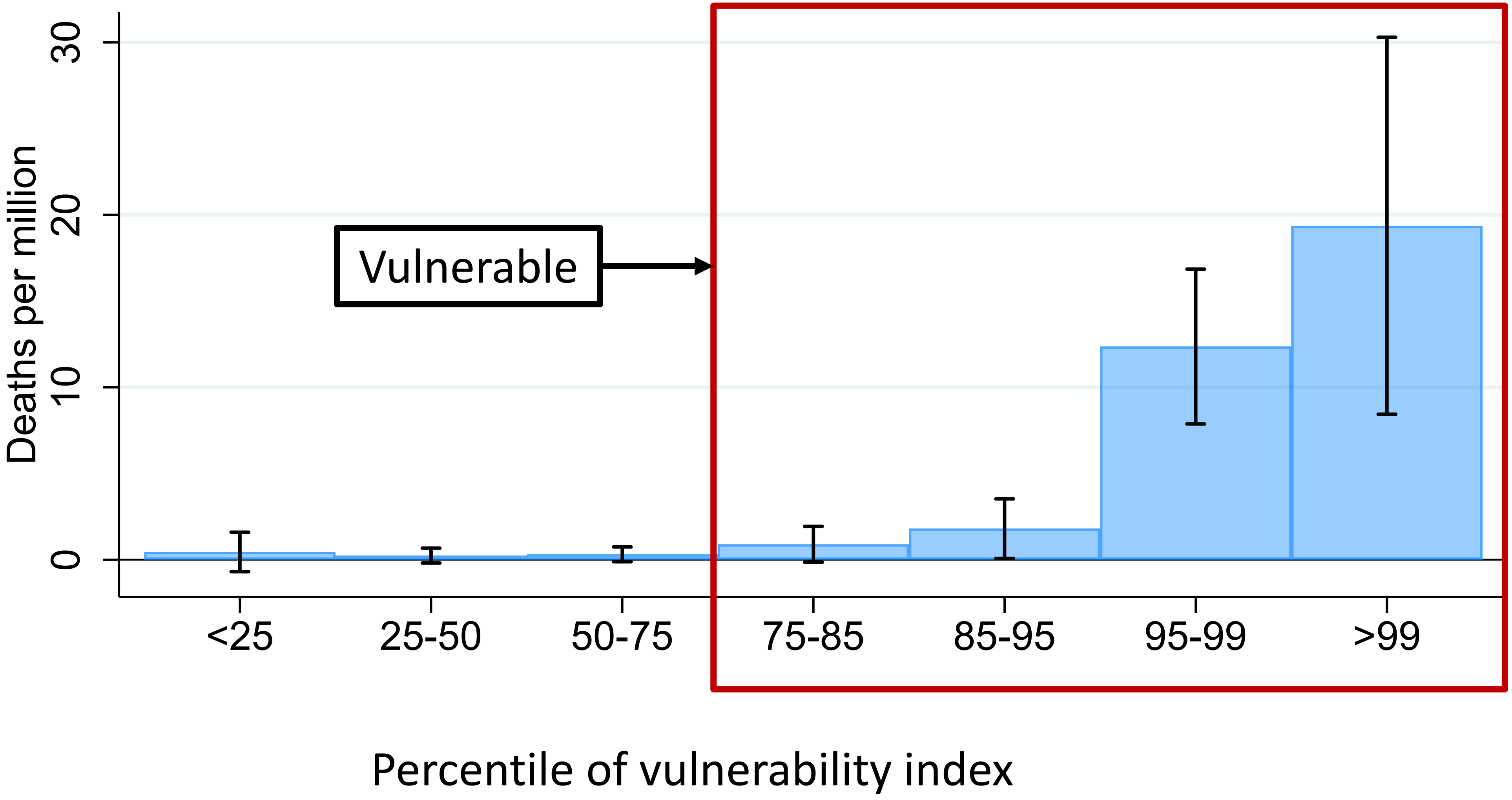
1. Construct sample of 23.6 million US elderly (2013 Medicare enrollees with sufficient data on health histories)
2. Train machine learning algorithm to separately predict elderly mortality on "high-pollution" and "low-pollution" days as function of many individual and local characteristics
   - e.g., local unemployment rate, income, presence of 27 chronic conditions, history of medical spending and medical events
3. Obtain differences in predictions at the individual level
   - Difference in the probability of dying on high-pollution versus low-pollution day is the individual's vulnerability measure

Effect of acute $PM_{2.5}$ exposure as function of vulnerability index



Percentile of vulnerability index

Effect of acute $PM_{2.5}$ exposure as function of vulnerability index



Who are the vulnerable?

Outcome	(1) Bottom 75%	(2) Top 25%	(3) Difference
Demographics			
Age (years)	75.3	79.7	4.42***
			(0.00435)
Male	0.421	0.461	0.0408***
			(0.000294)
Chronic conditions			
Alzheimer's or dementia	0.0946	0.239	0.144***
			(0.000197)
Chronic kidney disease	0.176	0.343	0.166***
			(0.000241)
COPD	0.197	0.386	0.189***
			(0.00025)
Heart failure	0.192	0.412	0.219***
			(0.00025)
Lung cancer	0.00965	0.0251	0.0155***
			(0.0000685)

Who are the vulnerable?

Outcome	(1) Bottom 75%	(2) Top 25%	(3) Difference
Demographics			
Age (years)	75.3	79.7	4.42***
			(0.00435)
Male	0.421	0.461	0.0408***
			(0.000294)
Chronic conditions			
Alzheimer's or dementia	0.0946	0.239	0.144***
			(0.000197)
Chronic kidney disease	0.176	0.343	0.166***
			(0.000241)
COPD	0.197	0.386	[illegible]
			[illegible]
Heart failure	0.192	0.412	0.219***
			(0.00025)
Lung cancer	0.00965	0.0251	0.0155***
			(0.0000685)

Not vulnerable

Vulnerable

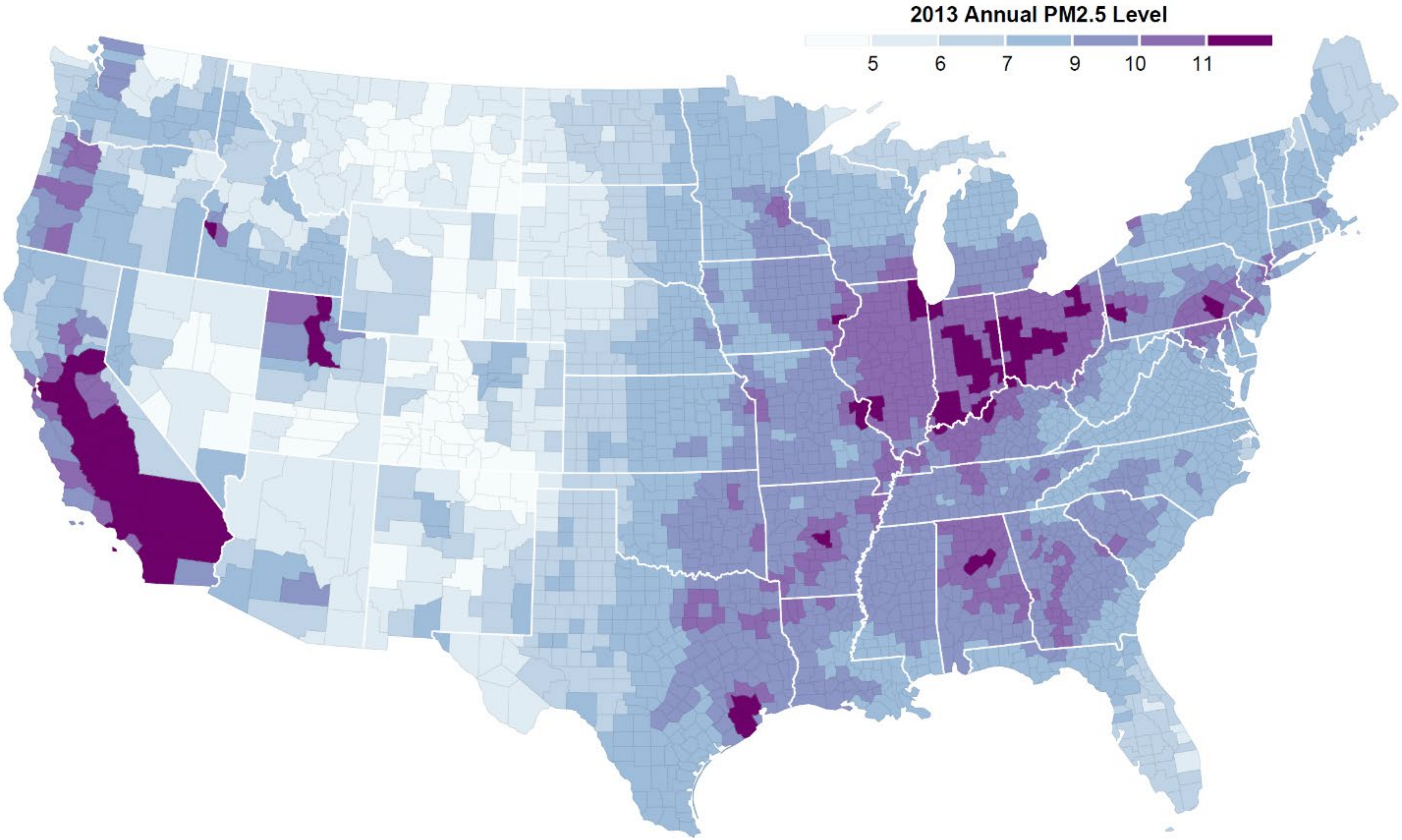
Who are the vulnerable?

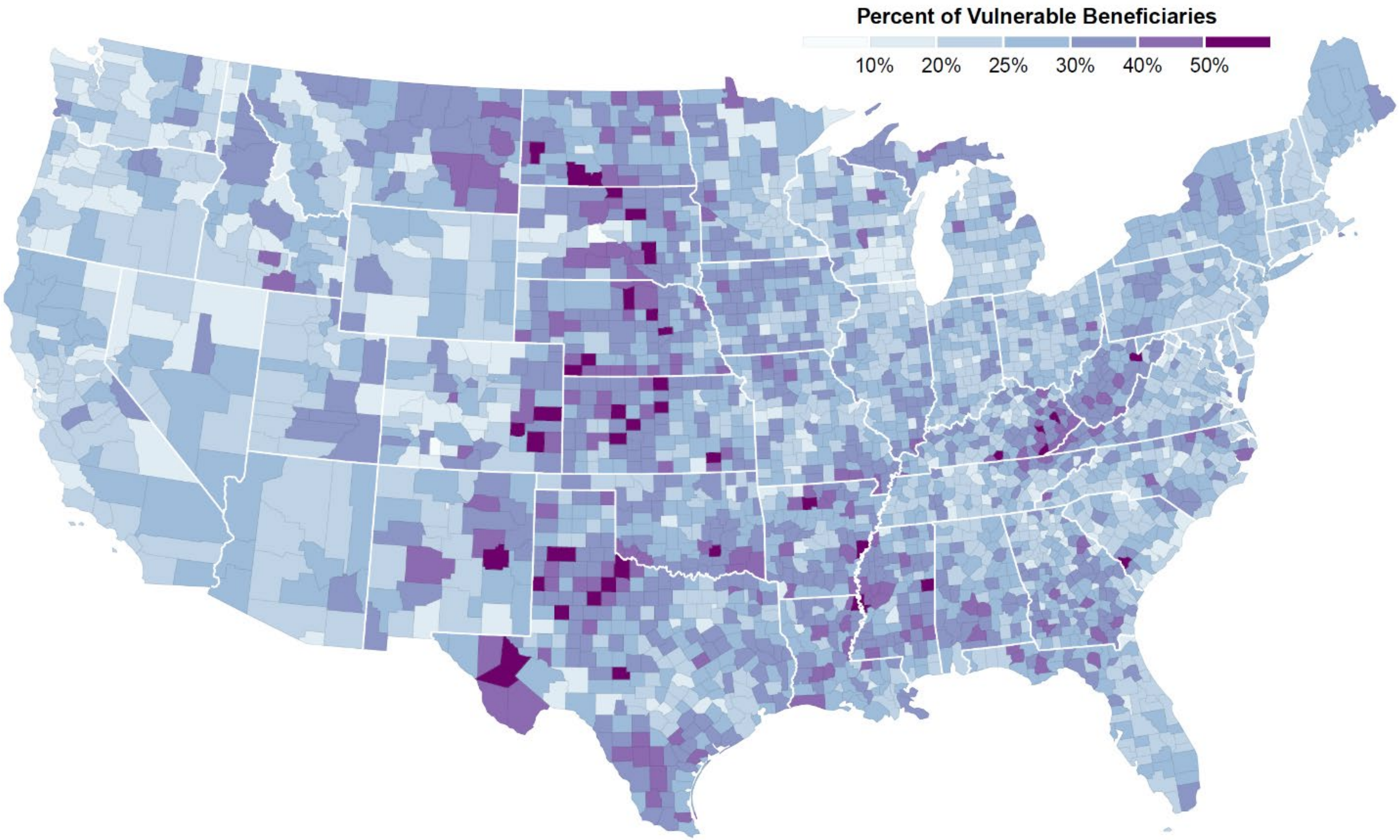
Outcome	(1) Bottom 75%	(2) Top 25%	(3) Difference
Demographics			
Age (years)	75.3	79.7	4.42***
			(0.00435)
Male	0.421	0.461	0.0408***
			(0.000294)
Chronic conditions			
Alzheimer's or dementia	0.0946	0.239	0.144***
			(0.000197)
Chronic kidney disease	0.176	0.343	0.166***
			(0.000241)
COPD	0.197	0.386	0.189***
			(0.00025)
Heart failure	0.192	0.412	0.219***
			(0.00025)
Lung cancer	0.00965	0.0251	0.0155***
			(0.0000685)

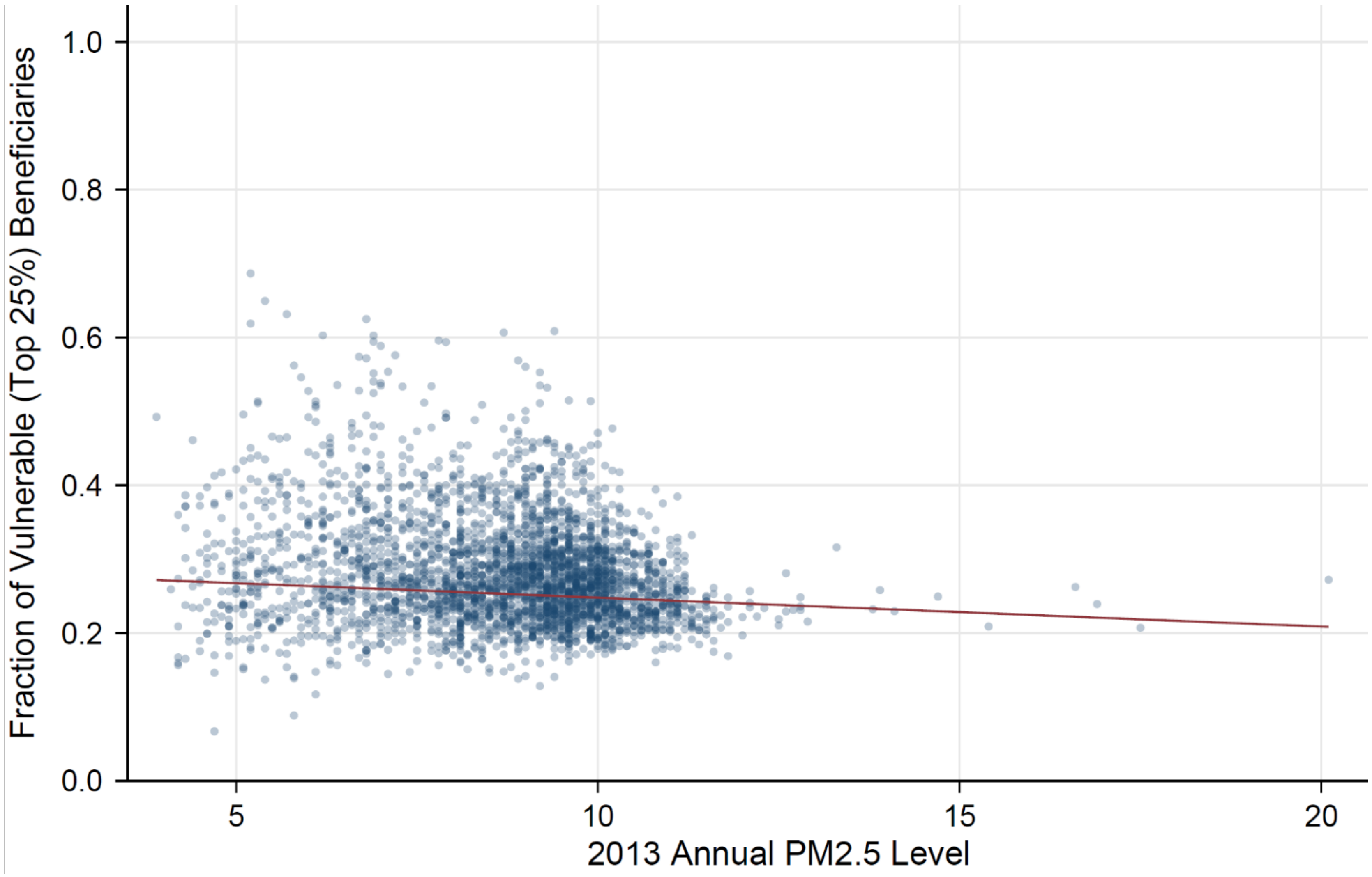
Differences in characteristics

Who are the vulnerable?

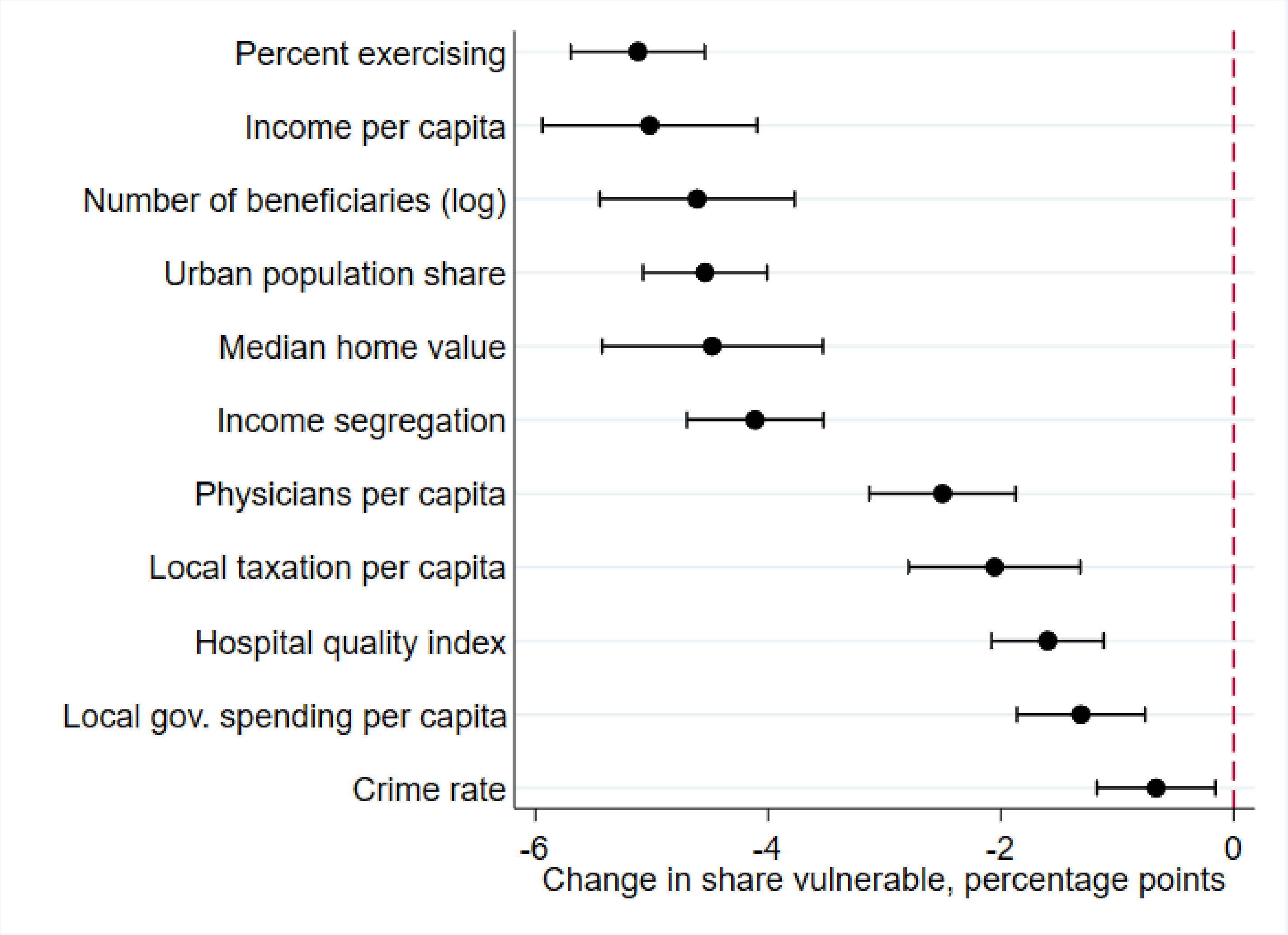
Outcome	(1) Bottom 75%	(2) Top 25%	(3) Difference
Medical spending (dollars)			
Durable medical equipment	160	333	173***
			(0.438)
Hospice	148	394	245***
			(1.69)
Hospital outpatient	1,177	2,246	1,068***
			(2.63)
Part B drug	277	637	360***
			(3.11)
Part B other	118	264	146***
			(0.710)


2013 Annual PM2.5 Level
5
6
7
9
10
11

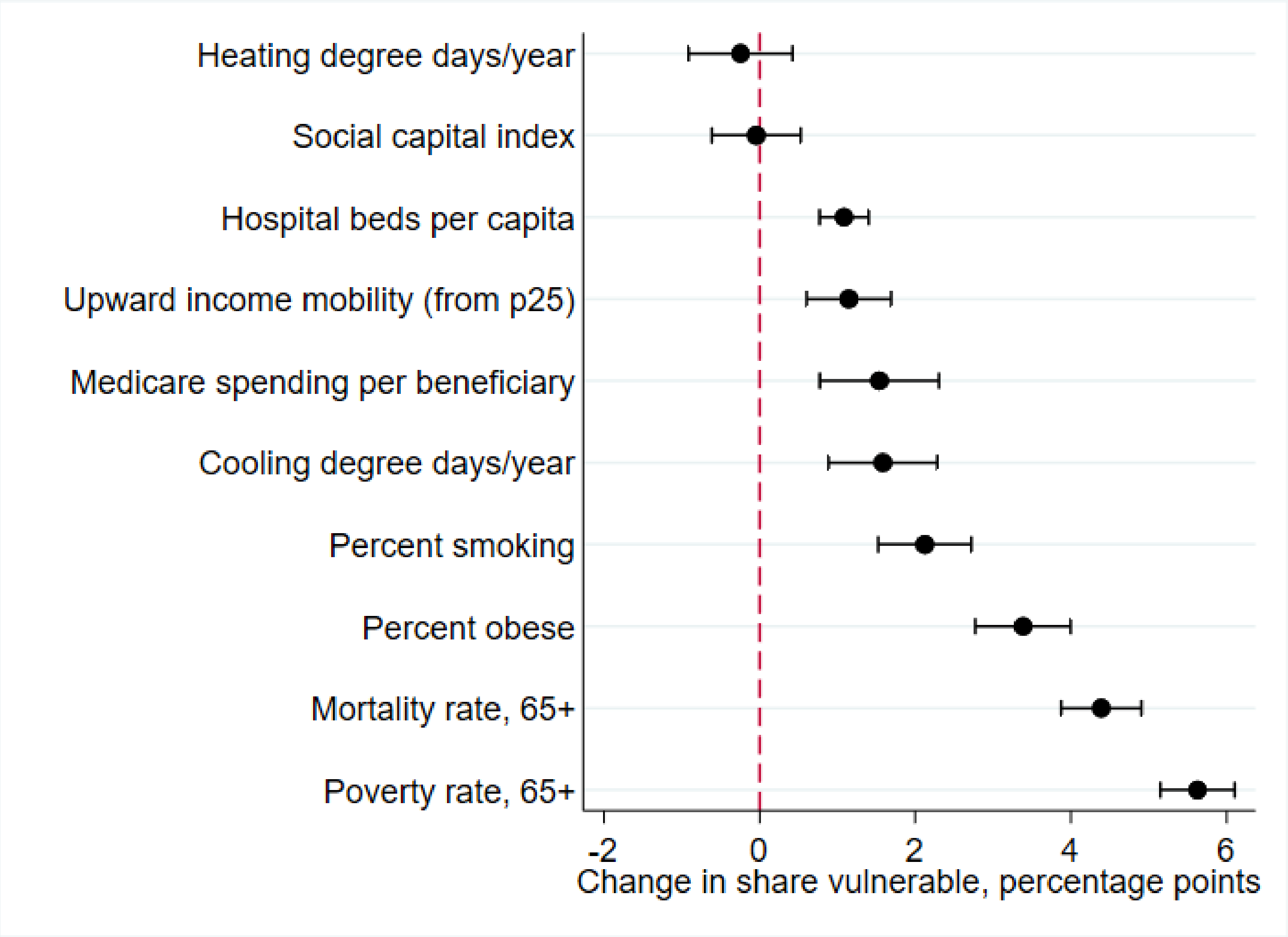

Percent of Vulnerable Beneficiaries
10%
20%
25%
30%
40%
50%

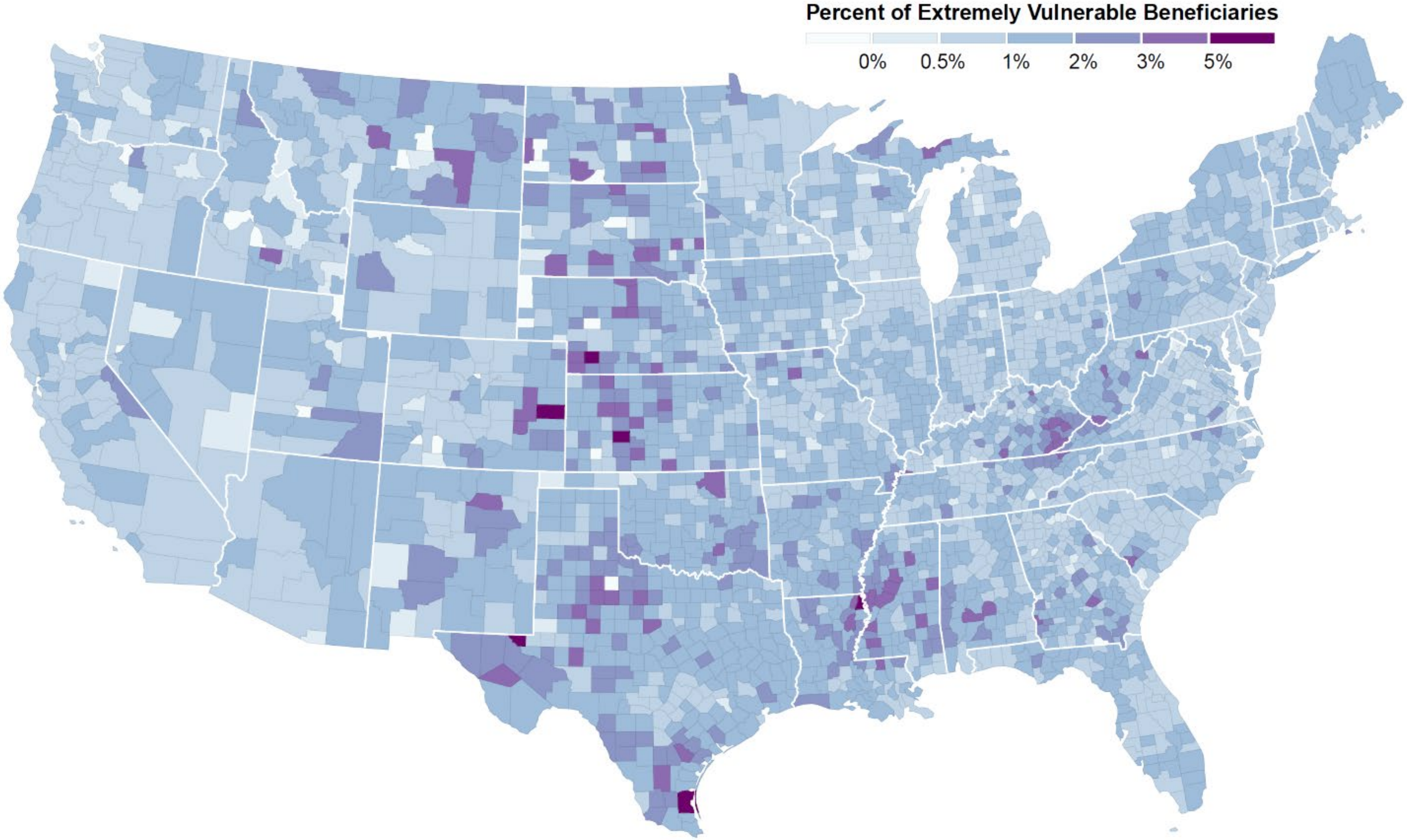

Fraction of Vulnerable (Top 25%) Beneficiaries
1.0
0.8
0.6
0.4
0.2
0.0
5
10
15
20
2013 Annual PM2.5 Level

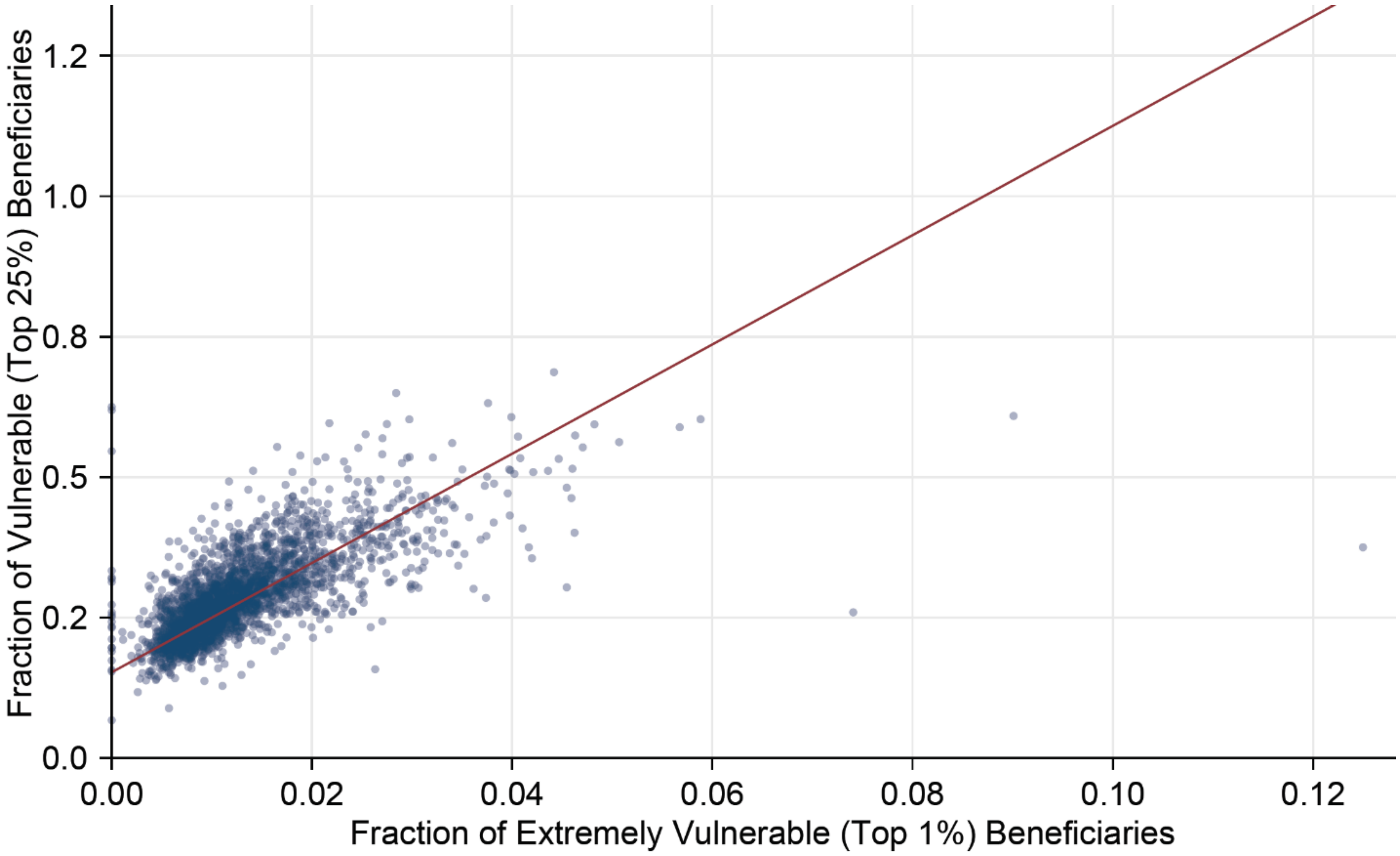
Predictors of lower vulnerability



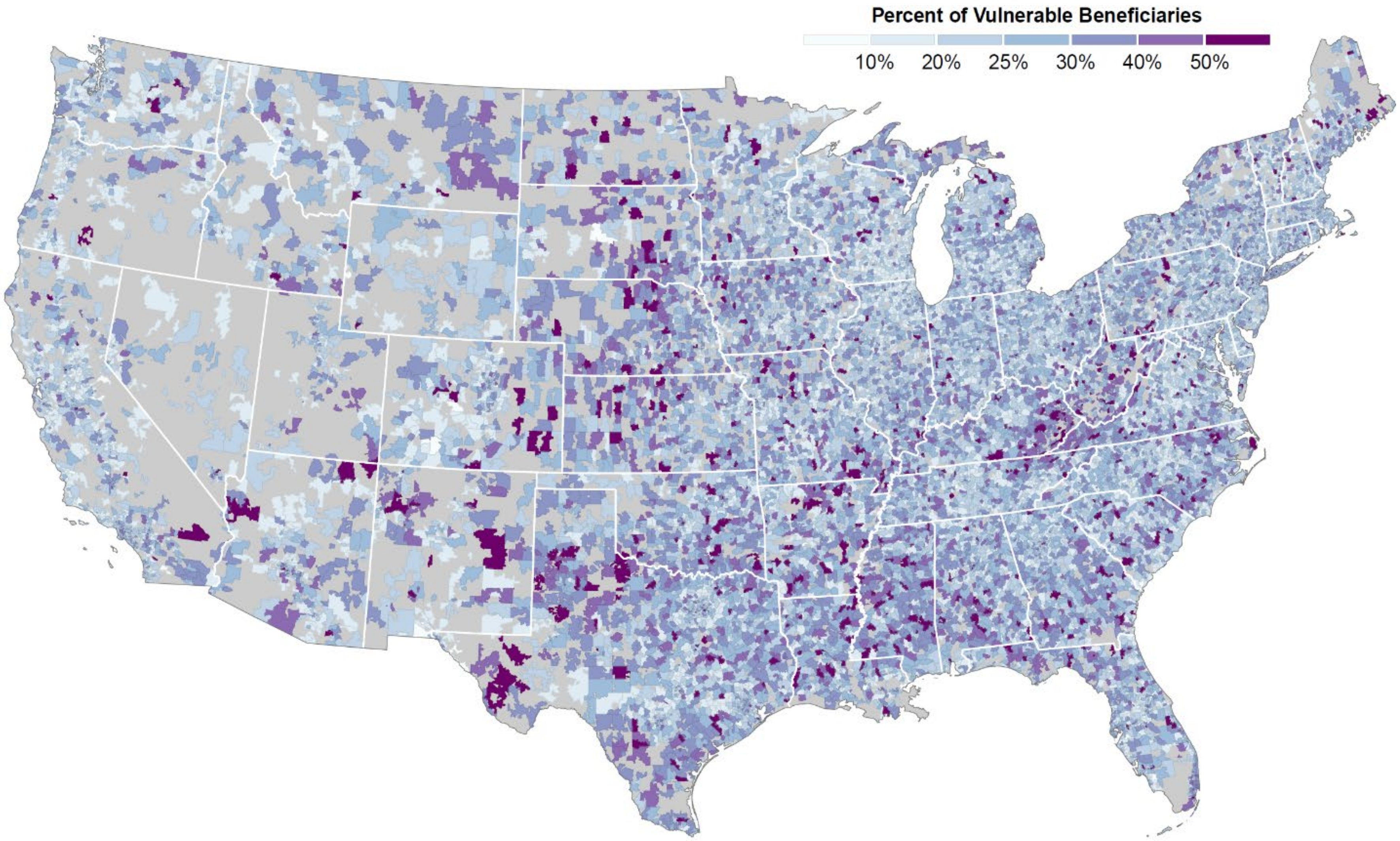
Predictors of greater vulnerability




Percent of Extremely Vulnerable Beneficiaries
0%
0.5%
1%
2%
3%
5%



Fraction of Vulnerable (Top 25%) Beneficiaries
1.2
1.0
0.8
0.5
0.2
0.0
0.00
0.02
0.04
0.06
0.08
0.10
0.12
Fraction of Extremely Vulnerable (Top 1%) Beneficiaries


Percent of Vulnerable Beneficiaries
10%
20%
25%
30%
40%
50%

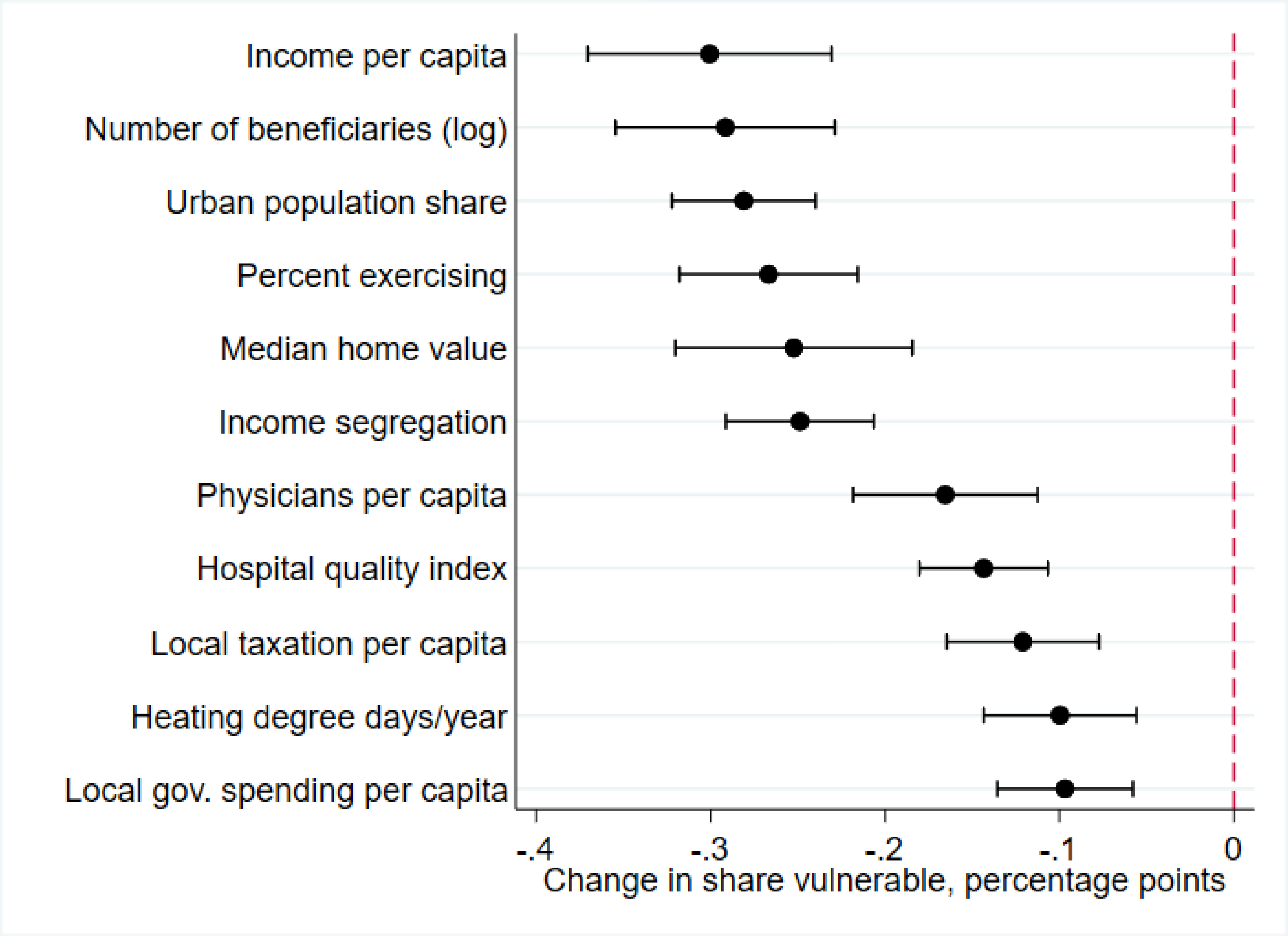
Conclusion

- There is substantial geographic variation in vulnerability to acute $PM_{2.5}$ exposure among the US elderly

- Vulnerability is negatively related to $PM_{2.5}$ levels, income, urbanicity, and exercising; it is positively related to cooling degree days, poverty, and smoking, obesity and mortality rates

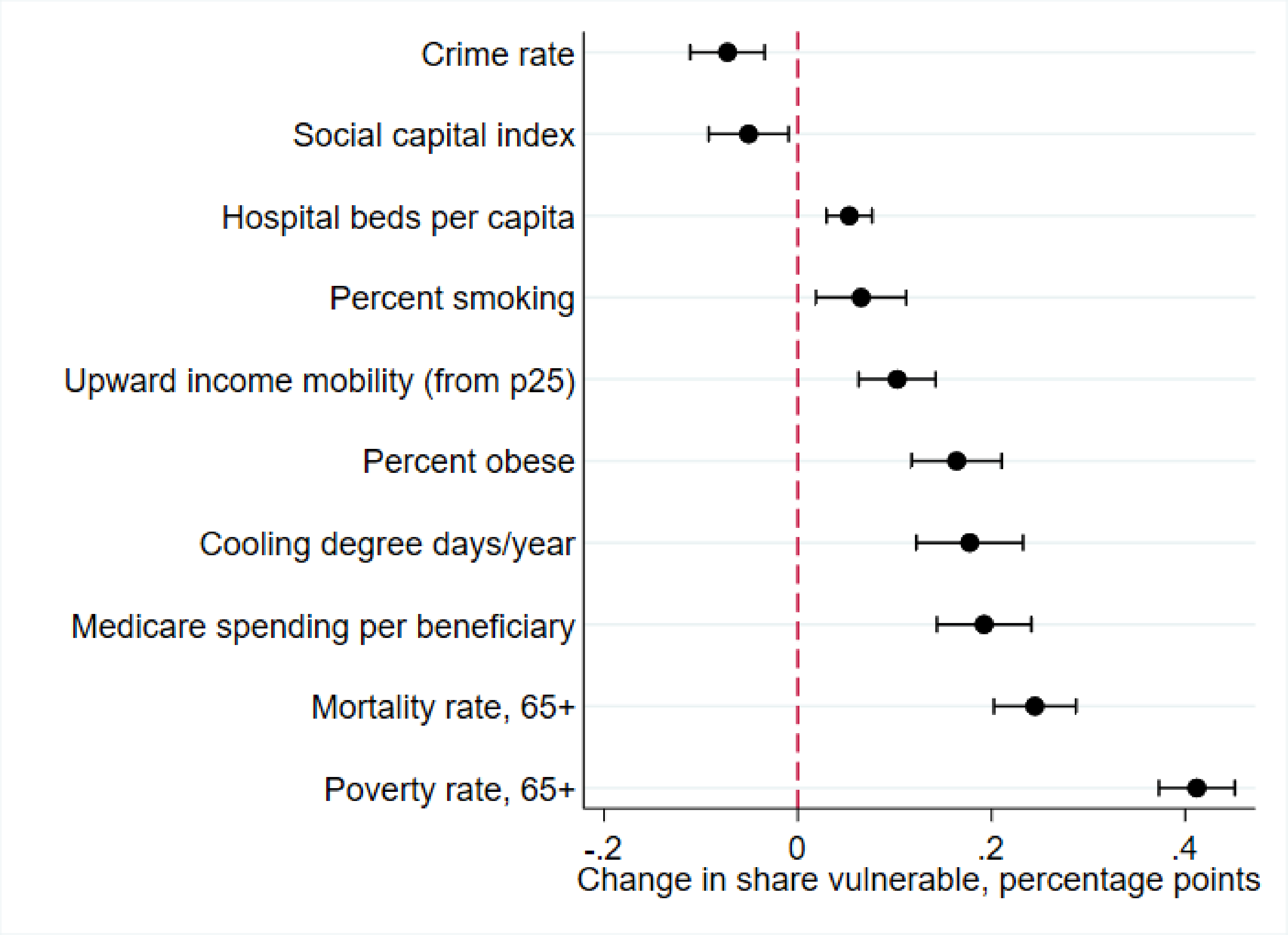
- Considerations of area characteristics other than pollution levels may improve efficiency of targeted pollution reduction strategies
  - Limitations: study population limited to elderly, chronic exposure not considered

Extra slides



Predictors of lower vulnerability, top 1%

Predictors of greater vulnerability, top 1%



Our first stage has 300 instruments

Allow pollution transport patterns to vary across 100 monitor groups, formed using a clustering algorithm.

- 100 different spatial regions (g)
- 3 different 90-degree bins (b) (1 omitted category)

First stage is group-specific relationship between wind direction and pollution:

$$PM2.5_{cdmy} = \sum_{g=1}^{100} \sum_{b=0}^{2} \beta_b^g 1[G_c = g] \times WINDDIR_{cdmy}^{90b} + [\ldots] + \epsilon_{cdmy}$$

c = county, d = day, m = month, y = year

Second stage

$$Y_{cdmy} = \beta PM2.5_{cdmy} + X'_{cdmy}\boldsymbol{\gamma} + \alpha_c + \alpha_{sm} + \alpha_{my} + \epsilon_{cdmy}$$

Y_{cdmy} = 3-day mortality rate

- Alternative specifications extend outcome window to 28 days

Flexibly control for interactions of temperature (min and max), precipitation, and wind speed

- Control for instruments and weather on days $d + 1$ and $d + 2$ to identify effect of 1-day shock
- Control for instruments on $d - 1$ and $d - 2$ to account for autocorrelation
- In total, about 28,000 control variables

Estimating the proxy predictor, $\hat{S}(Z_{it})$

Treatment group $\xrightarrow{50\%}$ Auxiliary subsample

Treatment group $\xrightarrow{50\%}$ Main subsample

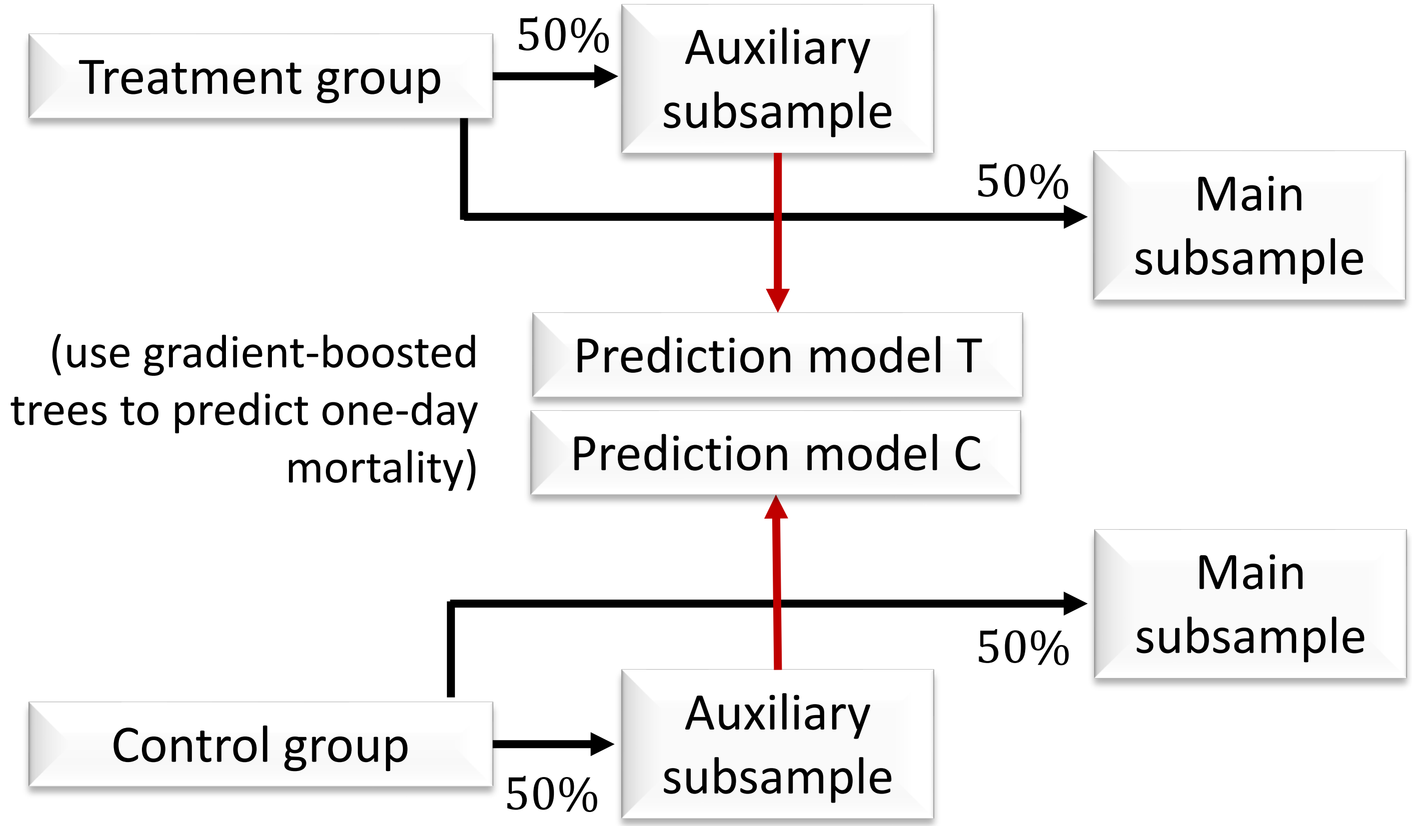
(individual-days where wind is blowing from “high-pollution” direction. PM 2.5 is 2.4 $\frac{\mu g}{m^3}$ higher than control group)

Control group $\xrightarrow{50\%}$ Main subsample

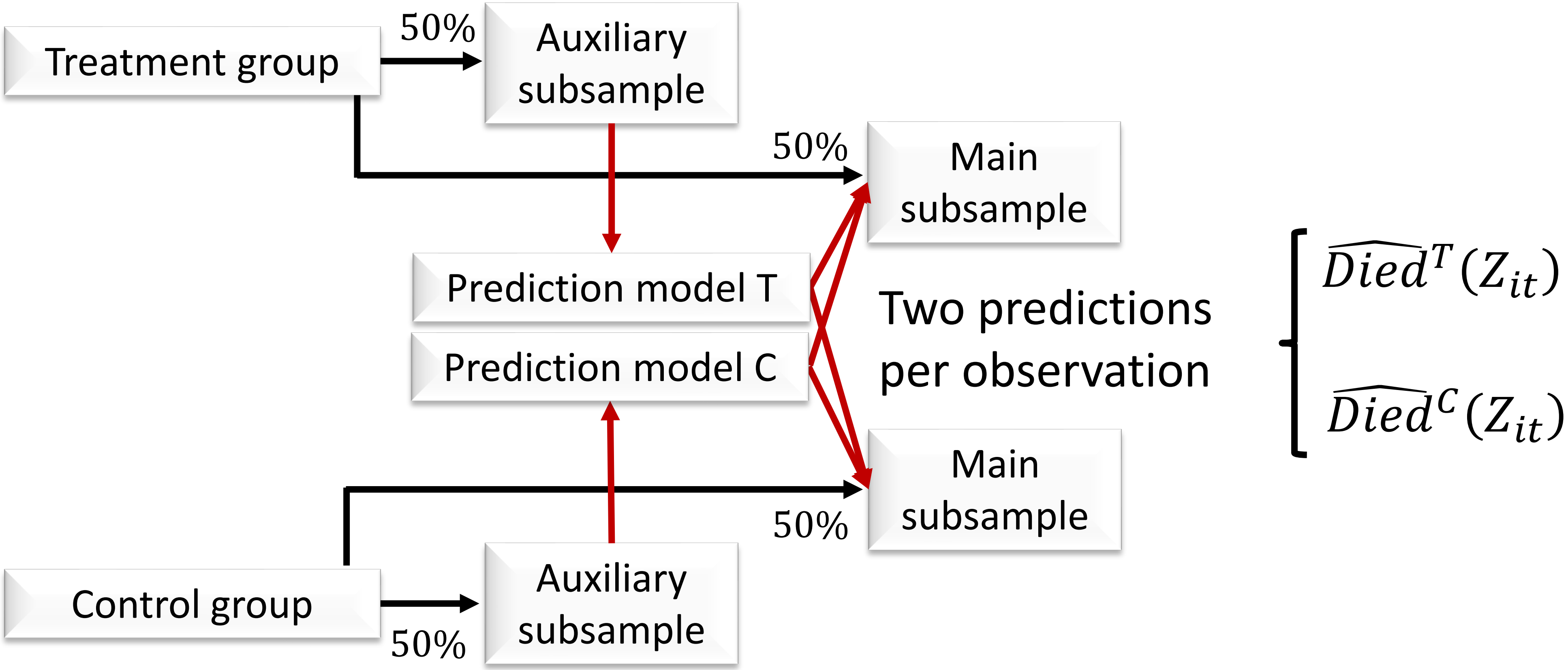
Control group $\xrightarrow{50\%}$ Auxiliary subsample

(“low-pollution” wind direction)

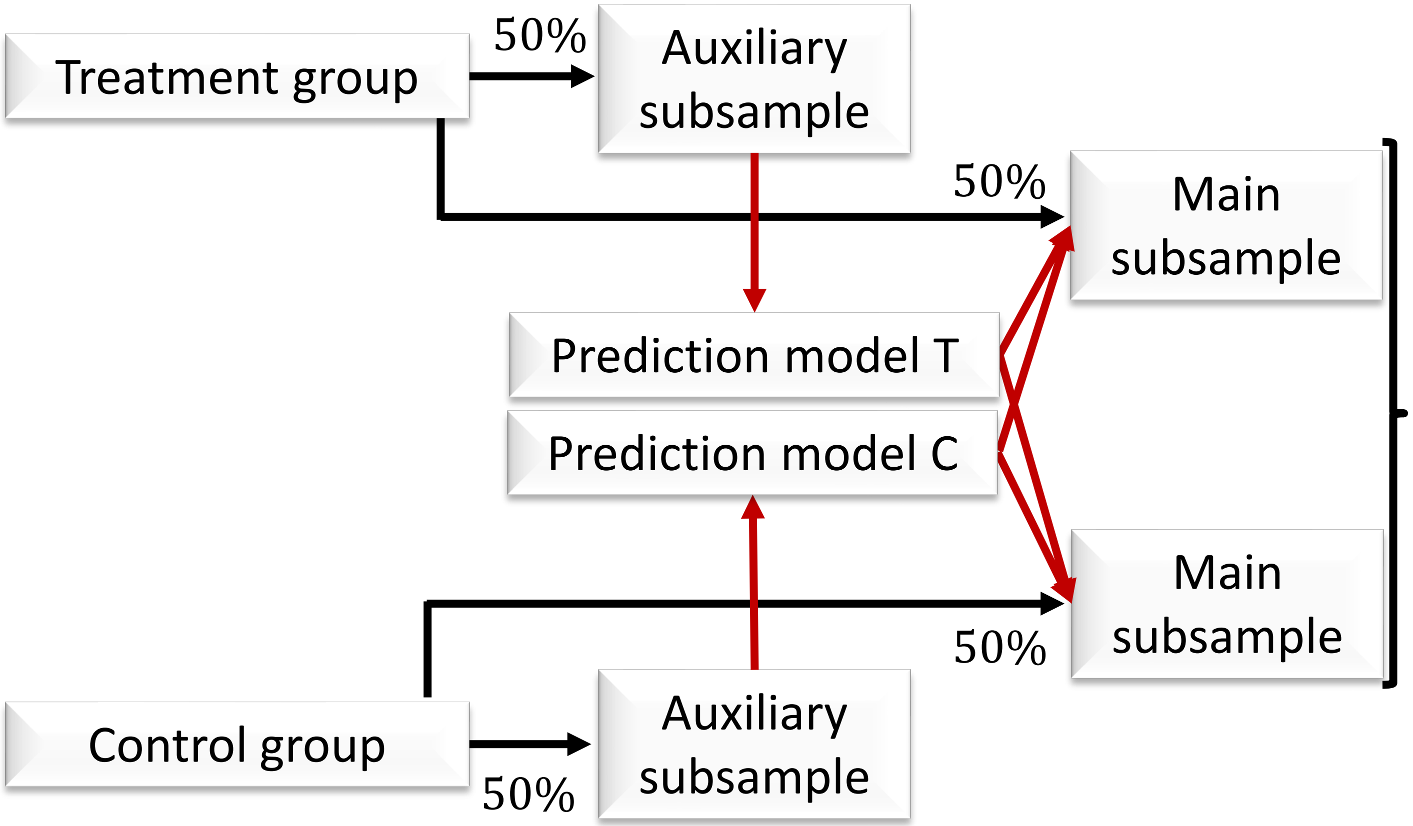
Estimating the proxy predictor, $\hat{S}(Z_{it})$



Estimating the proxy predictor, $\hat{S}(Z_{it})$



Estimating the proxy predictor, $\hat{S}(Z_{it})$



$$\hat{S}(Z_{it}) = \widehat{Died}^{T}(Z_{it}) - \widehat{Died}^{C}(Z_{it})$$

Using $\hat{S}(Z_{it})$ to characterize heterogeneity

1. Estimate average treatment effects for different groups, indexed by k

$$Died_{it} = \alpha + \sum_{k=1}^{K} \gamma_k \big(T_{it} - \hat{p}(Z_{it})\big) \cdot 1(G_k) + \theta \widehat{Died}^C(Z_{it}) + \epsilon_{it}$$

- $\hat{p}(Z_{it})$ is the estimated propensity score for person i at time t
- Coefficients $\gamma_k = E[s_0(Z_{it})|G = k]$
- Groups are percentiles of the proxy predictor $\hat{S}(Z_{it})$ → sorted group average treatment effects (GATES)

2. Compare mean characteristics of most-affected versus least-affected group

- Groups again defined by $\hat{S}(Z_{it})$