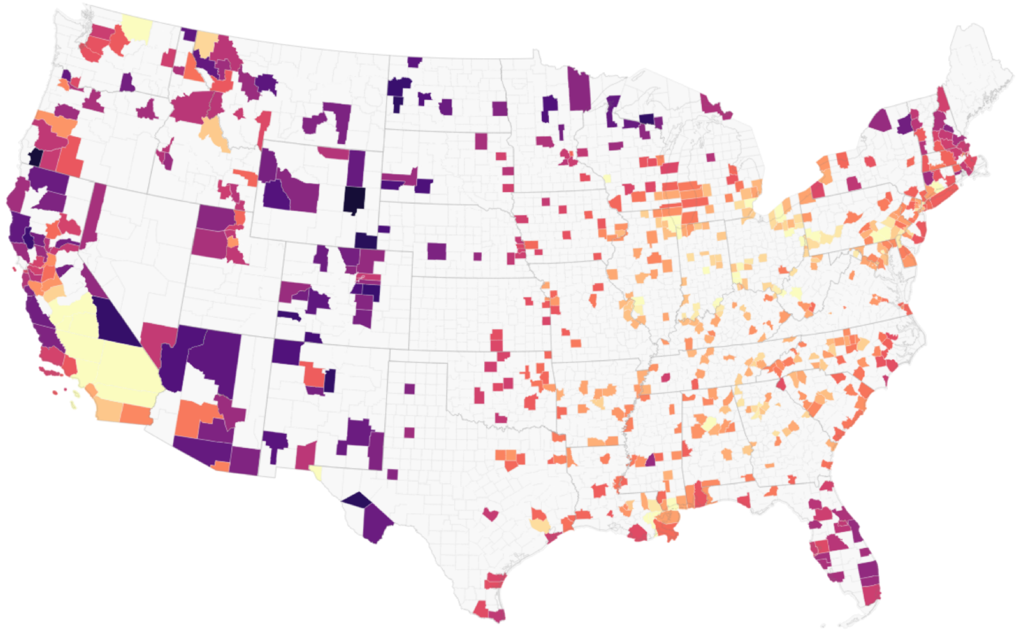
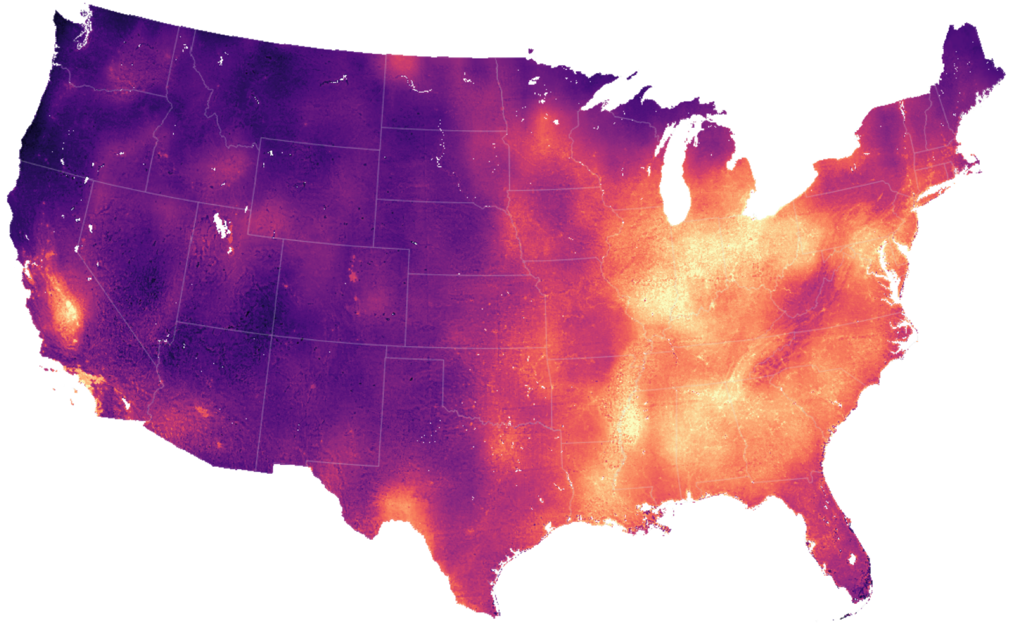
From Big Data to Big Decisions
better data + better analytics = better policy?

Meredith Fowlie
UC Berkeley
TWEEDS 2019



Sparse network of EPA PM2.5monitors



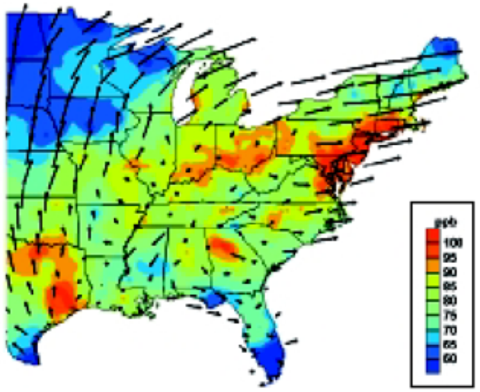
Satellite-based estimates (Van Donkelaar et al 2019)

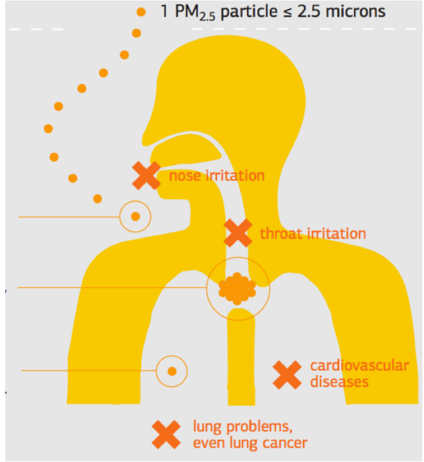
Data improvements -> welfare improvements?

- Policy is one essential lever for translating research insights into welfare improvements.

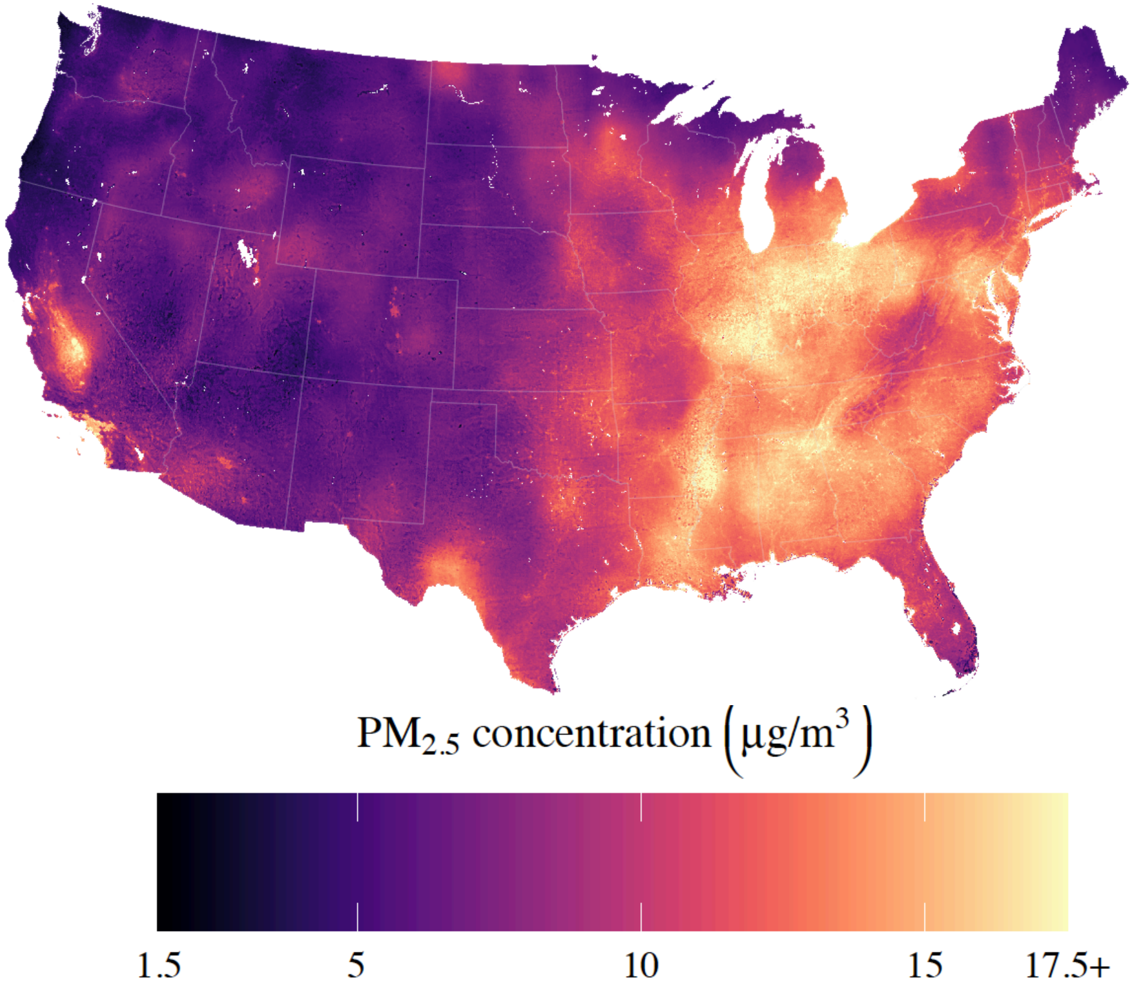
- But the causal relationship between environmental policy implementation and health/welfare is complicated.

Policy change



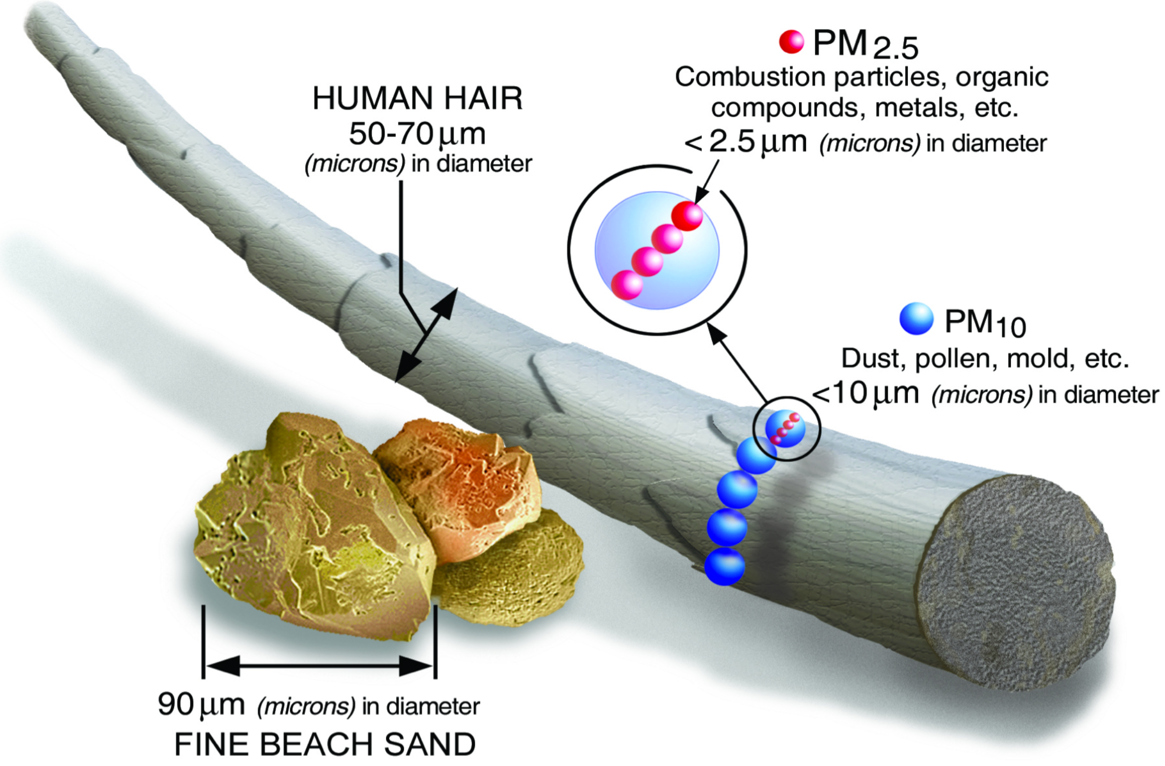


How to act on new data/analysis?



PM2.5 101

- Particulate matter (PM) with a diameter of less than 2.5 micrometers.
- Exposure measured in $\mu g/m^3$.
- Chemically ambiguous: volcanoes, dust storms, fires, human activity such as power plants and gasoline combustion.
- Composition (and health impact?) depends on the source.



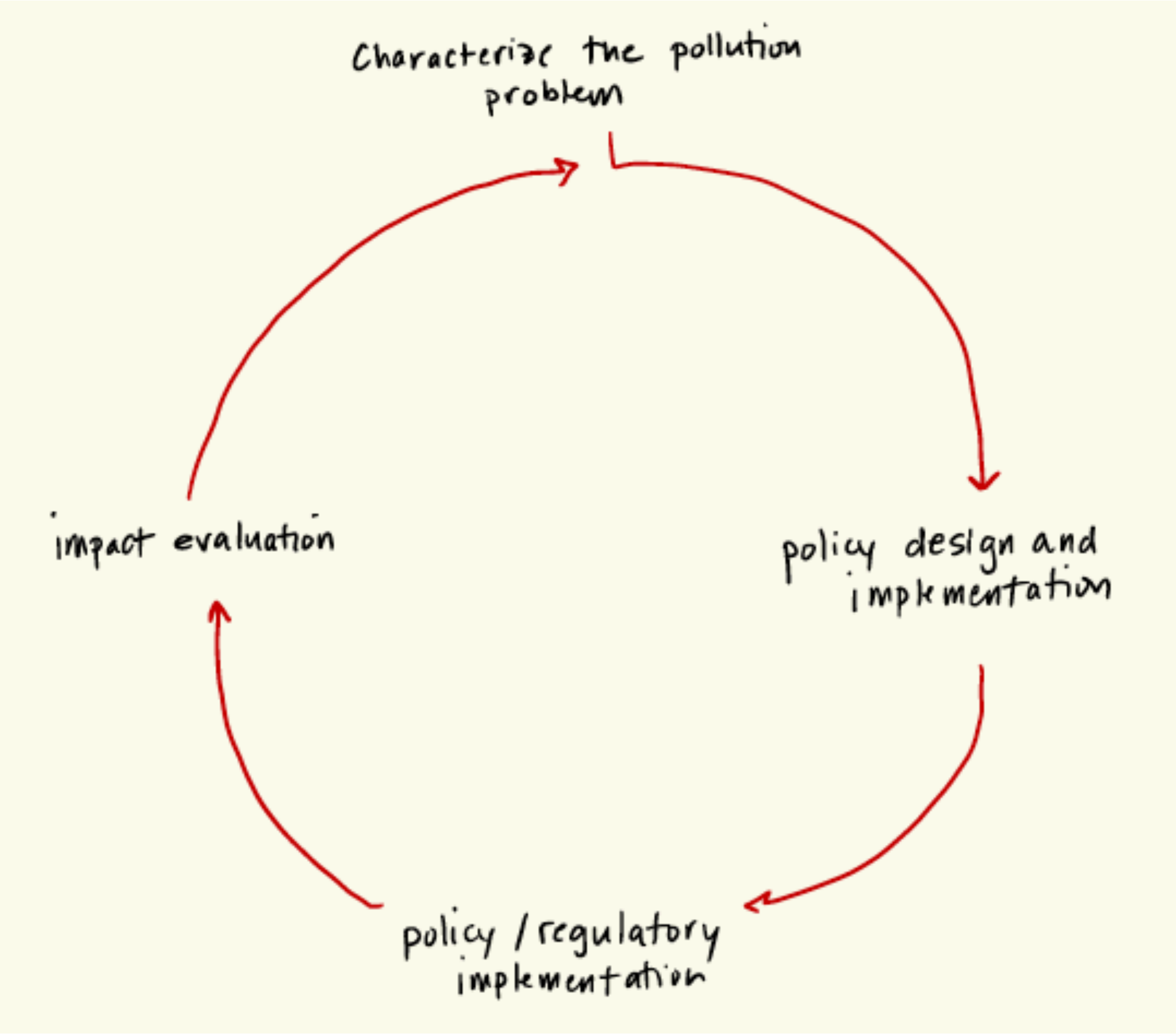
These small particulates are a big deal

- Muller and Mendelsohn (2007) estimate that health impacts of PM2.5 (primary + secondary) accounts for over 90% of total gross annual damages from air pollution (US)

Mortality impacts of PM2.5 exposure:

- >98% of benefits of last revision of PM2.5 national air quality standards.
- >94% of benefits from proposed “Affordable Clean Energy (ACE) rule”
- >99% of benefits from (now endangered) Mercury and Toxics rule.

My (wishful) model of how research can impact real-world policy outcomes



Beyond the glare of Washington, President Trump's retreat on the environment is unfolding in consequential ways for the health and safety of Americans.

Air Pollution Denial Could Become EPA Policy

For decades, the agency has said that inhaling soot in any amount is unsafe. The Trump administration might change that.

ENVIRONMENT

EPA Science Panel Considering Guidelines That Upend Basic Air Pollution Science

March 28, 2019 · 6:34 PM ET

Characterizing health impacts of PM2.5 exposure

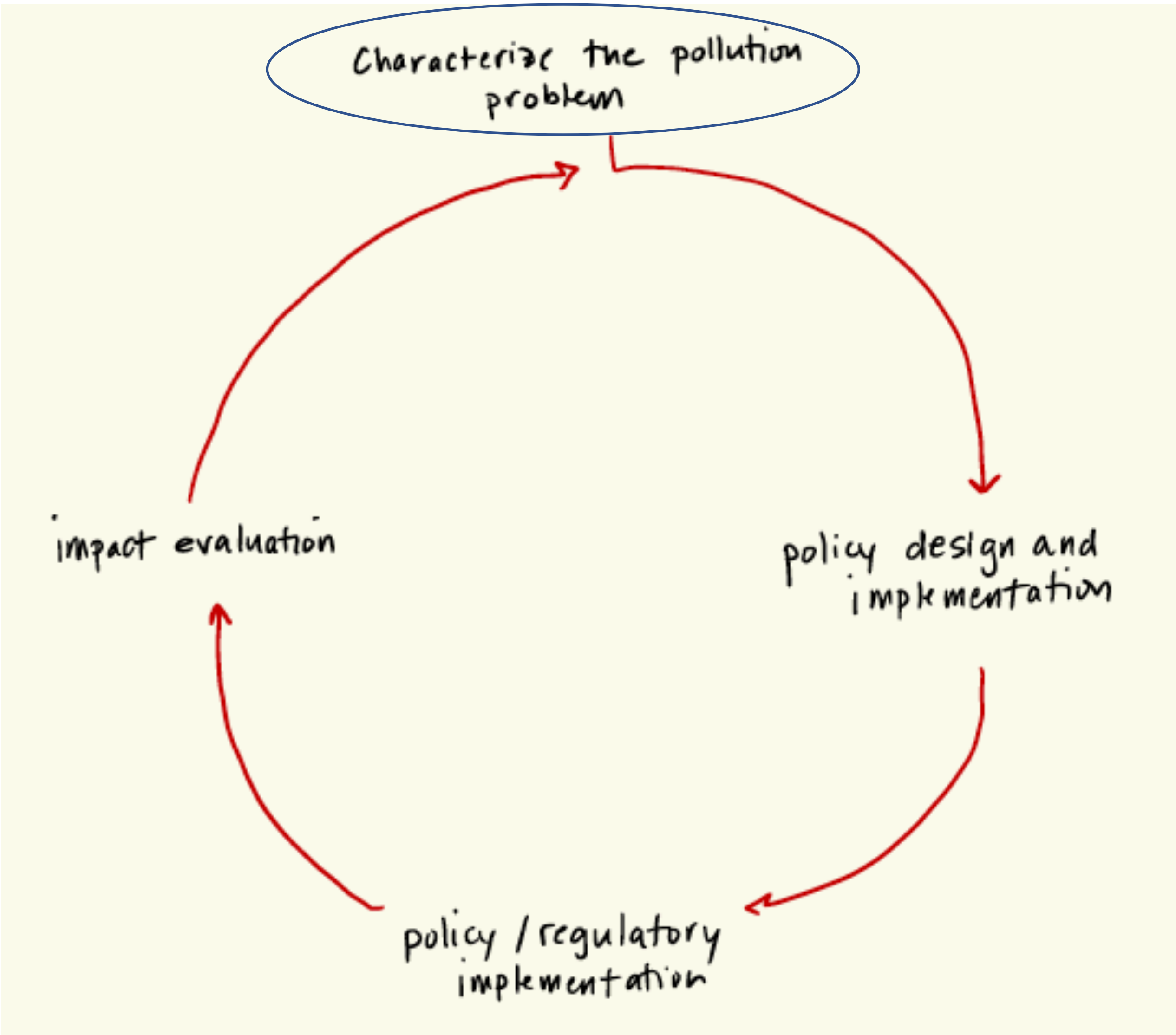


Table 5-21. Total Estimated Monetized Benefits of the for Revised and Alternative Annual Primary $PM_{2.5}$ Standards (Incremental to the Analytical Baseline) (billions of 2006$) [a,b]

Benefits Estimate	13 µg/m^3	12 µg/m^3	11 µg/m
Economic value of avoided $PM_{2.5}$-related morbidities and premature deaths using $PM_{2.5}$ mortality estimate from Krewski et al. (2009)			
3% discount rate	$1.3 + B	$4.0 +B	$13 + B
7% discount rate	$1.2 + B	$3.6 +B	$12 + B
Economic value of avoided $PM_{2.5}$-related morbidities and premature deaths using $PM_{2.5}$ mortality estimate from Lepeule et al. (2012)			
3% discount rate	$2.9 + B	$9.1 +B	$29 + B
7% discount rate	$2.6 + B	$8.2 +B	$26 + B

Source: Regulatory Impact Analysis for the Final Revisions to the National Ambient Air Quality Standards for Particulate Matter

Where do these estimates come from?

$$\Delta y = 1 - \left(e^{\beta \cdot \Delta x}\right) y_o \cdot Pop \qquad (5.1)$$

where y_0 is the baseline incidence rate for the health endpoint being quantified (for example, a health impact function quantifying changes in mortality would use the baseline, or background, mortality rate for the given population of interest); *Pop* is the population affected by the change in air quality; Δx is the change in air quality; and β is the effect coefficient drawn from the epidemiological study. Figure 5-1 provides a simplified overview of this approach.

Source: Regulatory Impact Analysis for the Final Revisions to the National Ambient Air Quality Standards for Particulate Matter

Where do these estimates come from?

- **Harvard 6 Cities Study**: Adults (25-74 years) randomly sampled from 6 cities in 1970s with PM2.5 monitors.

- **Krewski et al. (2009)** conduct extended (18 years) follow up and re-analysis to evaluate concerns about potential confounds, spatial auto-correlation
  - Hazard ratio elevated by 3-15% for each 10 $\mu g/m^3$ increase.
  - Intra-urban analysis in NYC and LA estimate “strikingly dissimilar” associations between exposure and mortality.

- **Lepeule et al. (2012)** incorporates lower exposures and evaluates alternative lags/CR-relationships.
  - 10 $\mu g/m^3$ associated with an increase in all cause (lung cancer) mortality of 14% (37%)
  - 1-3 year lag yields best fit.

Concerns and limitations?
(Krewski et al. 2009; Lepeule et al. 2012)

- **Omitted variable bias**: “there is potential for important residual confounding with unmeasured factors”
- **Heterogeneous treatment**: “Relative toxicity of particle elements remains highly controversial”
- **Divergent intra-urban results**: “our results argue for caution in extrapolating from on metro area to another”
- **Exposure assessment**: “Use of central monitoring stations as a proxy measure of mean personal exposure is prone to measurement error”

Cue the econometricians!

The Mortality and Medical Costs of Air Pollution: Evidence from Changes in Wind Direction

By TATYANA DERYUGINA, GARTH HEUTEL, NOLAN H. MILLER, DAVID MOLITOR, AND JULIAN REIF

March 2019

PM2.5 concentration instrument: *Daily* wind direction

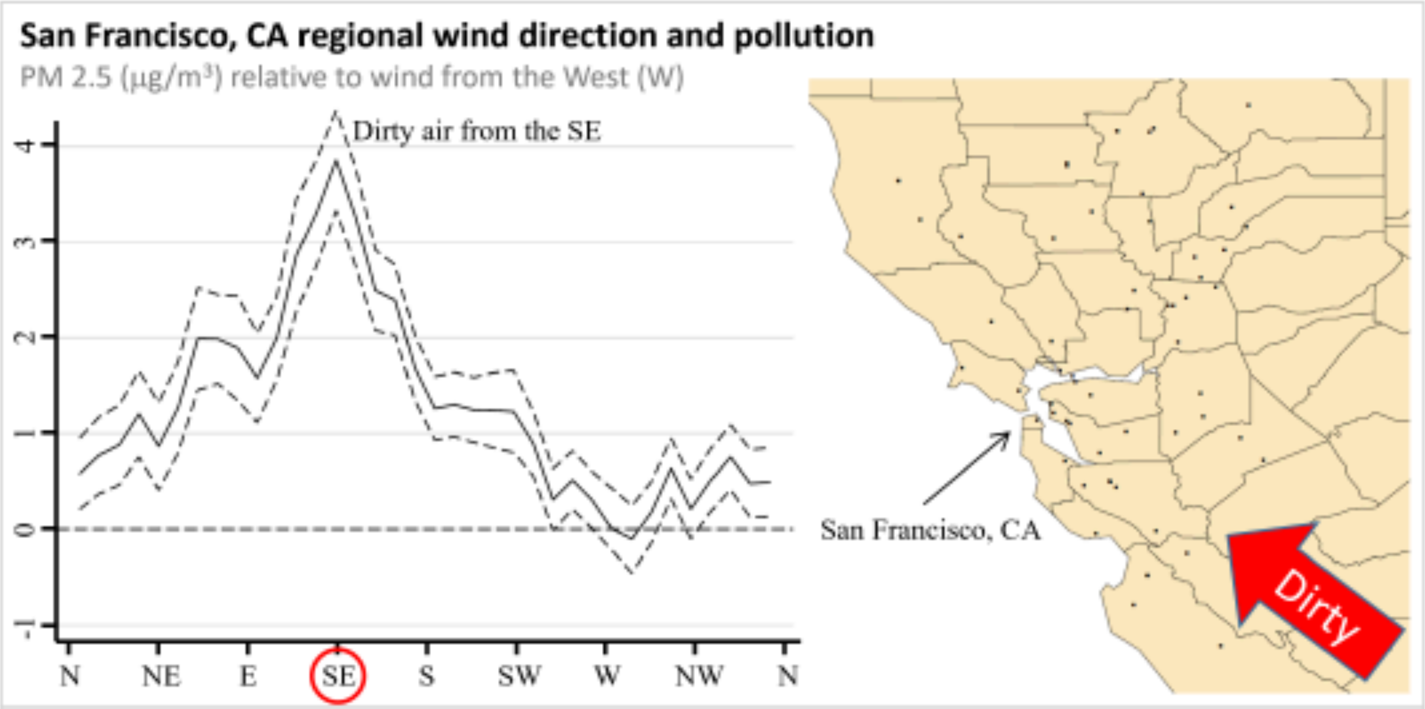


Figure 2. Relationship between daily average wind direction and PM 2.5 concentrations for counties in and around the Bay Area, CA. The left panel shows regression estimates of equation (A1) from the Online Appendix, where the dependent variable is the county average daily PM 2.5 concentration and the key independent variables are a set of indicators for the daily wind direction falling into a particular 10-degree angle bin. Controls include county, month-by-year, and state-by-month fixed effects, as well as a flexible function of maximum and minimum temperatures, precipitation, wind speed, and the interactions between them. The dashed lines represent 95 percent confidence intervals based on robust standard errors. The right panel shows the location of the PM 2.5 pollution monitors (black dots) in the Bay Area that provided the pollution measures for this regression.

Table 2: OLS and IV estimates of effect of PM 2.5 on elderly mortality, by age group

	(1) 65+	(2) 65–69	(3) 70–74	(4) 75–79	(5) 80–84	(6) 85+
Panel A: OLS estimates						
PM 2.5 ($\mu g/m^3$)	0.095***	0.041***	0.029	0.022	0.142***	0.425***
	(0.021)	(0.014)	(0.018)	(0.022)	(0.036)	(0.072)
Dep. var. mean	385	131	197	312	508	1,127
Effect relative to mean, percent	0.025	0.032	0.015	0.007	0.028	0.038
Observations	1,980,549	1,980,549	1,980,549	1,980,549	1,980,549	1,980,549
Adjusted R-squared	0.254	0.080	0.085	0.082	0.077	0.110
Panel B: IV estimates						
PM 2.5 ($\mu g/m^3$)	0.685***	0.267***	0.329***	0.348***	0.877***	2.419***
	(0.061)	(0.066)	(0.068)	(0.098)	(0.159)	(0.246)
F-statistic	298	285	292	303	309	315
Dep. var. mean	385	131	197	312	508	1,127
Effect relative to mean, percent	0.178	0.204	0.167	0.111	0.173	0.215
Observations	1,980,549	1,980,549	1,980,549	1,980,549	1,980,549	1,980,549

Notes: Table reports OLS and IV estimates of equation (1) from the main text. Dependent variable is the three-day mortality rate per million beneficiaries in the relevant age group. All regressions include county, month-by-year, and state-by-month fixed effects; flexible controls for temperatures, precipitation, and wind speed; and two leads of these weather controls. OLS (IV) estimates also include two lags and two leads of PM 2.5 (instruments). Estimates are weighted by the number of beneficiaries in the relevant age group. Standard errors, clustered by county, are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Unleash machine learning on health benefit estimation

$$Total\ Life\ Years = \sum_{i=1}^{n} LE_i \times M_i \tag{5.2}$$

where LE_i is the average remaining life expectancy for age interval i, M_i is the estimated change in number of deaths in age interval i, and n is the number of age intervals.

Source: Regulatory Impact Analysis for the Final Revisions to the National Ambient Air Quality Standards for Particulate Matter

- Estimating life-years lost is challenging because counterfactual life expectancy is unobserved.
- Standard approaches overstate years lost if individuals impacted by pollution have shorter life expectancies.
- Use machine learning techniques to predict individual life expectancy using 1,062 variables on health history
- Improved estimates reduce counterfactual life expectancy significantly!

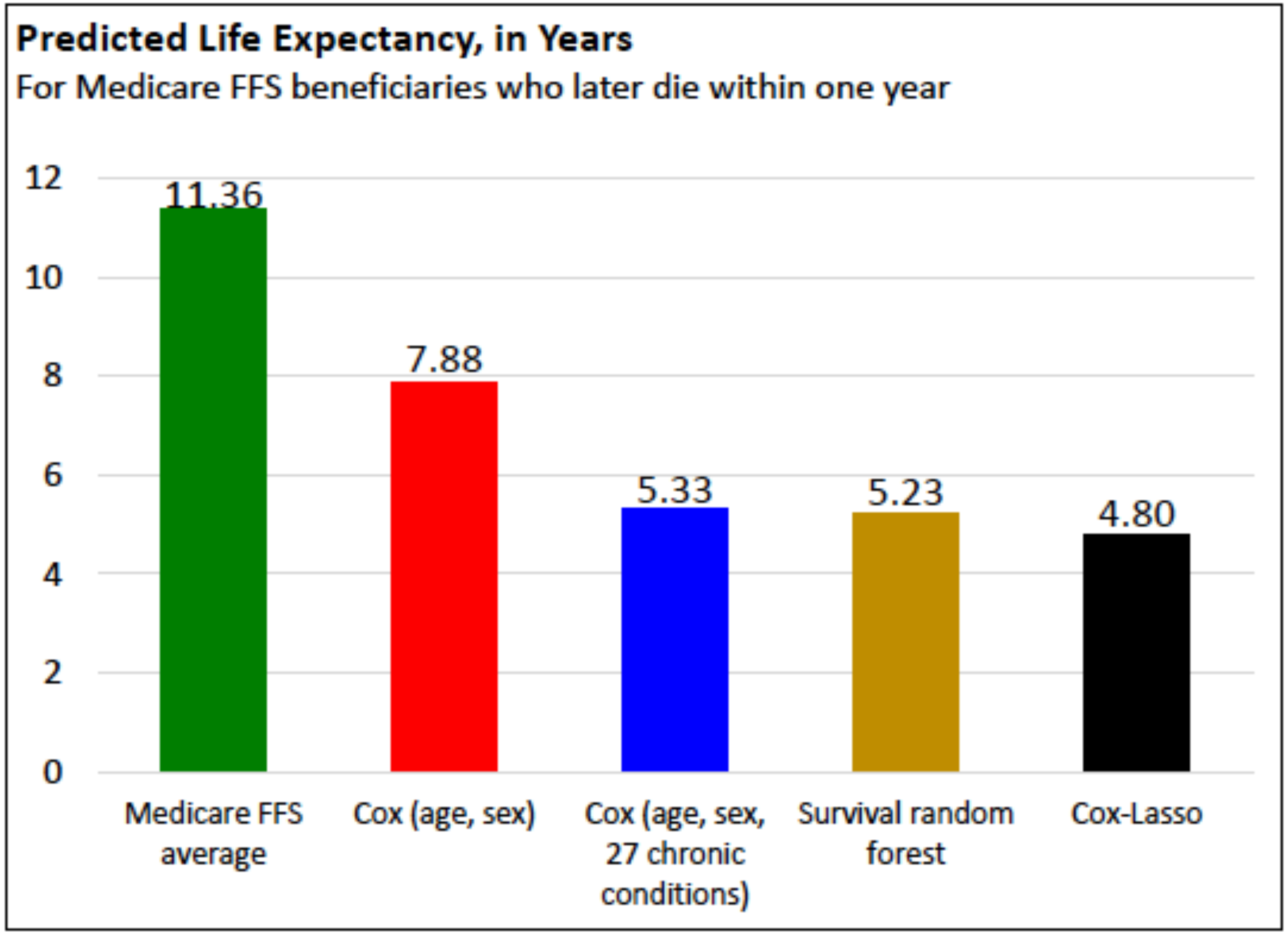


Figure 5. Average life expectancy for continuously enrolled FFS Medicare beneficiaries who later die within one year, 2001–2013. Life expectancy for each beneficiary is estimated as of January 1 of the calendar year of death. Estimates for "Medicare FFS average" are produced by MLE estimation of survival model (6) with no covariates. Estimates for "Cox (age, sex)" and "Cox (age, sex, cc) are produced by estimating the survival model (6) using age and sex, and age, sex and 27 chronic conditions, as predictors, respectively. Estimates for "Cox-Lasso" are produced by machine learning estimation of the survival model (7) with 1,062 predictors (including interactions). Estimates for survival random forest are produced by machine learning estimation using the same predictors as Cox-Lasso.

Policy relevant? Yes, but….

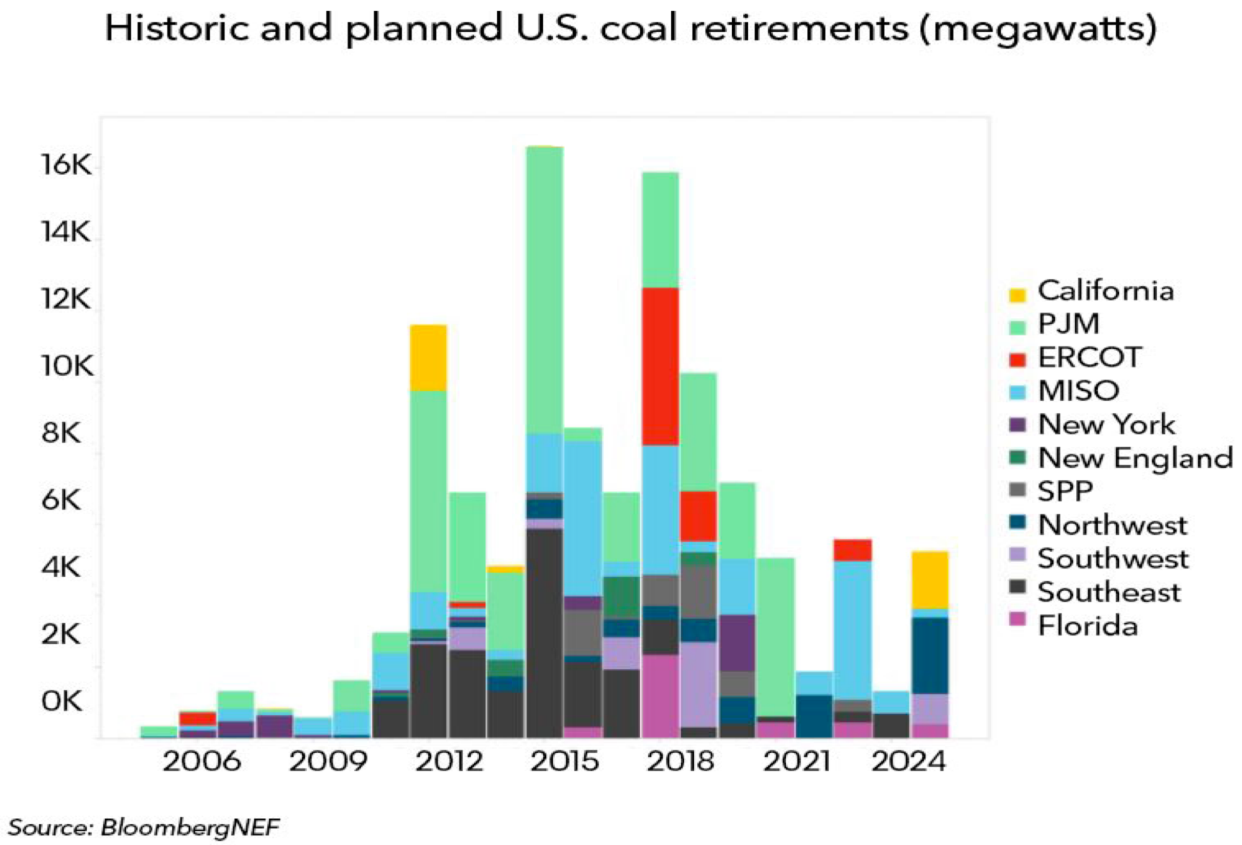
- Emphasis on short-run variation could limit the impact of these findings.

“When choosing between using short-term studies or cohort studies for estimating mortality benefits.. it is essential to use longer-run studies to capture the important effects”

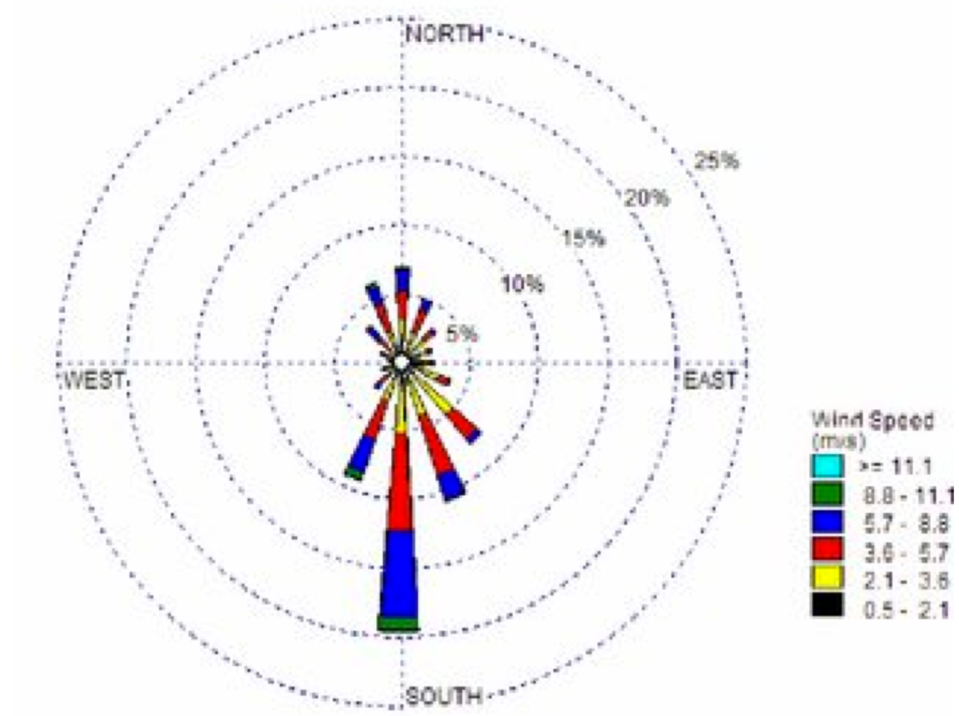
- Identification strategy complicates interpretation – are we capturing kale or KFC?

- Reliance on sparse monitors is also limiting….

Some related work in the works…



- Isolate effects of power plant emissions (main contributer)



Pollution from a coal-fired power plant
(Monticello in Titus County, Texas)
Closed in January 2018

Satellite-based estimates present challenges!

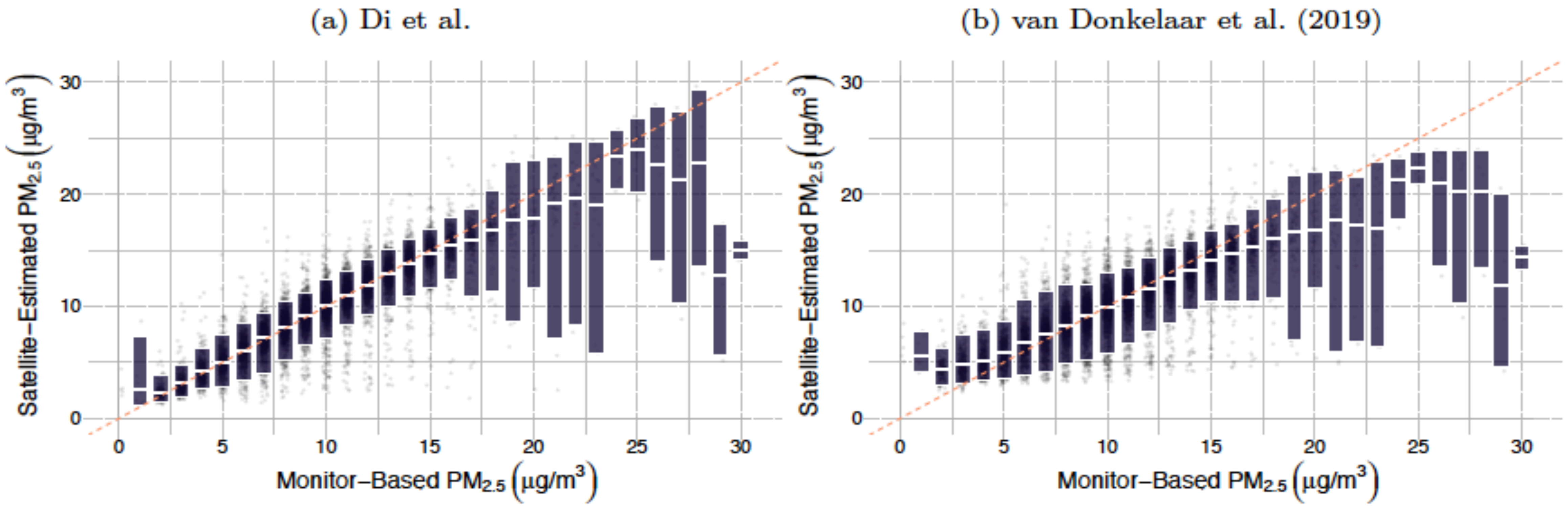
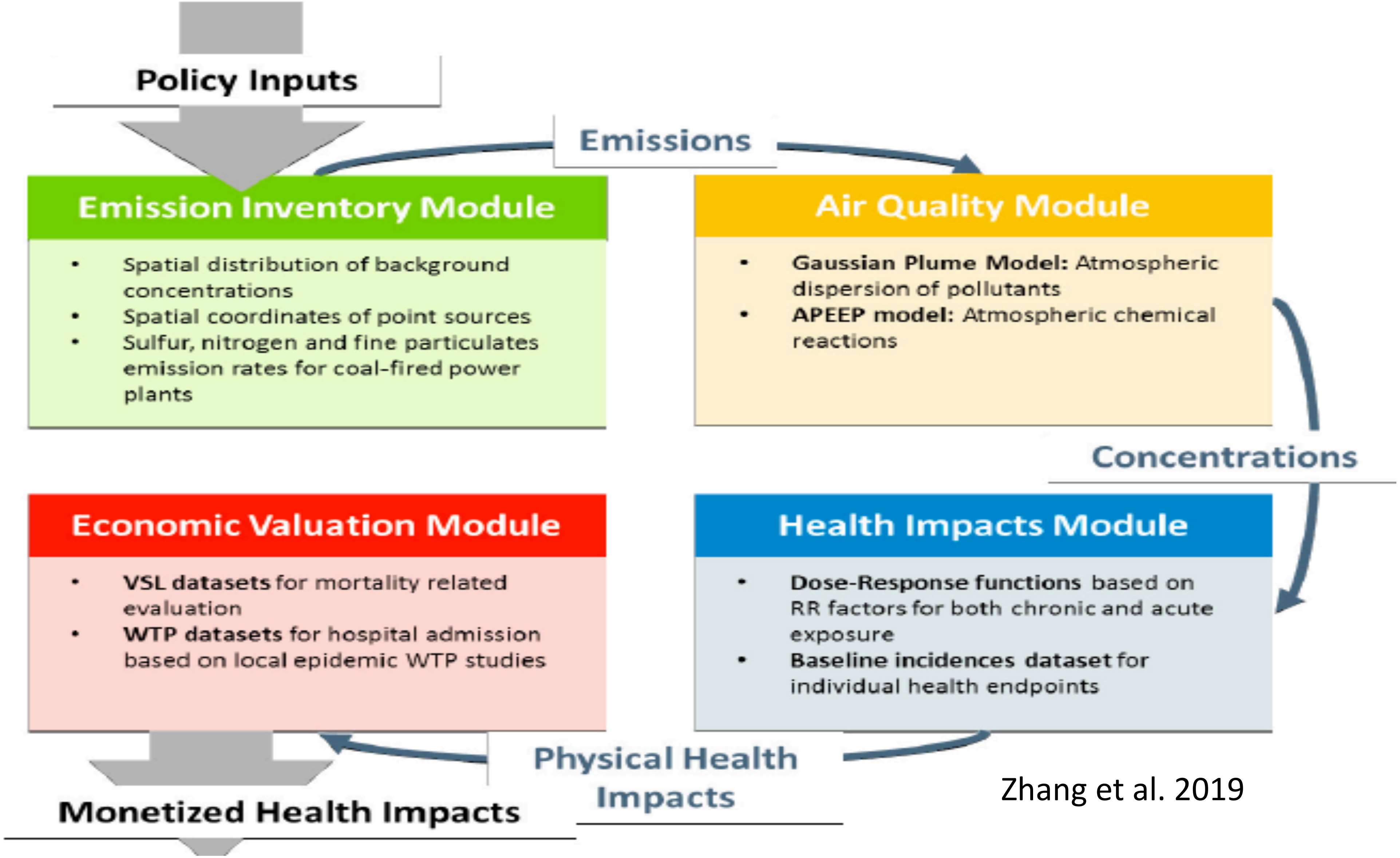


Figure 2. : Comparing $PM_{2.5}$: Monitors' Measurements vs. Satellite-Based Estimates

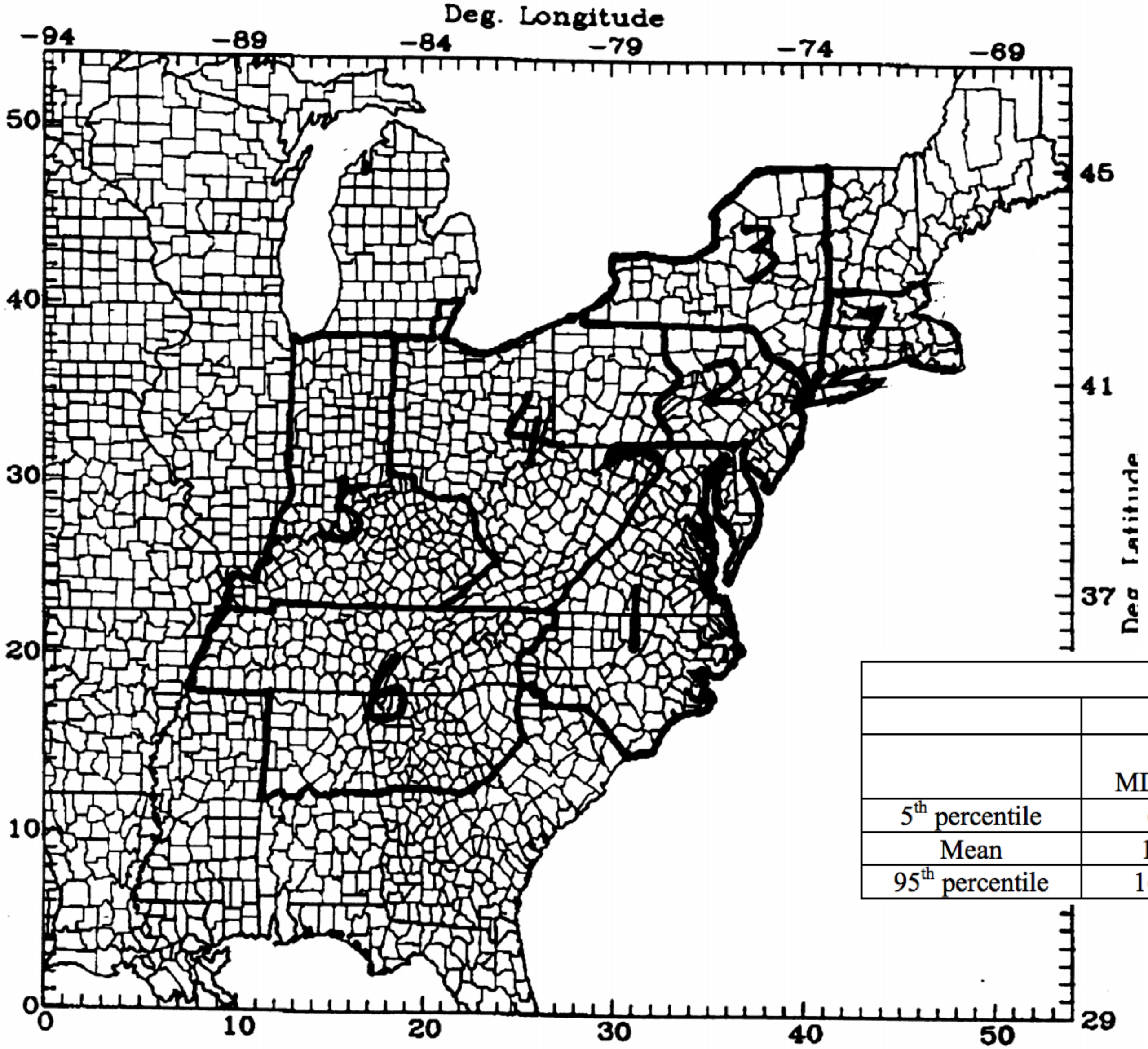
NOTES: These figures display the relationships between satellite-based pollution measurements and monitor based pollution measurements for the 911 census block groups that contain an EPA $PM_{2.5}$ monitor. The blue boxes depict the range of estimates (2.5^{th}–97.5^{th} percentiles) from the satellite-based datasets (y axis) for the given $PM_{2.5}$ level measured by the EPA-AQS monitor (x axis). Source: Authors, Di et al. (2016), van Donkelaar et al. (2019), EPA-AQS.

Mapping exposure to damages to $ is only half the battle!



Zhang et al. 2019

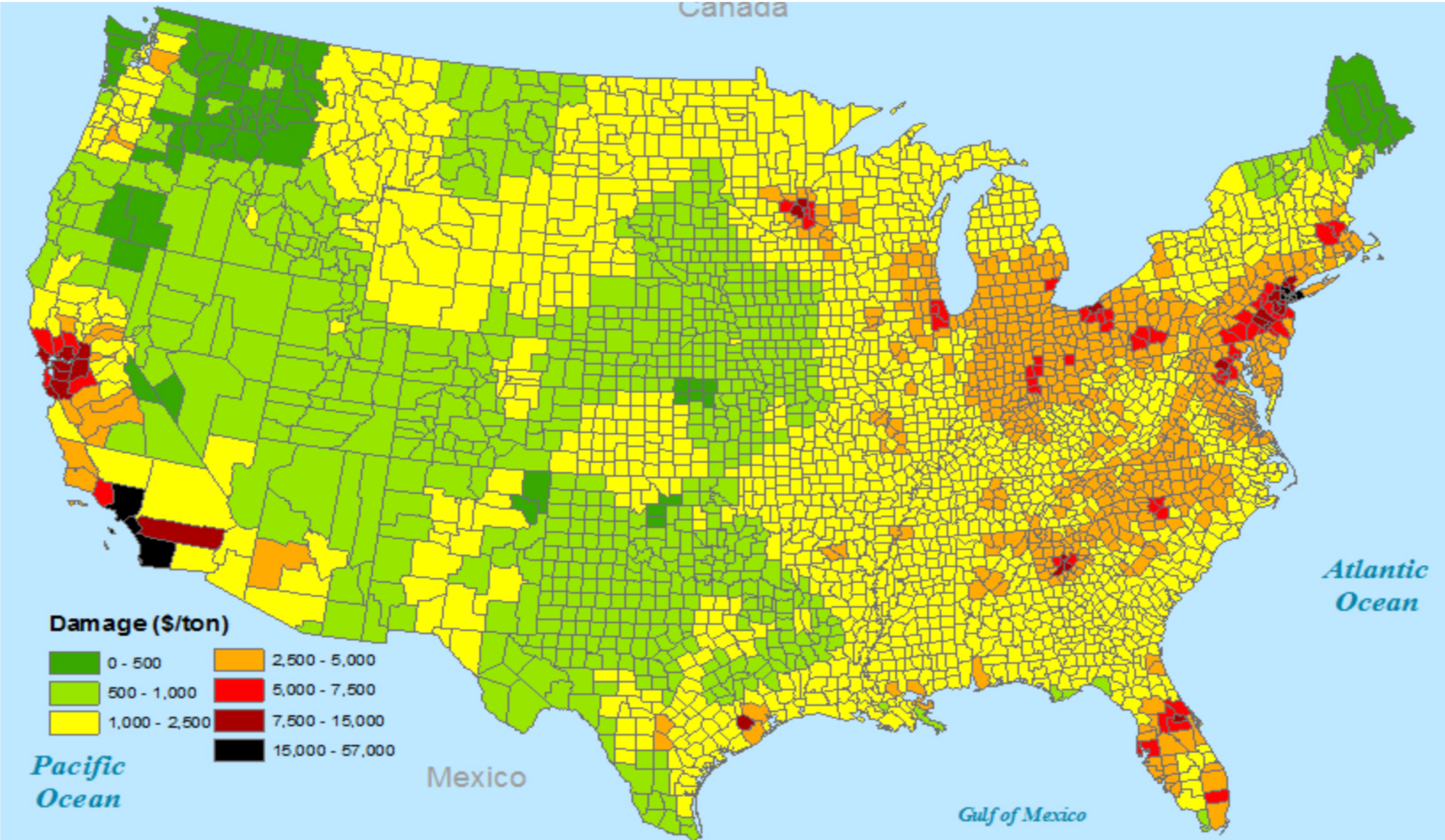
Ye olde fashioned integrated assessment model (Krupnick et al. 1998)



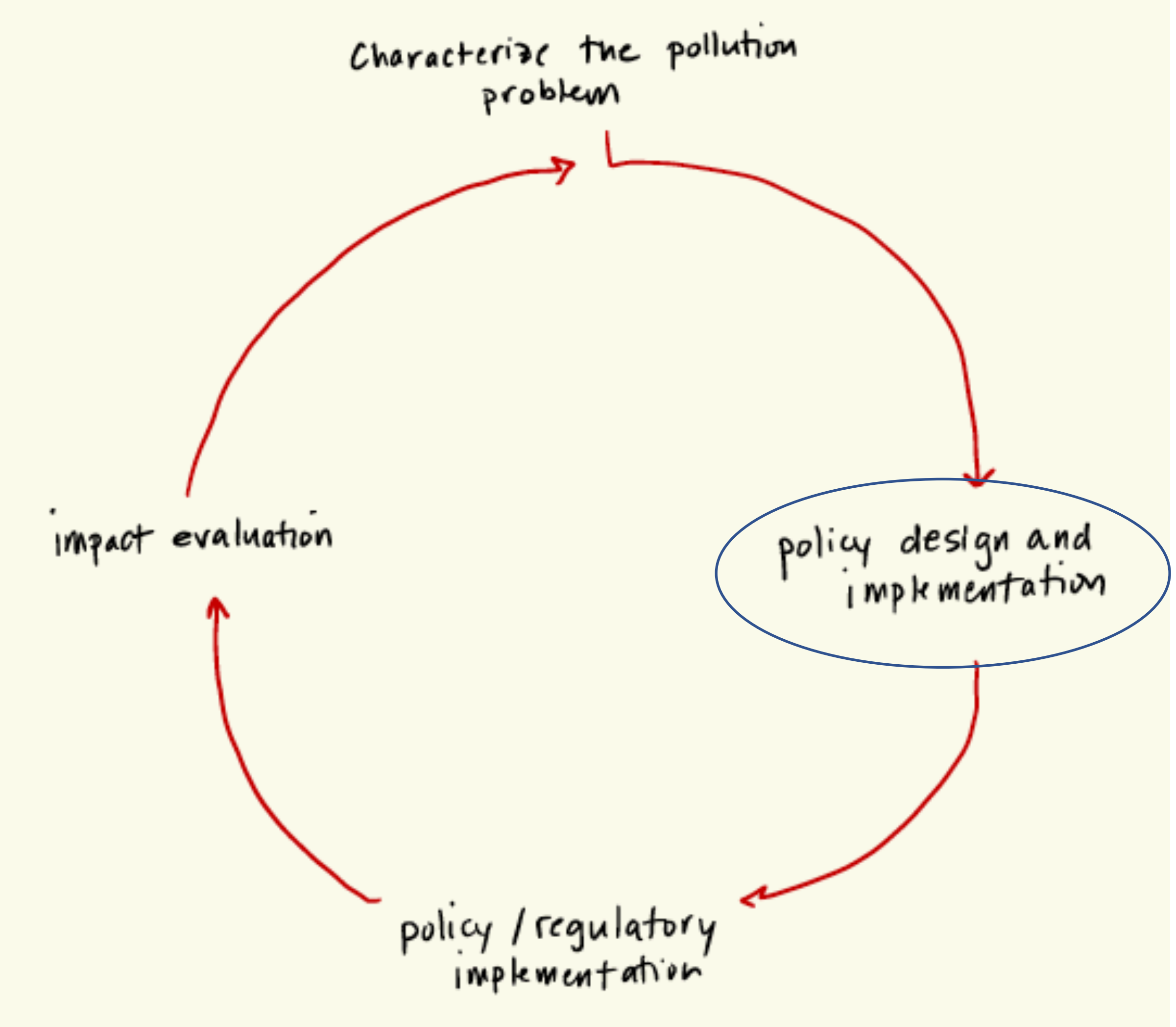
Monetary Benefits per Ton of Abatement by Region (1989 dollars)

Point Sources						
	Source Regions					
	MD-VA	E. PA-N.J.	NY	OH-W. PA-WV	IND-KY	TN-South
5th percentile	61	64	75	91	61	61
Mean	141	150	176	213	143	141
95th percentile	1656	1761	2062	2501	1671	1656

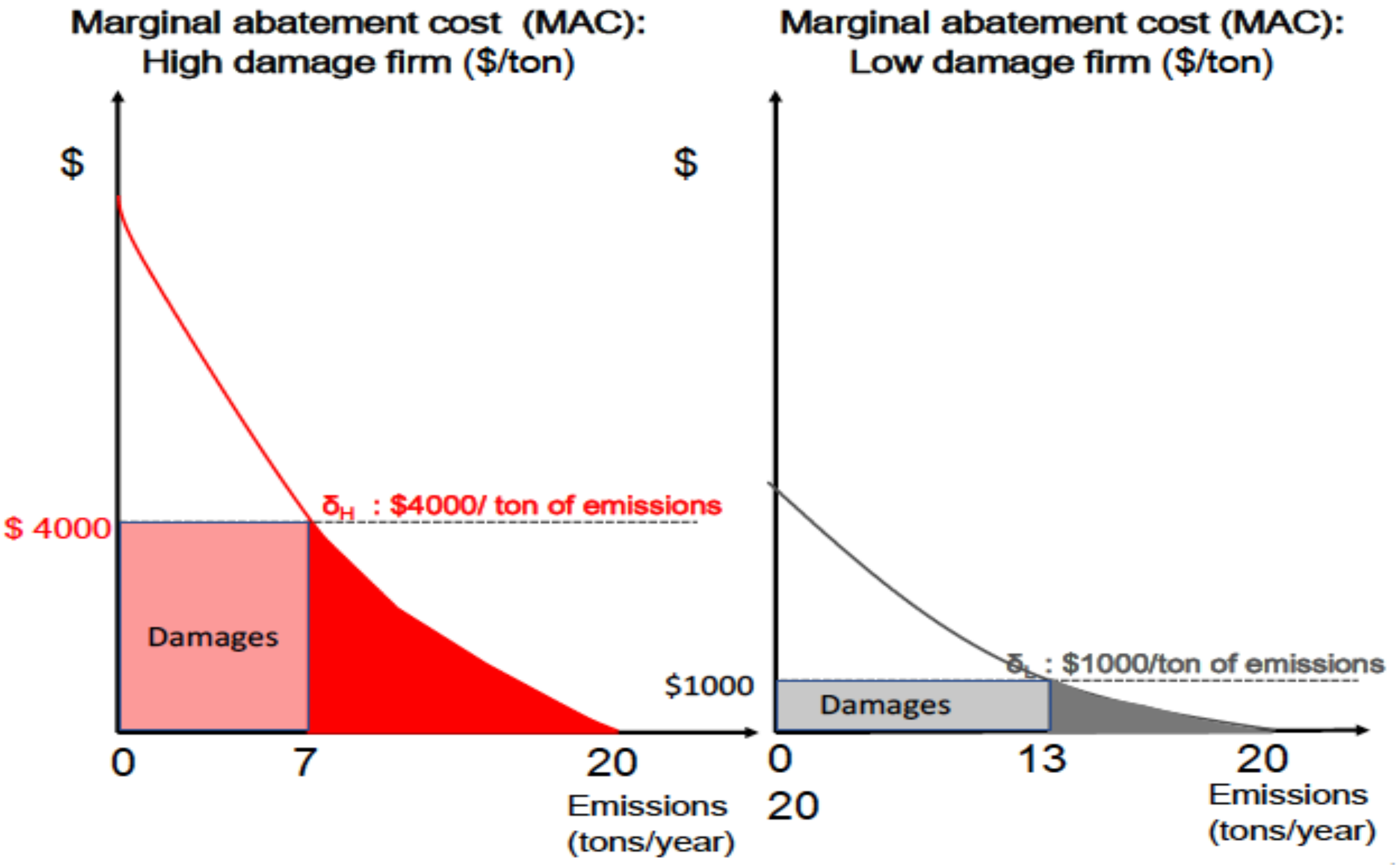
APEEP/AP2/A3 integrated assessment model



Policy design and implementation implications?



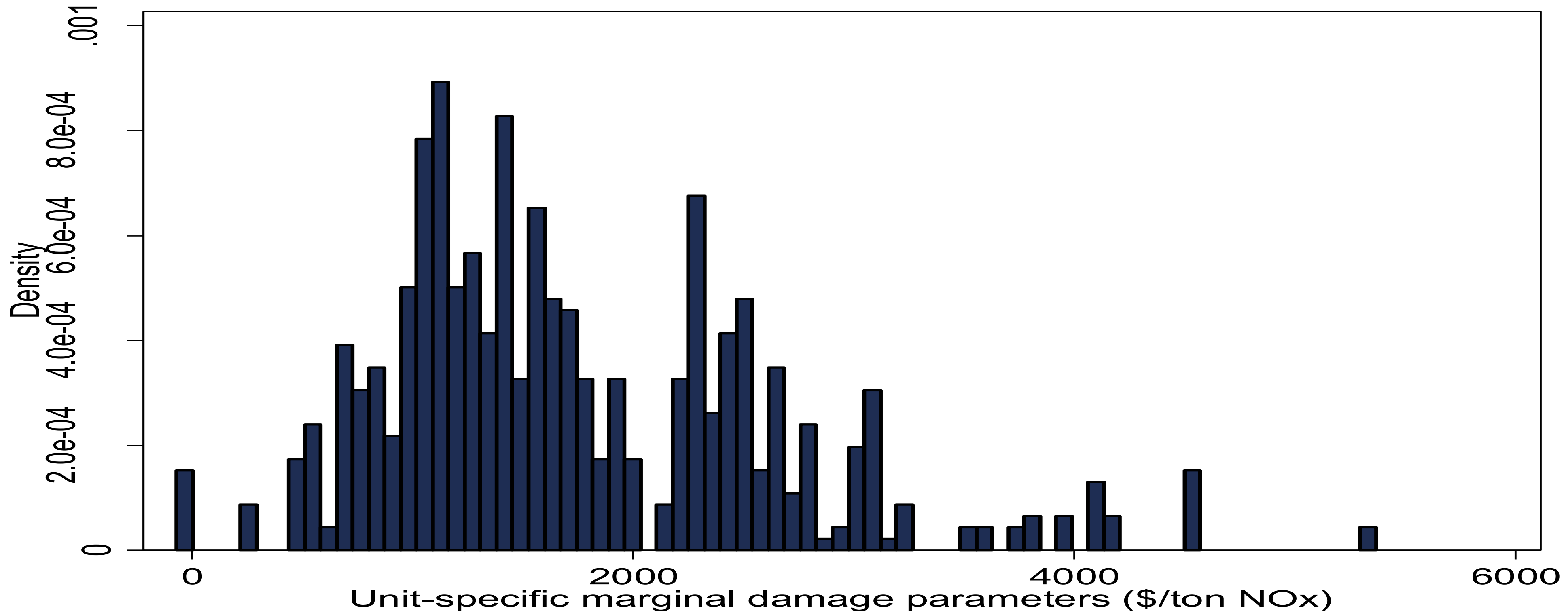
What does an economist want to do with source-specific damages?



Market-Based Emissions Regulation When Damages Vary across Sources: What Are the Gains from Differentiation?

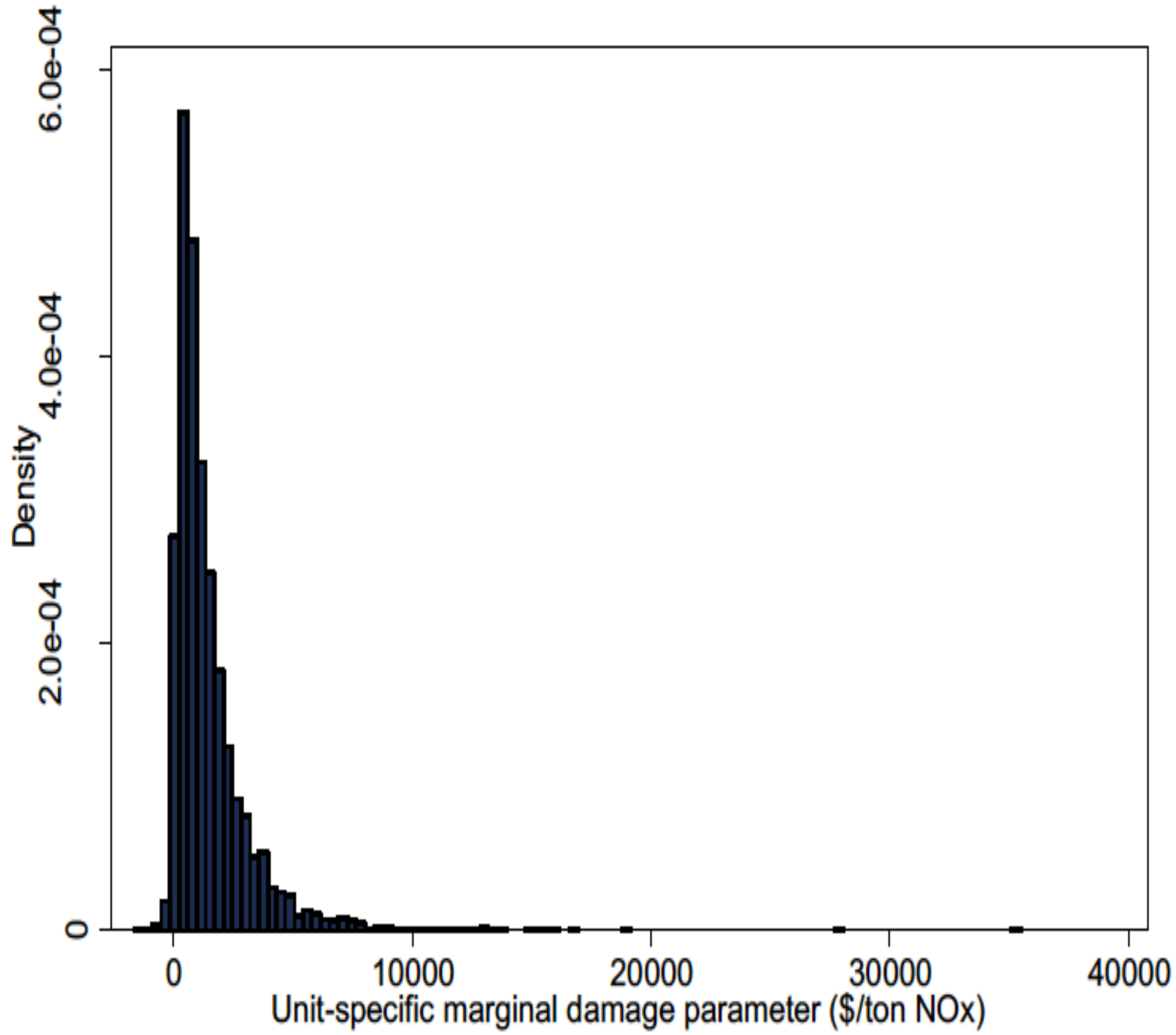
Meredith Fowlie and **Nicholas Muller**

RECEIVED: Feb 11, 2017 **ACCEPTED:** June 25, 2018 **ONLINE:** Mar 26, 2019



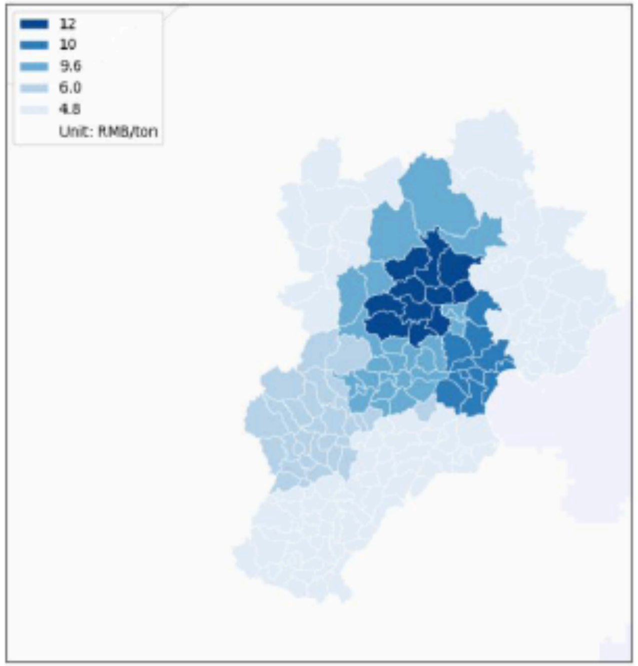
Some real-world complications….

- Differentiated policy designs welfare dominate under certain and complete information.
- Once uncertainty is introduced, the welfare implications of policy differentiation much more ambiguous.
- Uncertainty about abatement costs further complicates design of quantity-based instruments!



Policy experimentation in progress.....

“Community based” response to PM hotspots
California



First differentiated pollution tax
China

How can we best use the data and methods we’re talking about today to inform and improve the policy designs that can change outcomes?

Data -> policy change…

- Exciting innovations on multiple fronts – data collection, data analytics, policy incentive design, policy implementation.

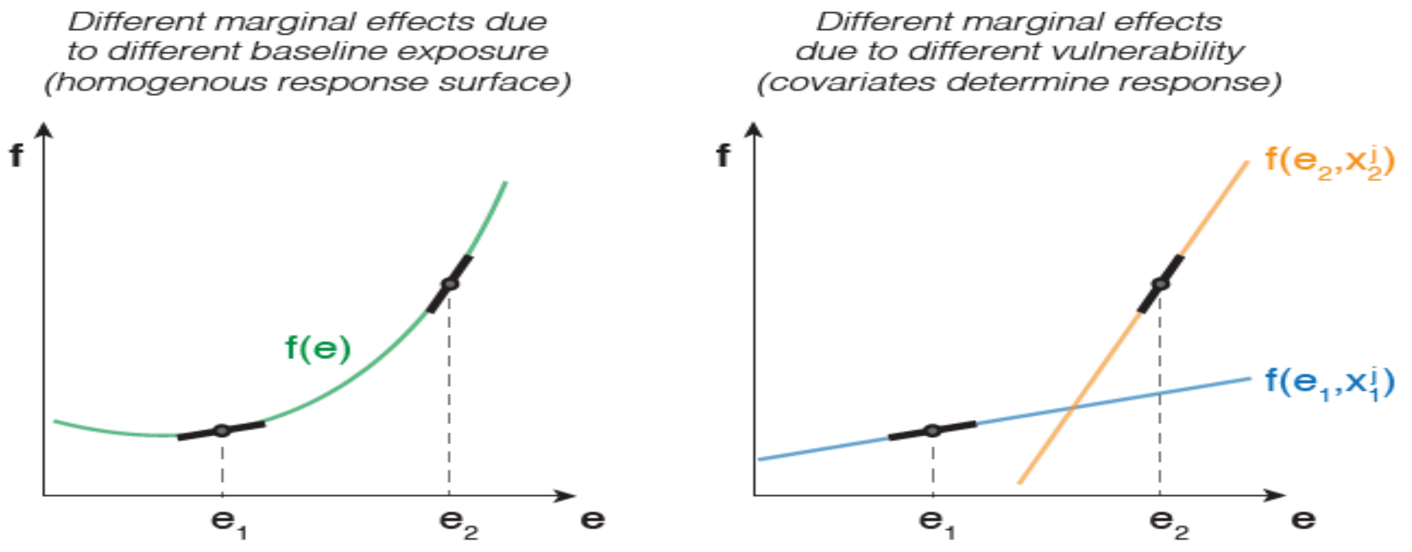
- Substantive interactions between data/methods/policy/communication types increases the chances that research can respond to policy needs and policy can respond to research insights.

- Great job, Ed Rubin et al.

Research -> policy change -> impact

Mapping exposures to impacts is complicated!

- Non-linear concentration-response (C-R) function?
- Differences in the underlying stock of health and/or defensive investments?
- Differences in composition?!



Which PM2.5 effects count?

- Short-run studies relate day-to-day changes in PM2.5 concentrations to changes in mortality rates several days following.
- Longer-run studies examine the relationships between PM2.5 exposure over multiple years and annual mortality rates that have been adjusted for individual level risk factors.

“When choosing between using short-term studies or longer-run studies for estimating mortality benefits.. it is essential to use the cohort studies to capture the important effects”

Regulatory Impact Analysis for the Final Revisions to the National Ambient Air Quality Standards for Particulate Matter, February 2013.

TABLE 10. Co-benefits of the MATS in 2016 (billions of 2007$)

Effect	Pollutant	Total 3%	Total 7%
Adult premature death (Pope)†	$PM_{2.5}$	$34	$30
Adult premature death (Laden)‡	$PM_{2.5}$	$87	$78
Infant premature deaths	$PM_{2.5}$	$0.2	$0.2
Chronic bronchitis	$PM_{2.5}$	$1.4	$1.4
Nonfatal heart attacks	$PM_{2.5}$	$0.5	$0.4
Hospital admissions (respiratory and cardiovascular)	$PM_{2.5}$	$0.04	$0.04
Minor restricted activity days	$PM_{2.5}$	$0.2	$0.2
CO_2-related co-benefits	CO_2	$0.36	$0.36
Total using Pope		$37	$33
Total using Laden		$90	$81

†Estimates from Pope et al. [12].

‡Estimates from Laden et al. [15].

Source: Regulatory Impact Analysis, Tables ES-3 and ES-4.1.

Direct mercury benefits of MATS in 2016

(*millions* of 2007%)

3%	$4.2–$6.2
7%	$0.47–$1.0

Noisy welfare gains!

	Tax
Change in annual costs ($M)	$13
Change in avoided annual damages ($M)	-$60 ($6, $143)
Net gains from differentiation (annual in $M)	$47 (-$7, $130)

2.5 and 97.5 percentile realizations of damage differences in parentheses.
Basis for comparison is undifferentiated policy