The Evolving Role of Causal Inference Methods for Informing Air Quality Policy

Corwin M. Zigler

Department of Statistics and Data Sciences
Department of Women's Health
University of Texas at Austin and Dell Medical School

January 7, 2020

Play/Pause

Public Health/Policy Question of Massive Scale

- Air pollution is a leading cause of global disease burden

Public Health/Policy Question of Massive Scale

- Air pollution is a leading cause of global disease burden
- U.S. regulations to limit exposure to air pollution
  - 58% to 80% of monetized benefits of all federal regulations
  - 44% to 54% of monetized costs
  - $\approx$ \$65 billion annually

Public Health/Policy Question of Massive Scale

- Air pollution is a leading cause of global disease burden
- U.S. regulations to limit exposure to air pollution
  - 58% to 80% of monetized benefits of all federal regulations
  - 44% to 54% of monetized costs
  - $\approx$ \$65 billion annually

Statistical methods for causal inference have emerged at the center of a very current and contentious debate about how maintain air quality policies npr

SIGN IN NPR SHOP DONATE

NEWS ARTS & LIFE MUSIC SHOWS & PODCASTS SEARCH

ENVIRONMENT

EPA Science Panel Considering Guidelines That Upend Basic Air Pollution Science

Listen now · 3:38 + QUEUE

March 28, 2019 · 6:34 PM ET

Heard on Morning Edition

npr EPA Science Panel Considering Guidelines That Upend Basic Air Pollution Science • 3:38

"[Committee] members have varying opinions on the adequacy of the evidence supporting the EPA's conclusion that there is a causal relationship between [particulate matter] exposure and mortality," said Cox, reading from the committee's draft recommendations before explaining that he is "actually appalled" at the lack of scientific evidence connecting particulate pollution to premature death.

EPA Science Panel Considering Guidelines That Upend Basic Air Pollution Science • 3:38

"[Committee] members have varying opinions on the adequacy of the evidence supporting the EPA's conclusion that there is a causal relationship between [particulate matter] exposure and mortality," said Cox, reading from the committee's

of

EPA Science Panel Considering Guidelines That Upend Basic Air Pollution Science • 3:38

overstates the scientific certainty around air pollution.

"If we don't know that X causes Y, then we should say we don't know," said Cox, who consults and lectures about various risk-related topics. He expressed concern that the EPA would move to reduce air pollution under the erroneous assumption that it would result in fewer premature deaths.

npr EPA Science Panel Considering Guidelines That Upend Basic Air Pollution Science • 3:38

"[Committee] members have varying opinions on the adequacy of the evidence supporting the EPA's conclusion that there is a causal relationship between [particulate matter] exposure and mortality," said Cox, reading from the committee's ... of

npr EPA Science Panel Considering Guidelines That Upend Basic Air Pollution Science • 3:38

overstates the scientific certainty around air pollution.

"If we don't know that X causes Y, then we should say we don't know," said Cox, who consults and lectures about various risk-related topics. He expressed concern that the

npr EPA Science Panel Considering Guidelines That Upend Basic Air Pollution Science • 3:38

explains, because no one study captures everything about a given pollutant.

"You can't randomize millions of [people] around the world to breathe higher pollution or lower pollution, so we have to rely on observational data," Dominici says.

POLICY FORUM | SCIENCE AND REGULATION

Don't abandon evidence and process on air pollution policy

Gretchen T. Goldman[1], Francesca Dominici[2]
+ See all authors and affiliations

Science 29 Mar 2019:
Vol. 363, Issue 6434, pp. 1398-1400
DOI: 10.1126/science.aaw9460

Air pollution kills—scientists have known this for many years. But how do they know? The global scientific community has developed and agreed upon a framework that draws on multiple lines of evidence across different scientific disciplines to assess the existence and strength of links between air pollution and health. In the United States, federal policies require use of this science-based framework to ensure that air pollution standards protect the public's health. But now this science-based policy process—and public health—are at risk. Recent developments at the U.S. Environmental Protection Agency (EPA) stand to quietly upend the time-tested and scientifically backed process the agency relies on to protect the public from ambient air pollution (*1*). One of these developments—changes in how the EPA handles causality between air pollutants and health effects—has received less attention but, if enacted, would alter the approach that the EPA has used for more than a decade to set health-based air pollutant standards. At the March meeting of the EPA's Clean Air Scientific Advisory

SHARE

POLICY FORUM | SCIENCE AND REGULATION

Don't abandon evidence and process on air pollution policy

Gretchen T. Goldman[1], Francesca Dominici[2]
+ See all authors and affiliations

Science 29 Mar 2019:
Vol. 363, Issue 6434, pp. 1398-1400
DOI: 10.1126/science.aaw9460

Science
Vol 363, Issue 6434
29 March 2019

ARTICLE TOOLS
Email · Print · Save to my folders · Alerts · Request Permissions · Citation tools · Share

Article | Figures & Data | Info & Metrics | eLetters | PDF

Air pollution kills—scientists have known this for many years. But how do they know? The global scientific community has developed and agreed upon a framework that draws on multiple lines of evidence across different scientific disciplines to assess the existence and strength of links between air pollution and health. In the United States, federal policies require use of this science-based framework to ensure that air pollution standards protect the public's health. But now this science-bas
developments at the
upend the time-teste
protect the public fr
changes in how the
effects—has receive
the EPA has used fo
standards. At the Ma

PHOTO: TETRA IMAGES/A

epidemiology can be viewed as a reasonable framework to address the general issue of confounding bias in individual studies. New statistical methods for the analysis of epidemiological studies on air pollution and health can inform and improve the EPA's ap-

rel
(e.g
exp
cor
of
and

Air Quality Policy

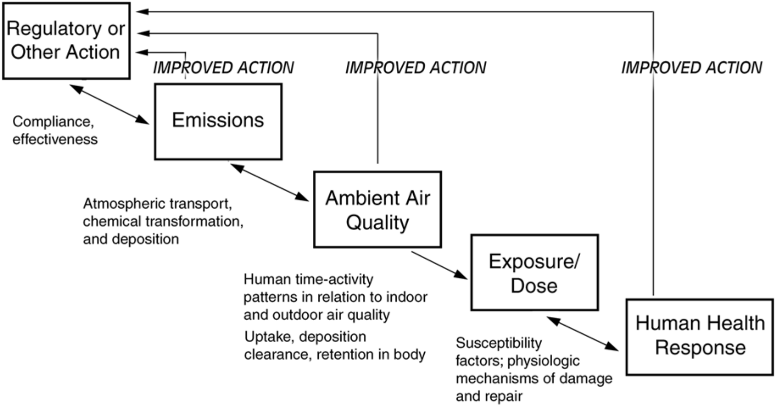
The Role of Data and Statistics

Major Themes of Statistical Methods Development

for air pollution policies and beyond

1. Importance of “causal inference”
2. Distinguishing between types of questions
   - “Health effects” of pollution exposure
   - “Health effects” of policy
3. Intersecting statistics with physical/mechanistic process models
   - E.g., atmospheric science/engineering
4. Methods for interference/networks for interventions that “spillover” or diffuse in time and space

“Chain of Accountability”
(from the Health Effects Institute)

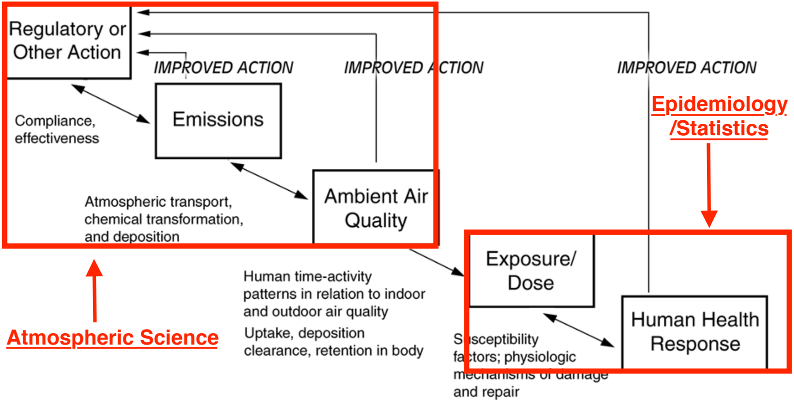


"Chain of Accountability"

(from the Health Effects Institute)

Statistical Methods with **Atmospheric Science**

(have historically been kept separate)



Existing Policy Evaluations Underuse Data and Statistics

lots of engineering, chemistry, exposure science etc.

- Heavy reliance on deterministic physical/chemical model output
  - E.g., CMAQ, GEOS-Chem, etc.
- Focus on prospective predictions
- Not empirically verified using observed data following implementation
- **Opportunities** for marrying the physical/mechanistic understanding with empirical data and statistics
  - Health and environmental data becoming increasingly rich and available

Statistical Methods Directions

Two Types of “Policy” Questions

Two Important Questions for Policy

Causal inference important for both!

Question #1

What are the “health effects” of pollution exposure?

- I.e., *exposure-response* estimation
- Most epidemiology focuses here

Question #2

What are the “health effects” of a specific policy or intervention?

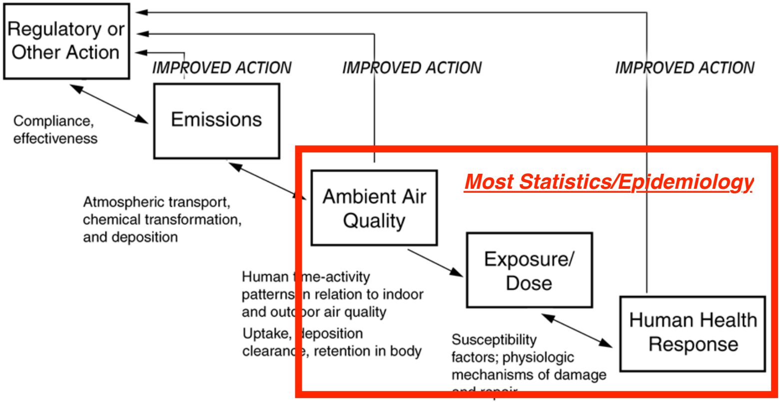
- I.e., *effectiveness of interventions*
- Less focus here
- Particular lack of appropriate statistical methods

Different questions $\leftrightarrow$ Different challenges $\leftrightarrow$ Different methods

Important Question #1

for informing policies

What is the causal effect of exposure to pollution on health?
(i.e.,**exposure-response** function)



Important Question #1
for informing policies

What is the causal effect of exposure to pollution on health?
(i.e.,**exposure-response** estimation)

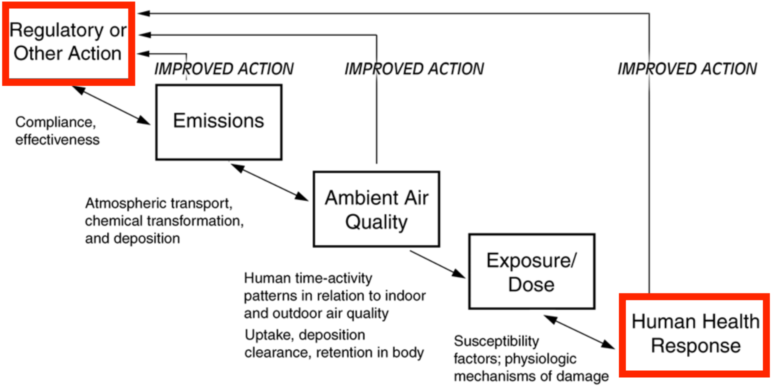
- Essentially the focus of air pollution epidemiology for decades
- Many open **statistical** problems
  - Multipollutant mixture exposure-response functions
  - Health effects at low exposures
  - Spatial confounding, overlap, effect heterogeneity, etc.
- Broadly familiar “causal inference” problems
  - Ongoing work in statistics, epidemiology, econometrics, etc.

Important Question #2

for informing policies

Did a specific policy causally affect health?

(i.e., **effectiveness** of a policy)



Important Question #2
for informing policies

Did a specific policy causally affect health?
(i.e., **effectiveness** of a policy)

- To what extent does a particular action *cause* health improvements?
- Which effects can be attributed to which policies?
- Which policies are most (cost) effective?

- Comparatively less focus from statisticians/epidemiologists
- **More opportunities** for new statistical methods development

Statistical Challenges for Estimating Air Pollution Policy Effectiveness

Example: Scrubbers on Coal Power Plants

Two Fundamental Features

Generic question: What are the (causal) health effects of an intervention taken to reduce pollution?

Two fundamental features:

1. *Interventions* occur at sources of air pollution
   - E.g., installing emissions controls on power plant smokestacks or enacting new vehicle emissions standards

2. Pollution *moves* from its originating source
   - Long-range pollution transport

Major Pollution Source: Power Plants

$\Rightarrow$ Many regulations to reduce emissions

- Some install “scrubbers” in response to regulations

$\rightarrow$ Reduce SO_2 emissions

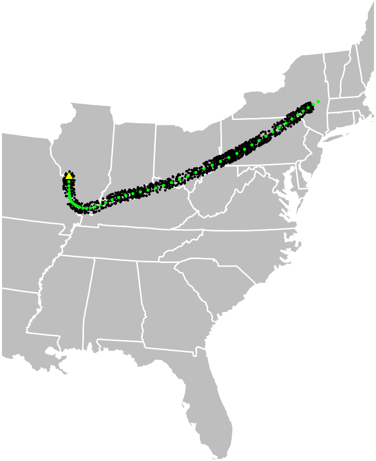
$\rightarrow$ Reduce ambient pollution

$\rightarrow$ Improve health

Example Question: Do scrubbers on coal-fired power plants causally affect health among Medicare beneficiaries?

Long-range Pollution Transport

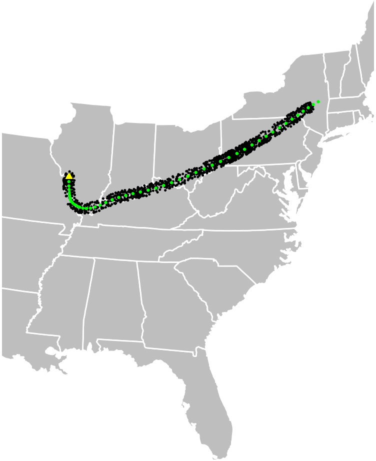
Movement of pollution from sources → populations



- Emissions originating at a power plant (**yellow triangle**) *move*
  - Long distances towards conversion to harmful pollution
- Lots of mechanistic knowledge of the physics/chemistry from atmospheric science fields

Long-range Pollution Transport

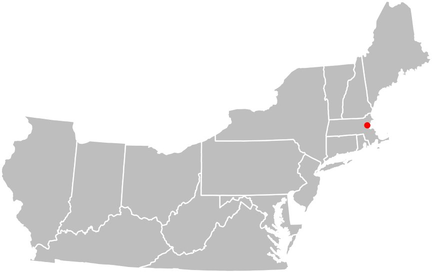
Movement of pollution from sources $\rightarrow$ populations



- Emissions originating at a power plant (**yellow triangle**) *move*
  - Long distances towards conversion to harmful pollution
- Lots of mechanistic knowledge of the physics/chemistry from atmospheric science fields
- Intervening at a power plant impacts health at *many* zip codes (**black dots**)
- $\Rightarrow$ inference on a “network” of interconnected power plants/population locations (**interference**)

Potential Outcomes amid Long-Range Pollution Transport

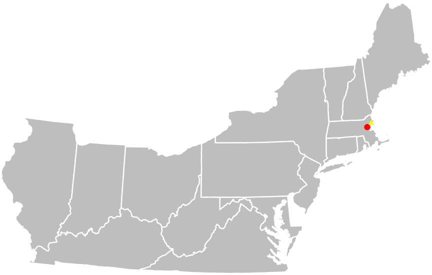
$Y \equiv$ health, $A \equiv$ treatment



Zip Code $j = 02138$ (**red dot**)
$Y_j(?) \equiv$ Health outcome that would be observed...

Potential Outcomes amid Long-Range Pollution Transport

$Y \equiv$ health, $A \equiv$ treatment

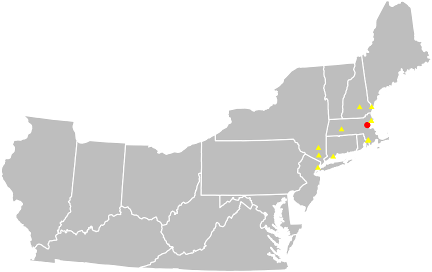


Zip Code $j = 02138$ (**red dot**)
$Y_j(A_i) \equiv$ Health outcome that would be observed...

if the *most influential (or closest)* power plant were intervened upon

Potential Outcomes amid Long-Range Pollution Transport

$Y \equiv$ health, $A \equiv$ treatment

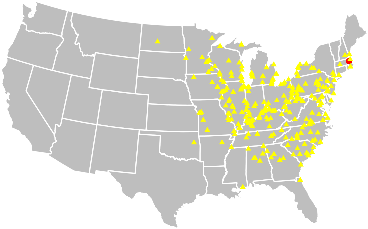


Zip Code $j = 02138$ (**red dot**)
$Y_j(A_1, \ldots, A_{10}) \equiv$ Health outcome that would be observed...

if the10 most influential power plants were intervened upon

Potential Outcomes amid Long-Range Pollution Transport

$Y \equiv$ health, $A \equiv$ treatment



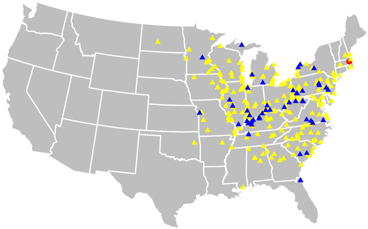
Zip Code $j = 02138$ (**red dot**)
$Y_j(A_1, \ldots, A_{278}) \equiv$ Health outcome that would be observed...

under a particular intervention allocation to all 278 plants that influence this location

Treating at any of these 278 plants affects health at 02138

Potential Outcomes amid Long-Range Pollution Transport

$Y \equiv$ health, $A \equiv$ treatment



Zip Code $j = 02138$ (**red dot**)
$Y_j(A_1, \ldots, A_{278}) \equiv$ Health outcome that would be observed...

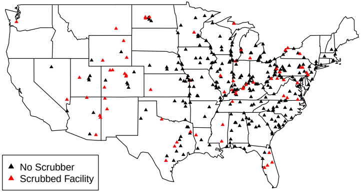
under a particular scrubber allocation to all 278 plants that influence this location

Treating any of these 278 plants affects health at 02138
48 observed to have intervention

Intervention Effects Propagate Across a Network

Network of 100s of power plants and 10Ks of population locations

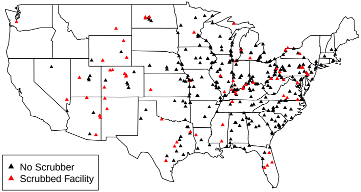
Interventions at Power Plants:
"Scrubber" Installation



Intervention Effects Propagate Across a Network

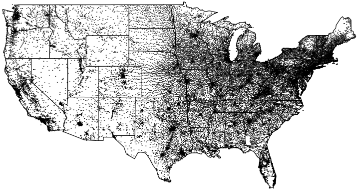
Network of 100s of power plants and 10Ks of population locations

Interventions at Power Plants: "Scrubber" Installation



Affect
⟶

Health at Population Locations (Zip Codes)



Interference Due to *Treatment Diffusion*

Interference introduces new causal quantities of interest:

- More than just accounting for or addressing structure of variability (e.g., spatial correlation)
- Budding research on interference in specialized settings
  - Infectious diseases
  - Social networks

- Here, interference arises due to **treatment diffusion** resulting in complex exposure dependencies
  - Governed by atmospheric transport of pollution
  - Treatment diffuses across a network
- **New methods** that are statistical/empirical, but make use of extant knowledge from atmospheric engineering

Statistical Challenges for Estimating Air Pollution Policy Effectiveness

Some Specific Statistical Methods Areas

Statistical Methods Challenges

- Causal inference with *interference*
  - Health outcomes depend on treatments applied at *many* different units
- Treatment diffusion across spatial networks

Statistical Methods Challenges

- Causal inference with *interference*
  - Health outcomes depend on treatments applied at *many* different units
- Treatment diffusion across spatial networks
- Mechanistic understanding of pollution transport
  - Leverage knowledge from atmospheric science

Statistical Methods Challenges

- Causal inference with *interference*
  - Health outcomes depend on treatments applied at *many* different units
- Treatment diffusion across spatial networks
- Mechanistic understanding of pollution transport
  - Leverage knowledge from atmospheric science
- Causal inference + spatial statistics

Statistical Methods Challenges

- Causal inference with *interference*
  - Health outcomes depend on treatments applied at *many* different units
- Treatment diffusion across spatial networks
- Mechanistic understanding of pollution transport
  - Leverage knowledge from atmospheric science
- Causal inference + spatial statistics
- Disentangling effects of multiple concurrent/overlapping policies
- Interventions impacting multiple pollutants (mixtures) simultaneously
- Complex mixtures of exposures

Closing Thoughts

Summary and Conclusions

Summary: Data Science for Evaluating Environmental Policies

- Policies of enormous consequence (huge costs and benefits)
- High quality data are increasingly available
- Particular focus on the role of causal inference methods
- Significant challenges that are unique(?) to air pollution
    - E.g., long-range pollution transport

- Key opportunities for statistical methods development:
    - Causal inference $+$ spatial statistics
    - Statistics $+$ mechanistic atmospheric models
    - Networks/interference