Estimating Causal Effects of Air Quality Regulations Using Principal Stratification for Spatially-Correlated Multivariate Intermediate Outcomes

Corwin M. Zigler

Department of Biostatistics, Harvard School of Public Health

May 24, 2012

Air Quality, Health, and Regulation

Accountability Assessment

- **Long term** exposure to air pollution is bad for health.
- EPA estimates $\approx$ \$25 billion per year on air quality management.
    - 1970 Clean Air Act.
- For a **specific** regulatory action:
    - What were the causal effects on air quality?
    - What were the causal effects on health?

1990 Clean Air Act Amendments (CAAA)

EPA designates **counties** as:

1. Attainment of air quality standards for PM_{10} .
2. Nonattainment of air quality standards for PM_{10} :
    - Required **states** to implement plans to achieve standards.

What were the **causal effects** of the 1990 nonattainment designations for PM_{10} on:

- Pollution
    - Ambient concentrations of PM_{10} and O_3 in 1999-2001.
- Health
    - All-cause Medicare mortality in 2001.

Causal Inference for Air Quality Regulations

1. Potential outcomes in the EPA regulatory environment.

Causal Inference for Air Quality Regulations

1. Potential outcomes in the EPA regulatory environment.
2. Air quality is a **posttreatment concomitant variable.**
   - $\Rightarrow$ Principal stratification.

Causal Inference for Air Quality Regulations

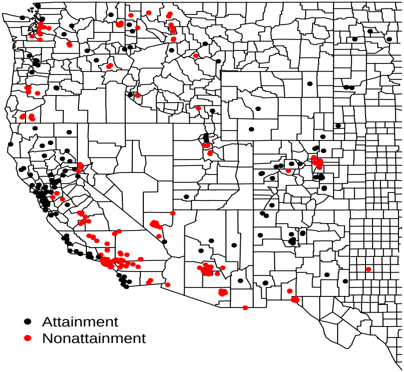
1. Potential outcomes in the EPA regulatory environment.
2. Air quality is a **posttreatment concomitant variable.**
   - ⇒ Principal stratification.
3. Regulations affect **multiple pollutants.**
   - ⇒ Multivariate continuous intermediate variable.

Causal Inference for Air Quality Regulations

1. Potential outcomes in the EPA regulatory environment.
2. Air quality is a **posttreatment concomitant variable.**
   - $\Rightarrow$ Principal stratification.
3. Regulations affect **multiple pollutants.**
   - $\Rightarrow$ Multivariate continuous intermediate variable.
4. Pollution is **spatially correlated.**
   - Hierarchical spatial model.
5. Interference between observations (no SUTVA).

Observed Regulation Program and Overall Causal Effect

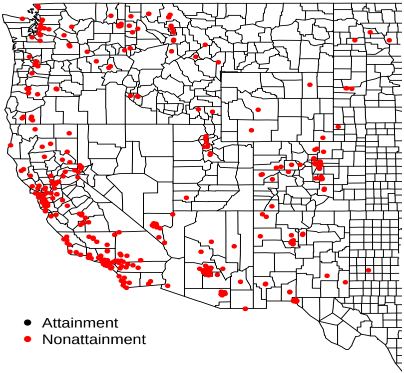
Observed Regulation Program



Observational unit: pollution monitor (point-referenced data)

Observed Regulation Program and Overall Causal Effect

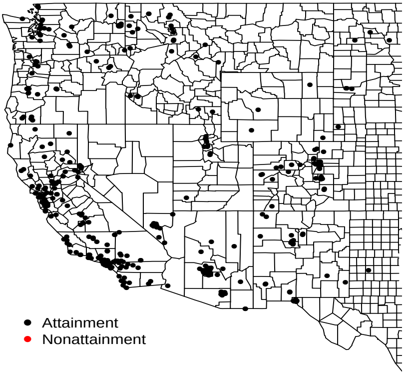
Overall Causal Effect



EPA regulates all locations

Observed Regulation Program and Overall Causal Effect

Overall Causal Effect



No EPA Action

Potential Outcomes

Regulation program vector: $\mathbf{A} = [A(s_i)]_{i=1}^{n}$

- $n = 362$ locations.
- $A(s_i) = 1 \Rightarrow i^{th}$ location nonattainment.
- Specific regulation program $\mathbf{A} = \mathbf{a}$.

Potential Outcomes

Regulation program vector: $\mathbf{A} = [A(s_i)]_{i=1}^{n}$

- $n = 362$ locations.
- $A(s_i) = 1 \Rightarrow i^{th}$ location nonattainment.
- Specific regulation program $\mathbf{A} = \mathbf{a}$.

Potential Outcomes

- Pollution $X_{\mathbf{a}}(s)$: vector of pollutant concentrations.
    - PM_{10} and O_3 .
- Mortality $Y_{\mathbf{a}}(s)$: all-cause mortality among Medicare beneficiaries living near a monitor.
    - $\approx$ 7 million people aged 65+.

Spatial Correlation $\rightarrow$ Interference

- Typical assumption of no interference (SUTVA) likely **violated**.
  - Regulations can affect air quality in other areas.

Spatial Correlation $\rightarrow$ Interference

- Typical assumption of no interference (SUTVA) likely **violated**.
  - Regulations can affect air quality in other areas.
- Full interference: every location interferes with every other location.
  - 2^n potential outcomes for each location.

Spatial Correlation $\rightarrow$ Interference

- Typical assumption of no interference (SUTVA) likely **violated**.
  - Regulations can affect air quality in other areas.
- Full interference: every location interferes with every other location.
  - 2^n potential outcomes for each location.
- Partial interference.
  - Some locations interfere with some other locations.
  - How to define the interference groups?

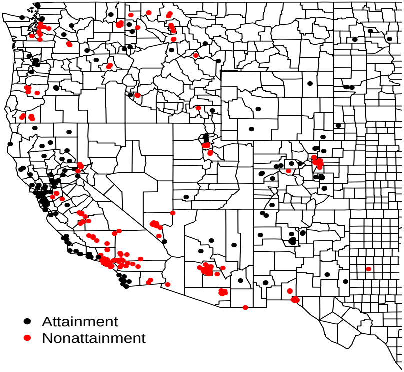
Never Observe Potential Outcomes Under Regulation Programs of Interest

			Observed Regulation		No Regulation		Full Regulation	
			$\mathbf{A} = \mathbf{a}^{\mathbf{obs}}$		$\mathbf{A} = \mathbf{0}$		$\mathbf{A} = \mathbf{1}$	
	i	$A^{obs}(s_i)$	$X_{\mathbf{a}^{\mathbf{obs}}}(s_i)$	$Y_{\mathbf{a}^{\mathbf{obs}}}(s_i)$	$X_0(s_i)$	$Y_0(s_i)$	$X_1(s_i)$	$Y_1(s_i)$
Observed	1	0	*obs*	*obs*				
Attainment	2	0	*obs*	*obs*				
(black dots)	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Observed nonattainment	$n-1$	1	*obs*	*obs*				
(red dots)	n	1	*obs*	*obs*				

Never Observe Potential Outcomes Under Regulation Programs of Interest

			Observed Regulation $\mathbf{A} = \mathbf{a}^{\mathbf{obs}}$		No Regulation $\mathbf{A} = \mathbf{0}$		Full Regulation $\mathbf{A} = \mathbf{1}$	
	i	$A^{obs}(s_i)$	$X_{\mathbf{a}^{\mathbf{obs}}}(s_i)$	$Y_{\mathbf{a}^{\mathbf{obs}}}(s_i)$	$X_0(s_i)$	$Y_0(s_i)$	$X_1(s_i)$	$Y_1(s_i)$
Observed	1	0	*obs*	*obs*	*obs**	*obs**	?	?
Attainment	2	0	*obs*	*obs*	*obs**	*obs**	?	?
(black dots)	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
Observed	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
nonattainment	$n-1$	1	*obs*	*obs*	?	?	*obs**	*obs**
(red dots)	n	1	*obs*	*obs*	?	?	*obs**	*obs**

Assignment Group Interference Assumption (AGIA)



Interference implicit in EPA decisions

$\Downarrow$

Black dots don't interfere with Red dots

Estimation

Models

- Potential pollution outcomes.
  - Multivariate spatial hierarchical model.
- Mortality outcomes, conditional on pollution.
  - Poisson regression.
- Estimation with data augmentation / MCMC.

Estimation

Models

- Potential pollution outcomes.
  - Multivariate spatial hierarchical model.
- Mortality outcomes, conditional on pollution.
  - Poisson regression.
- Estimation with data augmentation / MCMC.

Average Causal Effects

- Expected $\mathcal{K}$-Dissociative Effect:
  - Average effect on mortality in areas where regulation **did not** affect pollution.

Estimation

Models

- Potential pollution outcomes.
    - Multivariate spatial hierarchical model.
- Mortality outcomes, conditional on pollution.
    - Poisson regression.
- Estimation with data augmentation / MCMC.

Average Causal Effects

- Expected $\mathcal{K}$-Dissociative Effect:
    - Average effect on mortality in areas where regulation **did not** affect pollution.
- Expected $\mathcal{K}$-Associative Effect:
    - Average effect on mortality in areas where regulation **decreased** pollution.

Spatial Hierarchical Model

$$X(s) = Z^T(s)\beta + W(s) + \epsilon(s)$$

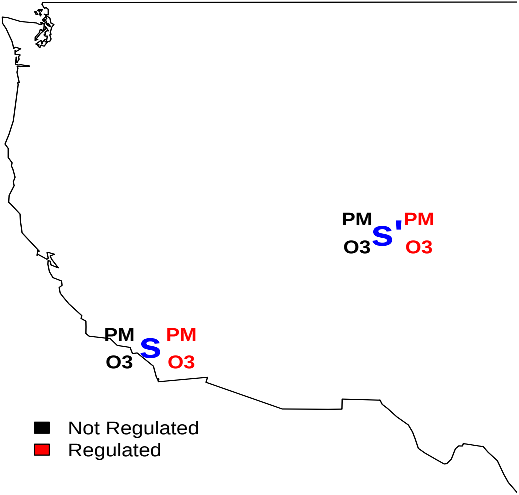
- $s \equiv$ specific location.
- $X(s) \equiv 4-$dimensional vector of pollution concentrations under both regulations $(X_{\mathbf{A}=0}(s), X_{\mathbf{A}=1}(s))$.
- $Z(s) \equiv$ covariates.
- $\epsilon(s) \equiv$ nonspatial ("nugget") error.
- $W(s) \equiv$ **spatially-varying random intercepts.**

Spatial Hierarchical Model

$$X(s) = Z^T(s)\beta + W(s) + \epsilon(s)$$

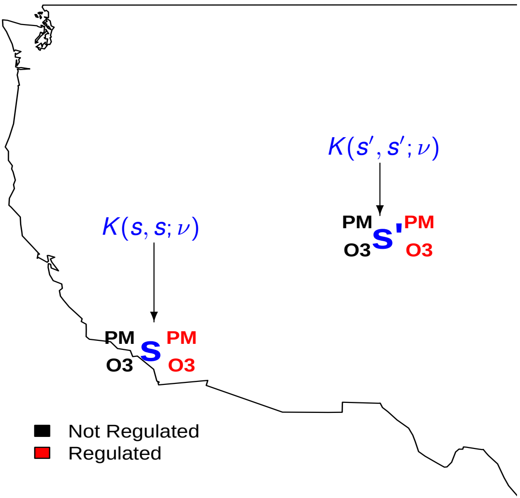
- $s \equiv$ specific location.
- $X(s) \equiv 4-$dimensional vector of pollution concentrations under both regulations $(X_{\mathbf{A}=0}(s), X_{\mathbf{A}=1}(s))$.
- $Z(s) \equiv$ covariates.
- $\epsilon(s) \equiv$ nonspatial ("nugget") error.
- $W(s) \equiv$ **spatially-varying random intercepts.**
  - $W(s) \sim$ Multivariate Gaussian Process (MVGP).
  - Cross-covariance: $K(s, s'; \nu)$.
    - ν governs spatial decay and smoothness.

$K(s, s'; \nu)$

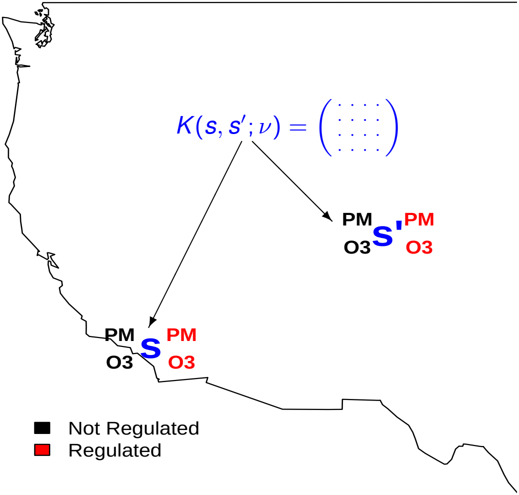


$K(s, s'; \nu)$

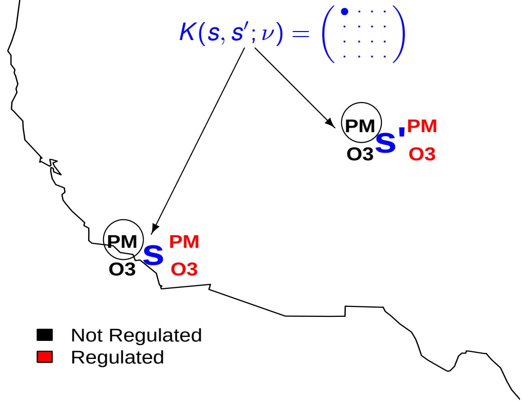
$K(s, s; \nu) = K(s', s'; \nu)$
$\Rightarrow$ stationary



$K(s, s'; \nu)$

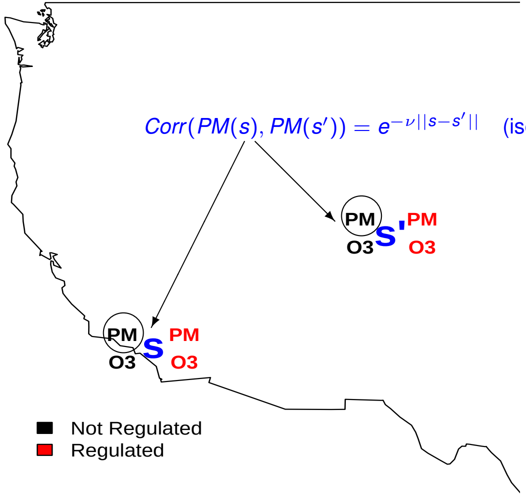

$K(s, s'; \nu) = \begin{pmatrix} \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{pmatrix}$
PM
O3
S'
PM
O3
PM
O3
S
PM
O3
Not Regulated
Regulated

$K(s, s'; \nu)$



$K(s, s'; \nu)$

$$Corr(PM(s), PM(s')) = e^{-\nu||s-s'||} \quad \text{(isotropic)}$$



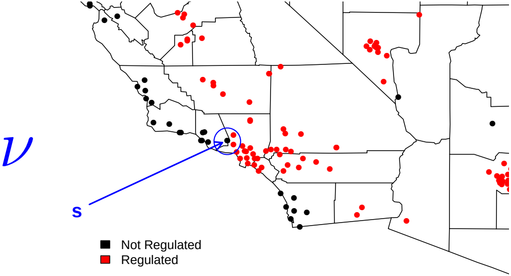
Specifying Spatial Structure

1. Specify $K(s, s; \nu)$
   - Covariance matrix for potential pollution concentrations within a location.
   - Nonidentifiability $\Rightarrow$ sensitivity parameter.
2. Specify spatial decay for each individual pollution concentration.
   - Separate isotropic exponential decay functions for each pollutant.
3. Combine 1. and 2. $\Rightarrow$ cross-covariance function for MVGP.
   - Computational feasibility.
   - Isoloate nonidentifiable associations between potential outcomes.

Why a Spatial Model?

Predicting potential outcomes:

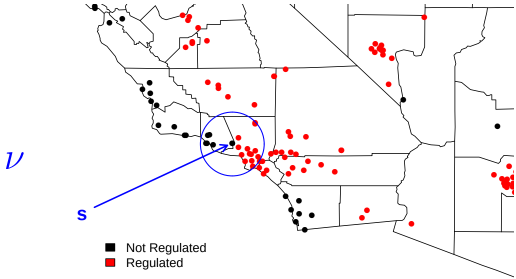
- Predict missing potential outcomes at s using information at surrounding locations.
- Use estimates of ν to assess interference assumption.



Why a Spatial Model?

Predicting potential outcomes:

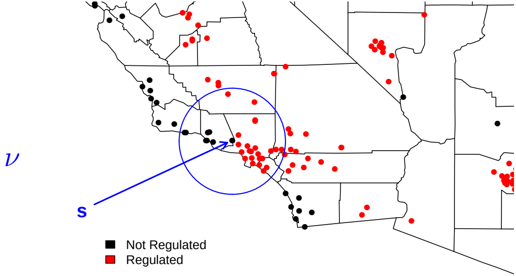
- Predict missing potential outcomes at s using information at surrounding locations.
- Use estimates of ν to assess interference assumption.



Why a Spatial Model?

Predicting potential outcomes:

- Predict missing potential outcomes at s using information at surrounding locations.
- Use estimates of ν to assess interference assumption.



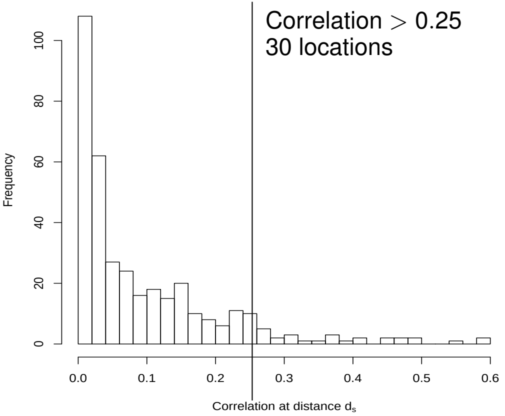
Assess Interference Assumption

$\hat{\nu}$ has implications of interference

- $\hat{\nu} \Rightarrow$ estimated correlation between measurements at two locations.
- Examine correlations between observations assumed not to interfere.
  - Substantial correlation $\Rightarrow$ violation of AGIA.

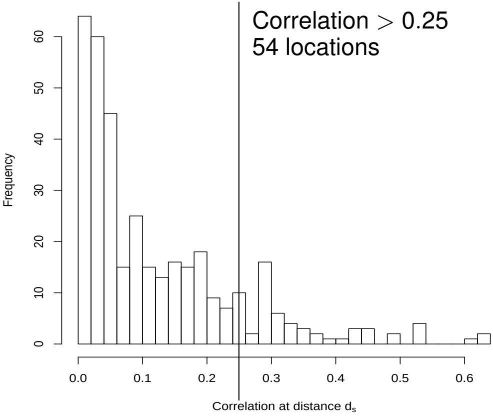
Assessment of AGIA for PM_{10}

Figure: PM_{10} , $\hat{\nu} = 3.13$



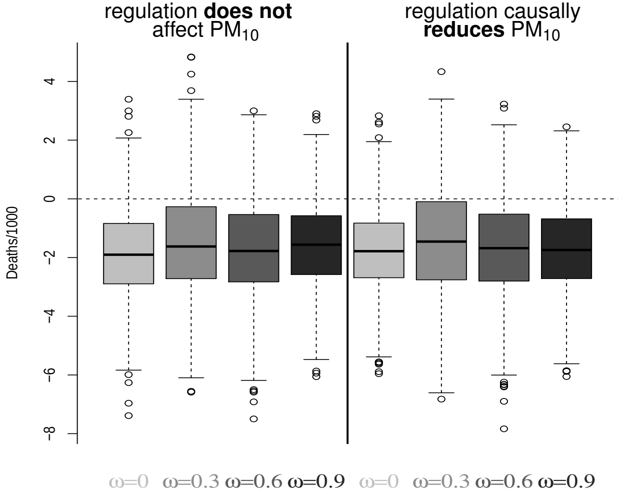
Assessment of AGIA for O_3

Figure: O_3 , $\hat{\nu} = 2.68$



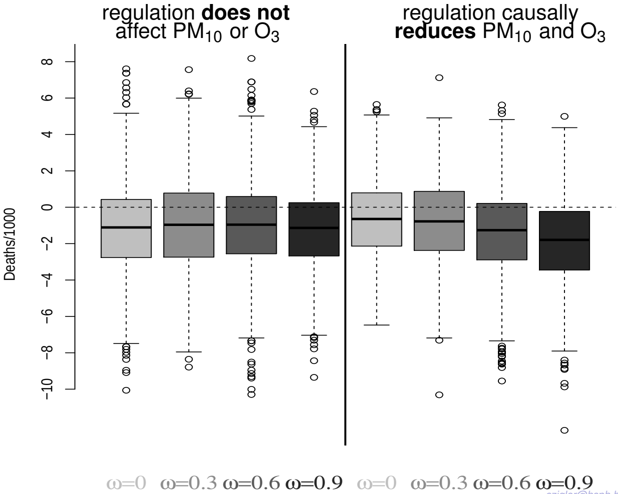
Associative and Dissociative Effects for PM_{10}

Causal effect on heath in areas where:



Associative and Dissociative Effects for joint effect on both PM_{10} and O_3

Causal effect on heath in areas where:



Summary

Causal inference for accountability assessment

- Complex regulatory environment.
- Causal inference with spatial data.
- Principal stratification
  - Multivariate intermediate variable.
  - Multipollutant approach.
- Assumptions about interference between observations.
  - What do we assume?
  - How do we assess?
  - What are the implications of violations?

Thank You

Acknowledgments

- Francesca Dominici
- Yun Wang
- Funding: Health Effects Institute.
- Publication: Zigler CM, Dominici F, and Wang Y. Estimating causal effects of air quality regulations using principal stratification for spatially-correlated multivariate intermediate outcomes. *Biostatistics* 2012; **13**(2): 289–302.