A Source-Oriented Approach to Coal Power Plant Health Effects

Kevin Cummiskey [1], Chanmin Kim [3], Christine Choirat [1], Lucas Henneman [1], Joel Schwartz [2], Corwin Zigler [1]

[1] Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA
[2] Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA
[3] Department of Biostatistics, Boston University School of Public Health, Boston, MA
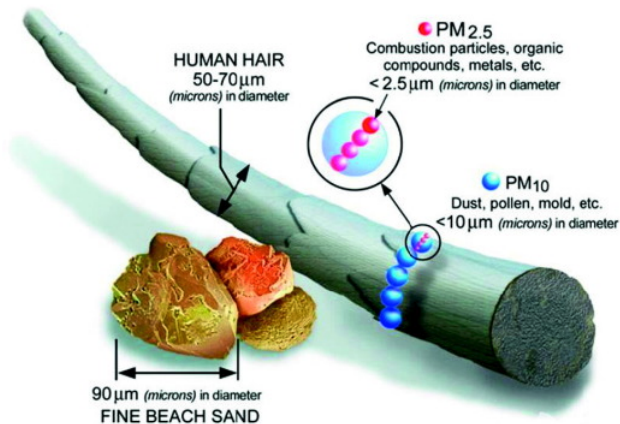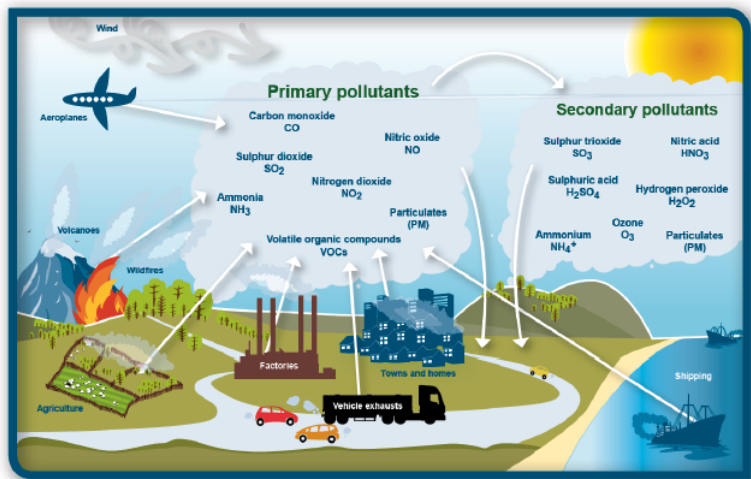
January 12, 2018

Image courtesy of U.S. EPA

Background: long-range pollution transport

Source: Scottish Parliament Information Center, *Air Quality in Scotland*

Objective

Investigate the effect of high-exposure to coal combustion emissions from 783 coal-fired power generating units on ischemic heart disease (IHD) hospitalizations in almost 20 million Medicare beneficiaries residing at 21,000 ZIP codes in the eastern United States in 2005.

Contributions of this study:

- uses emissions data from individual coal power plants
- large-scale analysis
- tightly controls for confounding

Data

Unit of analysis: U.S. ZIP code ($\approx 21,000$ in eastern U.S.)

Outcome:
Ischemic Heart Disease (IHD) hospitalizations (2005) in Medicare
population (≈ 20 million).

Emissions:
Total annual emissions (2005) from 783 coal-fired generating units
(AMPD, NEI)

Covariates:
Census 2000 socioeconomic and demographic variables, smoking rates,
weather characteristics

AMPD = Air Markets Program Data, NEI = National Emissions Inventory

Methods - Overview

1. Exposure classification
   - estimate "coal emissions influence" on each ZIP code using InMAP
   - classify ZIP codes as high-exposed or controls

2. Confounding adjustment with propensity score matching
   - propensity score

$$P(\text{high-exposed}|\text{covariates})$$

   - estimated using logistic regression
   - match high-exposed to controls with similar propensity scores

3. Fit outcome model
   - Poisson regression (by region)
   - quantity of interest: incidence rate ratio (IRR) for IHD hospitalizations comparing high-exposed to control locations

InMAP - An Intervention Model for Air Pollution (Tessum et al 2017)
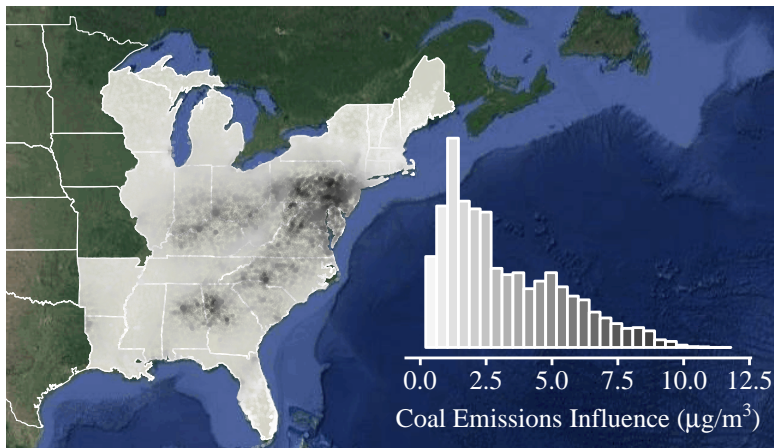
Results - Exposure Classification

Figure 1: InMAP estimates of the change in annual $PM_{2.5}$ concentration for a 100% reduction in coal emissions.
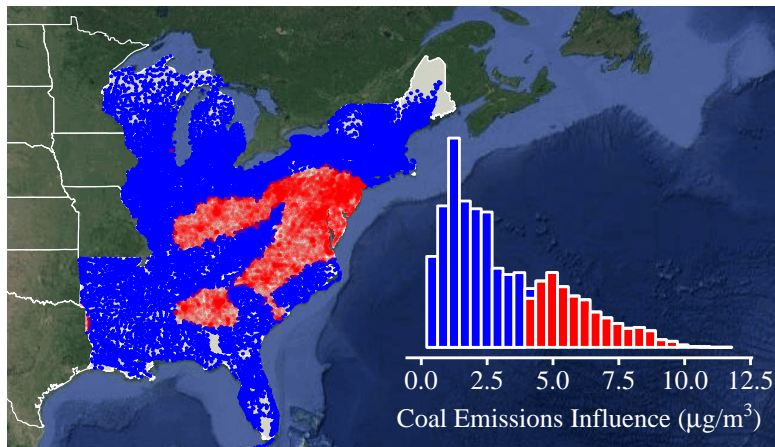
Results - Exposure Classification

Figure 2: High-exposed (red) and control (blue) locations.
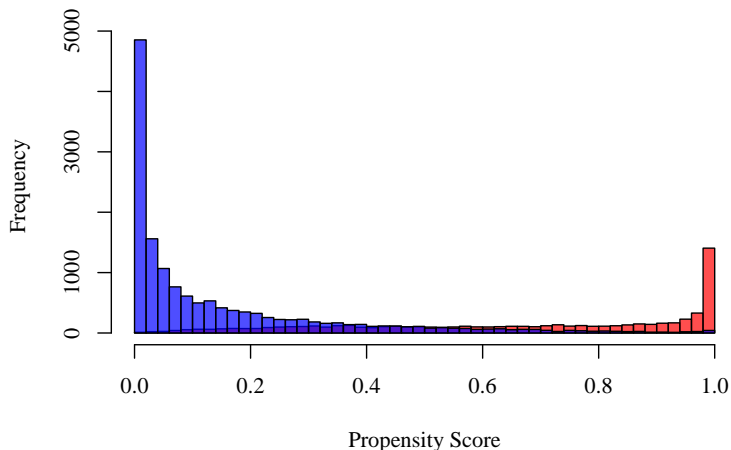
Results - Propensity scores

Figure 3: Histogram of propensity scores for high-exposed (red) and control (blue) locations.
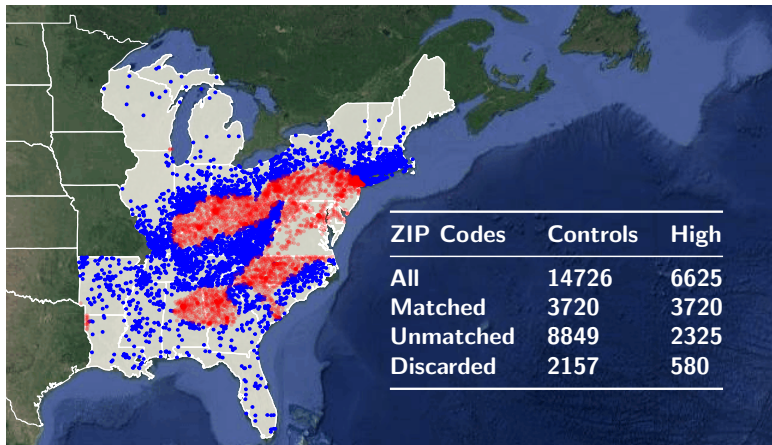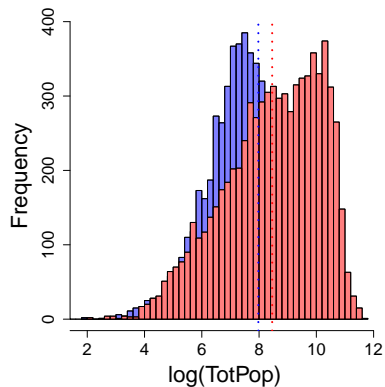
Results - Propensity Score Matched Dataset

ZIP Codes	Controls	High
All	14726	6625
Matched	3720	3720
Unmatched	8849	2325
Discarded	2157	580

Figure 4: High-exposed (red) and control (blue) locations.
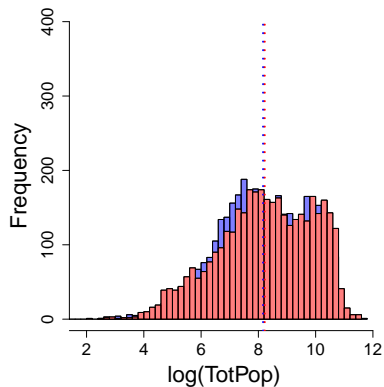
Covariate Balancing

Example: Total Population
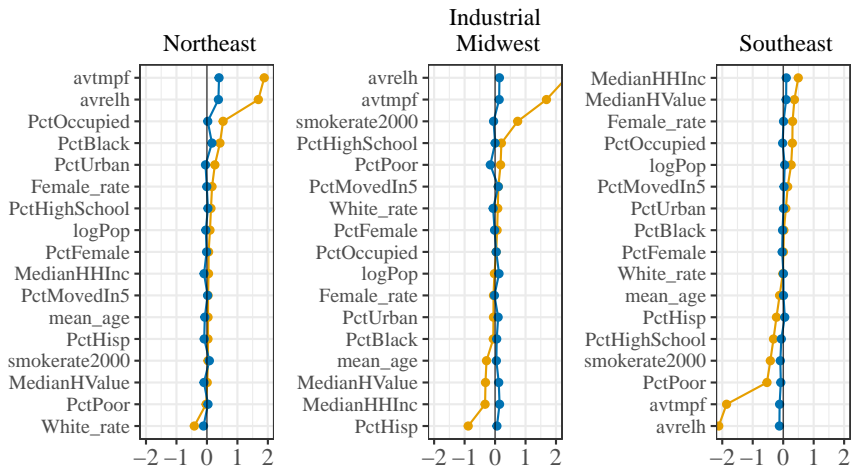
Results - Covariate Balancing

Figure 5: Standardized mean difference between high-exposed and controls - before (orange) and after (blue) matching.

Results - IHD model

Industrial Midwest	Northeast	Southeast
1.02	1.08	1.06
(1.00, 1.04)	(1.06, 1.09)	(1.04, 1.08)

Table 1: Estimated IRRs (95% C.I.) associated with IHD hospitalizations by region.

Conclusions

- significant associations in Northeast and Southeast
- uses emissions data from individual coal power plants
- large-scale analysis
- tightly controls for confounding

References I

1. Tessum, Christopher W., Jason D. Hill, and Julian D. Marshall. "InMAP: A model for air pollution interventions." PloS one 12.4 (2017): e0176131.

2. Austin, Peter C. "An introduction to propensity score methods for reducing the effects of confounding in observational studies." *Multivariate Behavioral Research* 46.3 (2011): 399-424.

3. Di, Qian, et al. "Assessing $PM_{2.5}$ exposures with high spatiotemporal resolution across the continental United States." *Environmental Science & Technology* 50.9 (2016): 4712-4721.

4. Papadogeorgou, Georgia, et al. "Adjusting for Unmeasured Spatial Confounding with Distance Adjusted Propensity Score Matching." (Submitted May 2017).

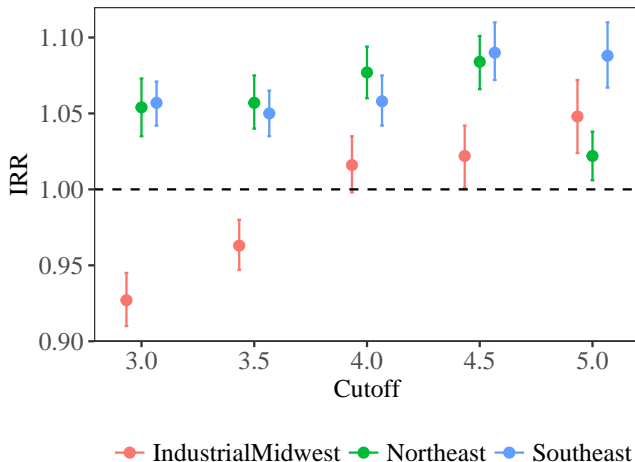# Results - Sensitivity to definition of high-exposed

Figure 6: IRRs for various cutoffs in coal emissions influence between the high-exposed and control locations