The Effect of Air Pollution Control on Life Expectancy in the United States

A Population-Based Analysis of Major Metropolitan Areas

Andrew Correia[1] (email: acorreia@hsph.harvard.edu), C Arden Pope III[2], Douglas Dockery[1], Majid Ezzati[3] & Francesca Dominici[1] (email: fdominic@hsph.harvard.edu)
1 Harvard School of Public Health, Boston, MA
2 Brigham Young University, Provo, UT
3 Imperial College London, London, UK

1. Background

Over the past few decades, there have been substantial and measureable improvements in ambient air quality in the United States (US), while at the same time, though highly variable across US counties, there have been improvements in population survival as well. Recent studies have suggested an association between these reductions in fine-particle air pollution ($PM_{2.5}$) and morbidity/mortality, as well as an association between $PM_{2.5}$ and life expectancy[1–4]. Using the most recent data available, we examine whether the relationship between $PM_{2.5}$ and life expectancy persists for hundreds of US counties in major metropolitan areas across the country using methodology from Pope et al, 2009.

2. Data

- Yearly life expectancy.
  - Calculated at the county level via mortality data obtained from the US National Center for Health Statistics (NCHS) for the years 2000 - 2007.
- Yearly average $PM_{2.5}$ levels .
  - County-level data spanning the years 2000 - 2007.
  - Obtained from US EPA's Air Quality System (AQS).
- Potentially confounding socioeconomic and demographic variables for 2000 - 2007.
  - Variables include population, income, and proportions of persons in the population who were high school graduates, had urban residences, had not lived in their current county of residence 5 years before the census, reported they were smokers, and reported that they were white, black, or Hispanic.
  - Obtained from decennial US Census (2000) and American Community Survey (2007).
- Matching $PM_{2.5}$ and LE data for 545 counties in 257 different Metropolitan Statistical Areas (MSA) for 2000 and 2007.

Select Summary Statistics for the 545 Counties (mean±se)

- Life expectancy (yr)
  - 2000: ..... 76.69 ± 1.72
  - 2007: ..... 77.53 ± 1.98
  - $\Delta LE_{2007-2000}$: ..... 0.84 ± 0.61
- $PM_{2.5}$ ($\mu g/m^3$)
  - 2000: ..... 13.22 ± 3.44
  - 2007: ..... 11.56 ± 2.81
  - $\Delta PM_{2000-2007}$: ..... 1.56 ± 1.48
  - Similar summary statistics available for SES and demographic variables.

4. Results

Results of Selected Regression Models ($\hat{\beta}$±SE)

Variable	Model 1	Model 2	Model 3	Model 4†	Model 5‡	Model 6¶	Model 7††	Model 8‡‡
Intercept	0.82±0.04§	1.08±0.08§	1.00±0.08§	0.97±0.10§	0.91±0.11§	0.84±0.15§	0.79±0.15§	0.67±0.15§
β_1	0.14±0.19	0.35±0.17⋆	0.35±0.16⋆	0.30±0.23	0.72±0.22§	0.95±0.31§	0.74±0.24§	0.96±0.28§
β_2	–	0.013±0.017	0.017±0.018	0.005±0.018	0.02±0.02	-0.01±0.03	0.03±0.02⋆	0.05±0.02§
β_3	–	0.13±0.05§	0.11±0.05§	0.07±0.05	0.06±0.04	0.02±0.04	0.07±0.06	0.34±0.12§
β_4	–	-9.12±1.61§	-7.98±1.56§	-7.27±1.95§	-4.42±2.60	-4.04±3.20	-1.94±3.35	-3.30±3.45
β_5	–	-6.55±2.05§	-6.34±1.97§	-7.86±3.07⋆	-12.56±3.59§	-8.12±2.84§	-11.14±3.00§	-6.21±2.97⋆
β_6	–	-2.16±0.47§	-2.03±0.47§	-2.12±0.59§	-0.95±0.62	5.28±3.58	-3.25±0.63§	-4.57±0.75§
β_7	–	–	-0.02±0.02	-0.02±0.02	-0.01±0.05	-0.05±0.05	-0.07±0.02§	-0.07±0.03⋆
β_8	–	–	-0.05±0.01§	-0.05±0.02§	-0.06±0.03	-0.06±0.05	-0.08±0.02§	-0.06±0.02§
No. of county units	545	545	545	257	307	169	545	545

β_1 = Reduction in $PM_{2.5}$ ($10\mu g/m^3$); β_2 = Change in income (in thousands of $); β_3 = Change in population (in hundreds of thousands);
β_4 = Change in high-school graduates (proportion of population); β_5 = Change in black population (proportion of population);
β_6 = Change in Hispanic population (proportion of population); β_7 = Change in lung cancer mortality rate (no./10,000 population);
β_8 = Change in COPD mortality rate (no./10,000 population).
§ Indicates $p < 0.01$.
⋆ Indicates $p < 0.05$.
† Indicates model included only counties with the largest 2000 population in the MSA.
‡ Indicates model included only counties with a 2000 population density > 200 ppl/sq. mile.
¶ Indicates model included only counties with a 2000 urban rate > 90%.
†† Indicates model was weighted by the square root of year 2000 population density.
‡‡ Indicates model was weighted by the inverse of county land area.

6. Conclusion

Controlling for potential confounders, our results show that a decrease of $10\mu g/m^3$ in $PM_{2.5}$ corresponds to 0.35 years (±0.16) gained in life expectancy. In more densely populated, urban areas, this effect was nearly 3 times higher (0.95 ± 0.31). Therefore we confirm that the benefit of reductions in $PM_{2.5}$ on life expectancy continue to be felt, even as reductions in $PM_{2.5}$ are becoming less drastic than they were over the 20 years prior to our study period.

References

[1] Pope CA, Ezzati M, Dockery D. Fine-particulate air pollution and life expectancy in the United States. N Engl J Med 2009;360:376 - 386.
[2] Dockery DW, Pope CA III, Xu X, Spengler JD, Ware JH, Fay ME, et al. An association between air pollution and mortality in six U.S. cities. N Engl J Med 1993;329:1753-9.
[3] Laden F, Schwartz J, Speizer FE, Dockery DW. Reduction in fine particulate air pollution and mortality: Extended follow-up of the Harvard Six Cities study. Am J Respir Crit Care Med 2006;173:667-72.
[4] Jerrett M, Burnett RT, Ma R, Pope CA III, Krewski D, Newbold KB, et al. Spatial analysis of air pollution and mortality in Los Angeles. Epidemiology 2005;16:727-36.
[5] Diggle PJ, Liang K, Zeger SL. Analysis of longitudinal data. Oxford: Clarendon Press, 1994.
[6] Chiang CL. Indroduction to stochastic processes in biostatistics. Wiley, New York, 1668.

Acknowledgements: This work was funded in part by the U.S. Environmental Protection Agency (R-834894 and RD-83479801)

3. Statistical Approach

- The outcome, difference in life expectancy (in yrs) over the years 2000 to 2007 ($\Delta LE_{2007-2000}$), is modeled via the regression:

$$\Delta LE_i = \alpha_0 + \beta \Delta \mathrm{PM}_{2.5i} + \gamma^T \boldsymbol{Z}_i + \varepsilon_i.$$

- For all models we use an ordinary linear regression approach with the cluster robust sandwich estimator for the variance of $\hat{\beta}$, which allows us to account for the potential lack of statistical independence for counties in the same MSA. Specifically:

$$Var(\hat{\beta}) = \left(X^T X\right)^{-1} \left(X^T \hat{V} X\right) \left(X^T X\right)^{-1}$$

- $\hat{V}$ is the block-diagonal matrix with non-zero blocks $\hat{V}_{0,MSA_j}$ defined by:

$$\hat{V}_{0,MSA_j} = (\boldsymbol{y}_{MSA_j} - \hat{\boldsymbol{\mu}}_{MSA_j})(\boldsymbol{y}_{MSA_j} - \hat{\boldsymbol{\mu}}_{MSA_j})^T$$

- $j \in 1, 2, \ldots, 257$; $\boldsymbol{y}_{MSA_j}$ is the vector of county-level observations for the counties in MSA_j; and $\hat{\boldsymbol{\mu}}_{MSA_j}$ is the vector of OLS estimates corresponding to the counties represented in $\boldsymbol{y}_{MSA_j}$.

5. Discussion

- By using data from the period 2000 to 2007, we provide evidence that more recent declines in $PM_{2.5}$ to relatively low levels continue to prolong life expectancy in the US.
- Also provide evidence that these benefits are largest among the most urban and densely populated counties.
  - Results are consistent with those previously reported.[1]
- Effect of changes in $PM_{2.5}$ on life expectancy were estimated to be 2 to 3 times larger in more densely populated and more urban areas, as well as counties with smaller land area.
  - Perhaps explained by differential exposure misclassification.
  - e.g. People in small counties and/or densely populated counties are more likely to be exposed to the same level of $PM_{2.5}$ than those in larger and/or less densely populated counties.
    - Perhaps more exposure misclassification in larger/less densely populated areas.
- 383 of 545 counties had BRFSS data on smoking prevalence.
  - In these counties, the $PM_{2.5}$ effect was virtually identical in models that included change in smoking variables and those that did not.
- Future work to include more sophisticated statistical modeling approaches to better control for and identify residual confounding.