

January 24, 2019

Peer Review of Manuscript BMJ-2018-048424 by James E. Enstrom, PhD, MPH, FFACE

“Air pollution and cause-specific risks and costs of hospital admissions”

By Yan Wang, ScD; Yaguang Wei, MS; Qian Di, ScD; Christine Choirat, PhD; Yun Wang, PhD; Petros Koutrakis, PhD; Antonella Zanobetti, PhD; Francesca Dominici, PhD (Dominici); and Joel D. Schwartz, PhD (Schwartz) from Harvard T.H. Chan School of Public Health, Boston, MA, USA with Dominici and Schwartz as Senior Authors

Key Sentences from Abstract and Text:

“Introduction: Short-term exposure to particulate matter with an aerodynamic diameter of <2.5 µm (PM_{2.5}) is associated with increased risks of deaths and hospital admissions.¹⁻⁸ The World Health Organization (WHO) set the air quality guideline for 24-hour average exposure to PM_{2.5} at 25 µg·m⁻³ in 2005.⁹ The WHO air quality guidelines are currently being reviewed with the date of expected publication in 2020, and scientific evidence supporting the update of the guidelines is subject to an unprecedented level of scrutiny.¹⁰”

“Results: We discovered statistically significant positive associations between short-term PM_{2.5} and hospitalizations for several prevalent but rarely studied diseases, including septicemia, fluid and electrolyte disorders, and acute and unspecified renal failure. We also found statistically significant positive associations for hospitalizations due to cardiovascular and respiratory diseases, Parkinson’s disease, diabetes mellitus with complications, phlebitis, thrombophlebitis, and thromboembolism, confirming previous results.”

“Conclusions: This study discovered new diseases and confirmed known diseases associated with short-term PM_{2.5} exposure, demonstrating substantial health benefits linked to a small reduction in short-term PM_{2.5}.”

Peer Review – Justification for Unequivocal Rejection of this Manuscript

- Originality - This work DOES NOT add enough to what is already in the published literature (references 1-8 by these same Senior Authors). This manuscript is latest addition to the massive effort by Schwartz and Dominici to promote the implausible and unproven hypothesis that many human health conditions, including premature death, are *caused* by short-term exposure to trace amounts of particulate matter, particularly fine particulate matter (PM_{2.5}). This effort dates back to at least the 1992 Am J Epi article by Schwartz “Particulate Air Pollution and Daily Mortality in Steubenville, Ohio” (<http://www.scientificintegrityinstitute.org/AJESchwartz1992.pdf>).”

- Importance of work to general readers – This work is NOT valuable to clinicians, patients, teachers, or policymakers because it is NOT scientifically transparent or valid for many reasons, as explained below. In addition, a general medical journal like BMJ is NOT the right place for a work that uses complex and non-transparent statistical analysis. This work is not written for the benefit of general readers. It is intended to influence the WHO and US EPA PM_{2.5} assessment and regulations.

- Scientific reliability – This research is not scientifically reliable because it is NOT transparent and NOT verifiable. Indeed, both Schwartz and Dominici co-signed a 60-page August 7, 2018 Harvard University anti-transparency letter by environmental lawyer Wendy B. Jacobs. This letter strongly opposes the currently proposed EPA rule “Strengthening Transparency in Regulatory Science” (<http://www.scientificintegrityinstitute.org/HELEPATrans080718.pdf>)

- Research Question – The research question is NOT appropriate and is NOT appropriately answered. The four lead Chinese co-authors, Wang, Wei, Di, and Wang, know that air pollution is a serious problem in China but not in the US. I filed a formal January 31, 2018 research misconduct complaint against the Duke/Chinese statistician Junfang Zhang, PhD, who wrote a deliberately incorrect editorial supporting the December 26, 2017 JAMA Di-Dominici-Schwartz study “Association of Short-term Exposure to Air Pollution With Mortality in Older Adults” (<http://www.scientificintegrityinstitute.org/Zhang013118.pdf>).

Zhang did not acknowledge that the Di-Dominici-Schwartz evidence on PM2.5 premature deaths in the Medicare population (as published in the June 29, 2017 NEJM and December 26, 2017 JAMA articles by Di et al) is severely flawed, as explained by Steve Milloy, JD, in his requests for retraction dated July 5, 2017 to the NEJM Editor (<https://junkscience.com/2017/07/retraction-request-made-fornejm-air-pollution-kills-study/>) and dated January 4, 2018 to the JAMA Editor (<https://junkscience.com/2018/01/junkscience-com-requests-jama-retract-new-harvard-pm2-5study-on-basis-of-scientific-misconduct/>). In addition, research misconduct complaints against Di et al have been filed with the US Office of Research Integrity by Mr. Milloy on September 5, 2017 (<https://junkscience.com/2017/09/junkscience-requests-federal-research-misconductinvestigation-for-air-pollution-study/>) and by John D. Dunn, MD, JD, on January 30, 2018.

- Overall design of study – The overall design of the study is NOT appropriate.

- Participants studied – The participants had 95 million Medicare inpatient hospital claims in the US during 2000–2012. I believe the participants include several of my relatives and friends without their knowledge or consent.

- Methods – The methods are described but it is impossible to use this description to independently replicate these findings. The manuscript contains this statement: “**Ethical approval:** This study was approved by the institutional review board at the Harvard T.H. Chan School of Public Health and was exempt from informed consent requirements as a study of previously collected administrative data.”

I content that the study does NOT have “Ethical approval”. I believe that the authors have obtained inappropriate access to 95 million Medicare hospitalization records during 2000-2012 for millions of Americans. Based on the authors’ description, the records of several of my relatives and friends are presumably included in this study without their knowledge or consent. The persons I know NEVER gave Medicare administrators approval to release their hospitalization records for epidemiological research of the scientifically questionable type done by Dominici and Schwartz.

I plan to submit a complaint to Medicare officials to stop the release of Medicare hospitalization records for epidemiologic research unless specific approval is given by the subjects. My formal complaint will go to US DHHS Centralized Case Management Operations, citing HIPAA regulations (<https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html>). Steve Milloy, JD, has already filed complaints requesting withdrawal of the 2017 NEJM and JAMA articles by Dominici and Schwartz, as well as complaints to the US DHH Office of Research Integrity, as cited above.

- Results – The results are NOT credible given all the objections described above. In any case, the relationships reported are too weak to qualify as a valid epidemiologic relationships.

- Interpretation and conclusions – The interpretation and conclusions are NOT based on an objective assessment of the data and its severe limitations. Instead, they are biased toward the authors’ predetermined conclusion that they are “demonstrating substantial health benefits linked to a small reduction in short-term PM_{2.5}.”

- References – The authors have selected only those references that support their research findings and there are glaring omissions. Following their prior pattern, Dominici and Schwartz have not addressed or cited the severe criticism of their 2017 NEJM and JAMA publications based on Medicare data. For example, the five omitted references shown below are all relevant to the validity of the findings presented in their current manuscript:
 - 1) October 12, 2017 NEJM letter by this reviewer, James E, Enstrom, PhD, MPH;
 - 2) May 22/29, 2018 JAMA letters by Air Pollution Expert Frederick Lipfert, PhD; EPA CASAC Chair Louis Anthony Cox Jr, PhD; and EPA Science Advisory Board Member S. Stanley Young, PhD;
 - 3) [June 1, 2011 JASA article](#) by Sonja Greven, Francesca Dominici, and Scott Zeger, “**An Approach to the Estimation of Chronic Air Pollution Effects Using Spatio-Temporal Information**”, with sentence “In either event, observational studies like these are subject to confounding by unmeasured variables.”;
 - 4) March 1, 2006 Am J Epi article by Roger D. Peng, Francesca Dominici, and Scott L. Zeger, “**Reproducible Epidemiologic Research**”, with sentences “Scientific evidence is strengthened when important findings are replicated by multiple independent investigators using independent data, analytical methods, laboratories, and instruments. Replication, as described here, has long been the standard in the biologic and physical sciences and is of critical importance in epidemiologic studies, particularly when they can impact broad policy or regulatory decisions.”
 - 5) April 11, 2015 Lancet Editorial by Richard Horton “**Offline: What is medicine’s 5 sigma?**”, with sentences “The apparent endemicity of bad research behaviour is alarming. In their quest for telling a compelling story, scientists too often sculpt data to fit their preferred theory of the world.”

- Abstract/summary/key messages/What this paper adds – These four items are severely flawed based on the reasons and evidence presented above.

No potential conflict of interest relevant to this letter was reported.

- Berger RE, Ramaswami R, Solomon CG, Drazen JM. Air pollution still kills. *N Engl J Med* 2017;376:2591-2.
- Aung N, Sanghvi MM, Zemrak F, et al. Impact of fine particulate matter air pollutant on cardiac atrial and ventricular structure and function derived from cardiovascular magnetic resonance (CMR) imaging — evidence from the UK Biobank. Presented at EuroCMR 2017, Prague, May 26, 2017. abstract.
- Kravchenko J, Akushevich I, Abernethy AP, Holman S, Ross WG Jr, Lyerly HK. Long-term dynamics of death rates of emphysema, asthma, and pneumonia and improving air quality. *Int J Chron Obstruct Pulmon Dis* 2014;9:613-27.
- Ross WG Jr. The North Carolina Clean Smokestacks Act. *N C Med J* 2011;72:128-31.

DOI: 10.1056/NEJMc1709849

TO THE EDITOR: The article by Di et al. contains weak noncausal evidence that PM_{2.5} is related to total mortality in the Medicare population. It does not cite the previous evidence reported by Zeger et al.¹ of a large, unexplained geographic variation in the risk of death associated with PM_{2.5} and of no risk of death associated with PM_{2.5} if the risk is based on a local regression coefficient that indicates the association between location-specific trends in pollution and mortality, as described in the detailed statistical analysis reported by Greven et al.²

The article by Di and colleagues also does not cite recent data showing no risk of death associated with PM_{2.5} in the National Institutes of Health–American Association of Retired Persons Diet and Health Study cohort³ and the Cancer Prevention Study cohort.⁴ We think that before the findings of the federally funded study by Di et al. are accepted as valid, the underlying Medicare data should be analyzed independently in accordance with the HONEST (Honest and Open New EPA Science Treatment) Act.⁵

James E. Enstrom, Ph.D., M.P.H.

University of California, Los Angeles
Los Angeles, CA
jenstrom@ucla.edu

No potential conflict of interest relevant to this letter was reported.

- Zeger SL, Dominici F, McDermott A, Samet JM. Mortality in the Medicare population and chronic exposure to fine particulate air pollution in urban centers (2000-2005). *Environ Health Perspect* 2008;116:1614-9.
- Greven S, Dominici F, Zeger S. An approach to the estimation of chronic air pollution effects using spatio-temporal information. *J Am Stat Assoc* 2011;106:396-406.
- Thurston GD, Ahn J, Cromar KR, et al. Ambient particulate matter air pollution exposure and mortality in the NIH-AARP Diet and Health cohort. *Environ Health Perspect* 2016;124:484-90.
- Enstrom JE. Fine particulate matter and total mortality in cancer prevention study cohort reanalysis. *Dose Response* 2017; 15:1559325817693345.

- H.R. 1430, Honest and Open New EPA Science Treatment (HONEST) Act of 2017 (<http://www.govtrack.us/congress/bills/115/hr1430/text>).

DOI: 10.1056/NEJMc1709849

THE AUTHORS REPLY: In response to Majeed and Majeed: our definition of warm-season ozone is consistent with that in the study by Jerrett et al., the results of which were also published in the *Journal*.¹ Although ozone levels peak over the summer, in recent decades, summer ozone levels have decreased, but spring and autumn ozone levels have increased. Using a statistical approach for causal inference, we have previously shown that exposure to high levels of ozone in the spring, summer, and fall is associated with an increased risk of death.²

In response to Raymond: given that there is no threshold for the relationship between PM_{2.5} and mortality, any reduction in air pollution is beneficial. Establishing a restriction on diesel idling would reduce air pollution without cost.

Enstrom points to some studies with null findings that we did not cite. Our conclusions would not have changed on the basis of which of the hundreds of studies of air pollution we might have cited. This is because our study is not a meta-analysis. It is an analysis of new nationwide data and an assessment of exposure with high spatial resolution (i.e., daily PM_{2.5} and ozone concentrations for nationwide grids that were 1 km by 1 km), and we reported strong, not weak associations. Sensitivity analyses showed that smoking and socioeconomic status are unlikely to confound the association, and we controlled for spatial variation (see the Supplementary Appendix, available with the full text of our article at NEJM.org). Moreover, meta-analyses of all published cohort studies show strong, robust associations of PM_{2.5} with mortality,³ and two recent studies have shown similar associations with the use of causal modeling techniques.^{4,5} The Medicare beneficiary denominator file from the Centers for Medicare and Medicaid Services is a publicly available data source, and therefore this study can be independently replicated.

Qian Di, M.S.

Francesca Dominici, Ph.D.

Joel D. Schwartz, Ph.D.

Harvard T.H. Chan School of Public Health
Boston, MA
fdominic@hsph.harvard.edu

Since publication of their article, the authors report no further potential conflict of interest.

ozone were associated with increased mortality in older adults using statistical analyses of a large database. However, they used incomplete exposure data and an inappropriate outcome measure, and they neglected regional variability.

The 1952 London fog established the lethality of air pollution.² The measure of risk used to investigate that event was the sum over lag days, but Di and colleagues relied on significance testing of individual lags of 0 or 1 day. Summing over lags from 0 to 4 days would increase the estimated risk.

Exposures were limited to outdoor air, although most individuals spend 85% of their time indoors, where PM_{2.5} can be augmented by indoor sources³ and ozone is adsorbed onto interior surfaces.

PM_{2.5} is not a pollutant per se but a regulatory construct largely based on facility of monitoring. It comprises a mixture of various particle sizes and composition, only some of which may be toxic, elemental carbon being more important than sulfate.⁴ When PM_{2.5} composition and toxicity vary, a counterintuitive dose-response function, as shown in Figure 5 in the article,¹ may result. Subgroup analyses of regions having typically different PM_{2.5} compositions would have been useful.

Persons most at risk and physiological mechanisms remain largely unknown. Di and colleagues found risks sharply increasing with age but posited that a random individual could succumb to a small perturbation in outdoor air quality. Another mortality model considered prior frailty and acute excursions of pollution and temperature combined.⁵ This model estimated that deaths among older persons in Chicago were limited to a small subset of frail individuals for which PM₁₀, ozone, nitrogen dioxide, sulfur dioxide, or carbon monoxide contributed losses of fewer than 2 days. It found an increased mortality risk for this frail population over 15 days of 0.83% with each increase of 10 µg/m³ of PM₁₀, similar to the results of the current study. In this alternative model, thresholds could occur with decreases in either individual frailty or pollution, but the former is unlikely because a day without frailty would be rare.

Frederick W. Lipfert, PhD

Author Affiliation: Greenport, New York.

Corresponding Author: Frederick W. Lipfert, PhD, 1500 Brecknock Rd, Apt 155, Greenport, NY 11944 (fred.lipfert@gmail.com).

Conflict of Interest Disclosures: The author has completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest and none were reported.

1. Di Q, Dai L, Wang Y, et al. Association of short-term exposure to air pollution with mortality in older adults. *JAMA*. 2017;318(24):2446-2456.
2. Lipfert FW. *Air Pollution and Community Health: A Critical Review and Data Sourcebook*. New York, NY: Wiley; 1994.
3. Lipfert FW. An assessment of air pollution exposure information for health studies. *Atmosphere*. 2015;6(11):1736-1752. doi:10.3390/atmos6111736
4. Lipfert FW, Wyzga RE, Baty JD, Miller JP. Air pollution and survival within the Washington University-EPRI veterans cohort: risks based on modeled estimates of ambient levels of hazardous and criteria air pollutants. *J Air Waste Manag Assoc*. 2009;59(4):473-489.
5. Murray CJ, Lipfert FW. Inferring frail life expectancies in Chicago from daily fluctuations in elderly mortality. *Inhal Toxicol*. 2013;25(8):461-479.

To the Editor In a case-crossover study, the authors noted that “In the US Medicare population from 2000 to 2012, short-term exposures to PM_{2.5} and warm-season ozone were significantly associated with increased risk of mortality. This risk occurred at levels below current national air quality standards, suggesting that these standards may need to be reevaluated.”¹ Such studies of association fail to address the key causal question: How would public health effects change if exposure concentrations were reduced? Instead, they addressed an easier, noncausal question: What are the estimated ratios (or slope factors or regression coefficients) of health effects to past pollution levels in selected models and data sets? Answers to the second question are inadequate substitutes for answers to the first question for several reasons.

First, published associations are often assumption and model dependent. Exposure may have a positive association with mortality in some regression models and a negative association in others; which is reported depends on the model selected.² Second, omitted confounders can create spurious exposure-response associations. Mr Di and colleagues¹ omitted lagged temperatures for days 2 to 7. Yet, in publicly available data,³ lagged temperatures were associated with both PM_{2.5} and mortality. PM_{2.5} predicts mortality only if lagged temperatures are omitted.

Third, model specification errors create spurious associations. In data from Los Angeles, PM_{2.5} predicted mortality using Poisson regression by reducing specification error; in nonparametric analyses, it was not a predictor.³ Fourth, ignored measurement errors can create spurious low-dose associations.⁴ The model used by Di and colleagues¹ omitted exposure measurement error. This can make even threshold exposure-response relations look linear at low doses,⁴ consistent with the finding in the study that responses “were almost linear, with no indication of a mortality risk threshold at very low concentrations.”¹

Evaluating adequacy of air quality standards requires addressing the first causal question. Causal analytics methods and software can help. Doing so will give regulators the scientific information they most need.

Louis Anthony Cox Jr, PhD

Author Affiliation: Cox Associates, Denver, Colorado.

Corresponding Author: Louis Anthony Cox Jr, PhD, Cox Associates, 503 N Franklin St, Denver, CO 80218 (tcxdenver@aol.com).

Conflict of Interest Disclosures: The author has completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest and reported receiving personal fees from the US Environmental Protection Agency (EPA), Society for Epidemiological Research, Society for Benefit Cost Analysis, and American Industrial Hygiene Council; fees to his company from the American Chemistry Council (ACC), American Petroleum Institute (API), and their members; salary for teaching from the University of Colorado; serving as the chair of the US EPA Clean Air Science Advisory Committee and a board member of EPA's Science Advisory Board; and being editor in chief of *Risk Analysis: An International Journal*.

Disclaimer: The contents of this letter are solely the views of the author and do not necessarily represent any view or opinion on behalf of EPA, its Clean Air Science Advisory Committee, API, ACC, *Risk Analysis*, or the Society for Risk Analysis.

1. Di Q, Dai L, Wang Y, et al. Association of short-term exposure to air pollution with mortality in older adults. *JAMA*. 2017;318(24):2446-2456.
2. Dominici F, Greenstone M, Sunstein CR. Science and regulation: particulate matter matters. *Science*. 2014;344(6181):257-259.
3. Cox LAT Jr. Do causal concentration-response functions exist? a critical review of associational and causal relations between fine particulate matter and mortality. *Crit Rev Toxicol*. 2017;47(7):603-631.
4. Rhomberg LR, Chandalia JK, Long CM, Goodman JE. Measurement error in environmental epidemiology and the shape of exposure-response curves. *Crit Rev Toxicol*. 2011;41(8):651-671.

To the Editor There are many aspects of the study by Mr Di and colleagues¹ that call into question their finding that air quality was associated with increased deaths: size of the effects, modeling bias, and flexibility of analysis. These methodological aspects are especially concerning given that recent large studies found no association between air quality and mortality.^{2,3}

The size of the effects in the study by Di and colleagues were small, 1% or less. Any small bias or model misspecification could produce such a small effect,⁴ as could aspects of the analysis, such as multiple testing and multiple modeling.⁵ For example, the baseline factors in Table 1 in the article¹ could produce 80 subgroup analyses. Consider the treatment of temperature and time lags. Each could exert an effect on the day in question or either of the previous 2 days, for $3 \times 3 = 9$ combinations. Modeling of exposure data could produce 720 possible analyses. Although the authors cited no negative studies, one offers a possible explanation for the positive results of the current analysis: confounding variables that differ across locations.³ Di and colleagues did not do a within- and across-location analysis.

When results are dependent on statistical analyses, it is the obligation of the authors to provide strong evidence, address conflicting studies, and make their data set and analysis code available. Considering the number of analysis options available, the results of this study could have been the result of the analysis choices made.

S. Stanley Young, PhD

Author Affiliation: CGStat, Raleigh, North Carolina.

Corresponding Author: S. Stanley Young, PhD, CGStat, 3401 Caldwell Dr, Raleigh, NC 27607 (genetree@bellsouth.net).

Conflict of Interest Disclosures: The author has completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest and none were reported.

1. Di Q, Dai L, Wang Y, et al. Association of short-term exposure to air pollution with mortality in older adults. *JAMA*. 2017;318(24):2446-2456.
2. Young SS, Smith RL, Lopiano KK. Air quality and acute deaths in California, 2000-2012. *Regul Toxicol Pharmacol*. 2017;88:173-184.
3. Greven S, Dominici F, Zeger S. An approach to the estimation of chronic air pollution effects using spatio-temporal information. *J Am Stat Assoc*. 2011;106(494):396-406.
4. Young SS. Statistical Analyses and Interpretation of Complex Studies. Medscape. 2008. <https://www.medscape.org/viewarticle/571523>. Accessed January 4, 2018.
5. Simonsohn U, Nelson LD, Simmons JP. P-curve: a key to the file-drawer. *J Exp Psychol Gen*. 2014;143(2):534-547.

In Reply Dr Lipfert criticizes our study for not including indoor-generated particles in our exposure. Indoor particles are a potential confounder, not an omitted part of outdoor particles. Differences in indoor exposure across persons cannot confound the results of our case-crossover study; the exposure contrast was within person, between case and control days. To confound, daily variation in particles from cooking, smoking, etc, must be correlated with daily variation in outdoor PM_{2.5} of the same person. The Medicare Beneficiary Survey showed that 86% of beneficiaries were nonsmokers. It seems implausible that smokers consistently smoked more or that people fried more food on higher pollution days.

We believe total early deaths is an appropriate and policy-relevant outcome. We agree that frailty is an issue and examined modification by sex, race/ethnicity, poverty, and age; in other studies, medical conditions were examined as predisposing.¹ We disagree that only frail people die from air pollution; multicity studies have shown that this is not the case.² It would be useful to examine regional variations to learn about particle composition, but that does not imply that our national estimate is biased as a national average estimate. Such examination requires defining region by pollution mix and not census categories, a substantial effort we will consider in the future. In addition, many studies have found that for all-cause mortality, lags 0 and 1 are the most relevant averaging period.

Dr Cox questions our finding of effects below the US Environmental Protection Agency's current standard because with a wide range of exposure and considerable exposure error, nonlinear relationships may look linear. However, this is not relevant to our analysis restricted to observations when PM_{2.5} was below 25 µg/m³, well below the current standard of 35 µg/m³. Because the root mean squared error of our exposure model was 2.7 µg/m³, a significant association in the restricted analysis cannot be due to exposure error. Cox also cites a report that the association between PM_{2.5} and daily deaths disappeared after control for more lags of temperature. Other studies have differed. In a 14-city case-crossover study,³ a larger effect size for particles controlling for temperature lags 0 through 4 was found than with lags 0 and 1. Cox argues that causal modeling methods would be useful in air pollution epidemiology. We agree, and several causal modeling analyses of this question have been published, with more planned.^{4,5} They all support the association we reported in this study.

Dr Young argues that our effect size estimate was small and that there are many possible choices that could affect our result, including subgroups analyzed, how covariates are modeled, and that the study could have been confounded by variables in Table 1 of our article that differed across locations. Because our primary analysis was all beneficiaries, the implication that choices of subgroups could affect the outcome is not correct. Variables that differ across locations cannot be confounders in a case-crossover analysis because the analysis was within person. Exposure contrasts were within participants, on a case day and nearby

An Approach to the Estimation of Chronic Air Pollution Effects Using Spatio-Temporal Information

Sonja GREVEN, Francesca DOMINICI, and Scott ZEGER

There is substantial observational evidence that long-term exposure to particulate air pollution is associated with premature death in urban populations. Estimates of the magnitude of these effects derive largely from cross-sectional comparisons of adjusted mortality rates among cities with varying pollution levels. Such estimates are potentially confounded by other differences among the populations correlated with air pollution, for example, socioeconomic factors. An alternative approach is to study covariation of particulate matter and mortality across time within a city, as has been done in investigations of short-term exposures. In either event, observational studies like these are subject to confounding by unmeasured variables. Therefore the ability to detect such confounding and to derive estimates less affected by confounding are a high priority.

In this article, we describe and apply a method of decomposing the exposure variable into components with variation at distinct temporal, spatial, and time by space scales, here focusing on the components involving time. Starting from a proportional hazard model, we derive a Poisson regression model and estimate two regression coefficients: the “global” coefficient that measures the association between national trends in pollution and mortality; and the “local” coefficient, derived from space by time variation, that measures the association between location-specific trends in pollution and mortality adjusted by the national trends. Absent unmeasured confounders and given valid model assumptions, the scale-specific coefficients should be similar; substantial differences in these coefficients constitute a basis for questioning the model.

We derive a backfitting algorithm to fit our model to very large spatio-temporal datasets. We apply our methods to the Medicare Cohort Air Pollution Study (MCAPS), which includes individual-level information on time of death and age on a population of 18.2 million for the period 2000–2006.

Results based on the global coefficient indicate a large increase in the national life expectancy for reductions in the yearly national average of $PM_{2.5}$. However, this coefficient based on national trends in $PM_{2.5}$ and mortality is likely to be confounded by other variables trending on the national level. Confounding of the local coefficient by unmeasured factors is less likely, although it cannot be ruled out. Based on the local coefficient alone, we are not able to demonstrate any change in life expectancy for a reduction in $PM_{2.5}$. We use additional survey data available for a subset of the data to investigate sensitivity of results to the inclusion of additional covariates, but both coefficients remain largely unchanged.

KEY WORDS: Backfitting algorithm; Environmental epidemiology; Particulate matter; Spatio-temporal data; Specification test.

1. INTRODUCTION

The Clean Air Act (Environmental Protection Agency, last amended in 1990) requires the U.S. Environmental Protection Agency (EPA) to set National Ambient Air Quality Standards for seven pollutants considered harmful. Air quality standards for several air pollutants have since also been adopted by the European Union. Implementation of these standards led to decreases in air pollution concentrations in the United States (Bachmann 2008). From a public policy and public health perspective, it is of importance to assess whether these decreases have also led to an improvement in morbidity and mortality for the general population (Health Effects Institute 2003). Standards are reviewed periodically, with evidence from epidemiologic studies playing a large role in the public policy process (Kaiser 1997; Greenbaum et al. 2001; Samet et al. 2003). While there is substantial observational evidence that long-term

exposure to particulate air pollution is associated with premature death in urban populations, confounding by unmeasured variables remains a large concern in observational studies. The ability to detect such confounding and to derive estimates less affected by confounding thus are of great importance.

Evidence on the magnitude of the chronic effects of long-term exposure to air pollution on mortality stems mostly from cohort studies (see, e.g., Dockery et al. 1993; Pope et al. 2002; Laden et al. 2006; Eftim et al. 2008). These studies compare across locations long-term average air pollution concentrations and time-to-death in cohorts. Cohort studies allow the estimation of life expectancy lost due to air pollution (Künzli et al. 2001; Rabl 2003). They have been criticized (Moolgavkar 1994; Vedal 1997; Gamble 1998), due to the difficulty of fully accounting for all potential confounders, including individual risk factors and location-specific characteristics such as socioeconomic factors.

An alternative approach is to study covariation of particulate matter and mortality across time within a predefined geographical location (e.g., county or city), as has been done in investigations of health effects associated with short-term exposures. Time series studies (see, e.g., Schwartz and Dockery 1992; Spix et al. 1993; Kelsall et al. 1997) estimate acute effects of short-term exposure to air pollutants, comparing day-to-day variations in mortality with those in air pollution concentrations. Multisite time series studies (Katsouyanni et al. 1997;

Sonja Greven is Emmy Noether Junior Research Group Leader, Department of Statistics, Ludwig-Maximilians-Universität München, 80539 Munich, Germany (E-mail: sonja.greven@stat.uni-muenchen.de). Francesca Dominici is Professor, Department of Biostatistics, Harvard University, Boston, MA 02115. Scott Zeger is Professor, Department of Biostatistics, Johns Hopkins University, Baltimore, MD 21205. Funding was provided by the U.S. Environmental Protection Agency (EPA) (grants RD-83241701 and RD-83362201). Although the research described in this article has been funded wholly or in part by the U.S. EPA, it has not been subjected to the agency's required peer and policy review and does not necessarily reflect the views of the agency, and no official endorsement should be inferred. The first author was also funded by Emmy Noether grant GR 3793/1-1 from the German Research Foundation. The authors thank the referees, associate editor, and editor for helpful comments, and Aidan McDermott for help with the datasets.



Commentary

Reproducible Epidemiologic Research

Roger D. Peng, Francesca Dominici, and Scott L. Zeger

From the Biostatistics Department, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD.

Received for publication November 4, 2005; accepted for publication January 10, 2006.

The replication of important findings by multiple independent investigators is fundamental to the accumulation of scientific evidence. Researchers in the biologic and physical sciences expect results to be replicated by independent data, analytical methods, laboratories, and instruments. Epidemiologic studies are commonly used to quantify small health effects of important, but subtle, risk factors, and replication is of critical importance where results can inform substantial policy decisions. However, because of the time, expense, and opportunism of many current epidemiologic studies, it is often impossible to fully replicate their findings. An attainable minimum standard is “reproducibility,” which calls for data sets and software to be made available for verifying published findings and conducting alternative analyses. The authors outline a standard for reproducibility and evaluate the reproducibility of current epidemiologic research. They also propose methods for reproducible research and implement them by use of a case study in air pollution and health.

air pollution; information dissemination; models, statistical

Abbreviation: NMMAPS, National Morbidity, Mortality, and Air Pollution Study.

Determinants of human disease are commonly investigated by epidemiologic studies focused on a particular subpopulation, time frame, and geographic location. Findings from such studies can play an important role in policy decisions affecting public health (1). Yet epidemiologic research has been criticized as being increasingly unreliable. One review of the field a decade ago raised questions about the reliability of observational epidemiologic studies when quantifying the health effects of important, but subtle, risk factors such as second-hand smoke, air pollution, and diet (2).

Scientific evidence is strengthened when important findings are *replicated* by multiple independent investigators using independent data, analytical methods, laboratories, and instruments. Replication, as described here, has long been the standard in the biologic and physical sciences and is of critical importance in epidemiologic studies, particularly when they can impact broad policy or regulatory decisions. Because of the time and expense involved with epidemiologic studies, many are often not fully replicable,

and policy decisions must be made with the evidence at hand.

An attainable minimum standard is *reproducibility*, where independent investigators subject the original data to their own analyses and interpretations. Reproducibility calls for data sets and software to be made available for 1) verifying published findings, 2) conducting alternative analyses of the same data, 3) eliminating uninformed criticisms that do not stand up to existing data, and 4) expediting the interchange of ideas among investigators. Ultimately, all scientific evidence should be held to the standard of full replication and the confirmation of important findings by independent investigators. However, the desire to quantify small health effects and the significant weight placed on epidemiologic findings in the policy-making process create a need for epidemiologic studies to meet a minimum standard. We propose reproducibility to be this minimum standard.

There are a number of new developments that are intensifying the need for reproducible epidemiologic research.

Offline: What is medicine's 5 sigma?



Stuart Clarke/Revea Features

"A lot of what is published is incorrect." I'm not allowed to say who made this remark because we were asked to observe Chatham House rules. We were also asked not to take photographs of slides. Those who worked for government agencies pleaded that their comments especially remain unquoted, since the forthcoming UK election meant they were living in "purdah"—a chilling state where severe restrictions on freedom of speech are placed on anyone on the government's payroll. Why the paranoid concern for secrecy and non-attribution? Because this symposium—on the reproducibility and reliability of biomedical research, held at the Wellcome Trust in London last week—touched on one of the most sensitive issues in science today: the idea that **something has gone fundamentally wrong with one of our greatest human creations.**

*



View Pictures/ULG via Getty Images

The case against science is straightforward: much of the scientific literature, perhaps half, may simply be untrue. Afflicted by studies with small sample sizes, tiny effects, **invalid exploratory analyses**, and flagrant conflicts of interest, together with an obsession for pursuing fashionable trends of dubious importance, science has taken a turn towards darkness. As one participant put it, **"poor methods get results"**. The Academy of Medical Sciences, Medical Research Council, and Biotechnology and Biological Sciences Research Council have now put their reputational weight behind an investigation into these questionable research practices. The apparent endemicity of bad research behaviour is alarming. In their quest for **telling a compelling story**, scientists too often sculpt data to fit their preferred theory of the world. Or they retrofit hypotheses to fit their data. Journal editors deserve their fair share of criticism too. We aid and abet the worst behaviours. Our acquiescence to the impact factor fuels an unhealthy competition to win a place in a select few journals. Our **love of "significance" pollutes the literature with many a statistical fairy-tale.** We reject important confirmations. Journals are not the only miscreants. Universities are in a perpetual struggle for money and talent, endpoints that foster reductive metrics, such as high-impact publication. National assessment procedures, such as the Research Excellence Framework, incentivise bad practices. And individual scientists, including their

most senior leaders, do little to alter a research culture that occasionally veers close to misconduct.

*

Can bad scientific practices be fixed? Part of the problem is that **no-one is incentivised to be right.** Instead, scientists are incentivised to be productive and innovative. Would a Hippocratic Oath for science help? Certainly don't add more layers of research red-tape. Instead of changing incentives, perhaps one could remove incentives altogether. Or insist on replicability statements in grant applications and research papers. Or emphasise collaboration, not competition. Or **insist on preregistration of protocols.** Or reward better pre and post publication peer review. Or improve research training and mentorship. Or implement the recommendations from our Series on increasing research value, published last year. One of the most convincing proposals came from outside the biomedical community. Tony Weidberg is a Professor of Particle Physics at Oxford. Following several high-profile errors, the particle physics community now invests great effort into intensive checking and re-checking of data prior to publication. **Filtering results through independent working groups, physicists are encouraged to criticise. Good criticism is rewarded. The goal is a reliable result, and the incentives for scientists are aligned around this goal. Weidberg worried we set the bar for results in biomedicine far too low. In particle physics, significance is set at 5 sigma—a p value of 3×10^{-7} or 1 in 3.5 million.** If the result is not true, this is the probability that the data would have been as extreme as they are). The conclusion of the symposium was that something must be done. Indeed, all seemed to agree that it was within our power to do that something. But as to precisely what to do or how to do it, there were no firm answers. Those who have the power to act seem to think somebody else should act first. And every positive action (eg, funding well-powered replications) has a counterargument (science will become less creative). The good news is that science is beginning to take some of its worst failings very seriously. The bad news is that nobody is ready to take the first step to clean up the system.

Richard Horton
richard.horton@lancet.com



Richard Horton