



## Point-Counterpoint

### Point: Clarifying Policy Evidence With Potential-Outcomes Thinking—Beyond Exposure-Response Estimation in Air Pollution Epidemiology

Corwin Matthew Zigler\* and Francesca Dominici

\* Correspondence to Dr. Corwin Matthew Zigler, 655 Huntington Avenue, Building 2, 4th Floor, Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115 (e-mail: czigler@hsph.harvard.edu).

Initially submitted January 8, 2014; accepted for publication May 19, 2014.

The regulatory environment surrounding policies to control air pollution warrants a new type of epidemiologic evidence. Whereas air pollution epidemiology has typically informed policies with estimates of exposure-response relationships between pollution and health outcomes, these estimates alone cannot support current debates surrounding the actual health effects of air quality regulations. We argue that directly evaluating specific control strategies is distinct from estimating exposure-response relationships and that increased emphasis on estimating effects of well-defined regulatory interventions would enhance the evidence that supports policy decisions. Appealing to similar calls for accountability assessment of whether regulatory actions impact health outcomes, we aim to sharpen the analytic distinctions between studies that directly evaluate policies and those that estimate exposure-response relationships, with particular focus on perspectives for causal inference. Our goal is not to review specific methodologies or studies, nor is it to extoll the advantages of “causal” versus “associational” evidence. Rather, we argue that potential-outcomes perspectives can elevate current policy debates with more direct evidence of the extent to which complex regulatory interventions affect health. Augmenting the existing body of exposure-response estimates with rigorous evidence of the causal effects of well-defined actions will ensure that the highest-level epidemiologic evidence continues to support regulatory policies.

accountability; air pollution; clean air act; health outcomes; particulate matter

Abbreviations: CAA, Clean Air Act; EPA, Environmental Protection Agency.

**Editor's note:** A counterpoint to this article appears on page 1141.

A new regulatory environment invites a new brand of epidemiologic evidence. The claim that exposure to ambient air pollution is harmful to human health is hardly controversial in this day and age, largely because of the evidence amassed through decades of epidemiologic research on air pollution. This body of research focused historically on hazard identification and more recently on estimation of exposure-response (or, more formally, concentration-response) functions relating how health outcomes differ with spatial and/or temporal variations in ambient pollution exposure (1–9). Although considerable uncertainty remains with regard to essential finer-grade issues, such as the specific shape of the exposure-

response functions, the mechanics of exactly how pollution harms the human body, and the achievement of an adequate margin of safety dictated by the US Clean Air Act (CAA), evidence of the exposure-response relationship between pollution and health has motivated a vast array of air quality-control policies in the United States and abroad. The collection of these measures has undeniably improved ambient air quality over the past several decades (10, 11).

Despite the success of such regulatory policies for cleaning the air, an evolving regulatory and political environment is placing new demands on input from the scientific community. With the prospect of increasing costs resulting from proposed tightening of air quality standards, the evidence that motivated these policies is being subject to unprecedented scrutiny, and the scientific community must adapt by providing new types of evidence to support current and future

regulatory strategies (11, 12). Policy makers, legislators, leaders of industry, and the public increasingly emphasize questions of whether past efforts have actually yielded demonstrable improvements to public health, whether the costs associated with implementation of control policies such as the CAA (e.g., annual costs of the 1990 amendments are estimated to reach \$65 billion by 2020 (13)) are justified, and which existing strategies have provided the greatest health benefits. These considerations reflect a shifting demand toward evidence of effectiveness of specific regulatory interventions. Starting most notably with a 2003 report from the Health Effects Institute (14), questions of so-called accountability assessment—assessment of the extent to which regulatory

actions taken to control air quality affect health outcomes—have been propelled to the forefront of policy debates. A National Research Council report commissioned by the US Congress recommended that an enhanced air quality–management system strive to take a more performance-oriented approach by tracking the effectiveness of specific control policies and creating accountability for results, with similar calls for the importance of accountability echoed by others, including the Environmental Protection Agency (EPA) (15–18). Increased emphasis on the direct study of the effectiveness of specific actions is one essential avenue to ensuring that epidemiologic research continues to inform air quality–control policies amid the current regulatory climate.

**Table 1.** Existing Accountability Studies Classified According to the Causal Question of Interest, 1993–2013

Category	First Author, Year (Reference No.)	Factors Studied	Direct or Indirect Accountability	Causal Analysis Questions
A	Dockery, 1993 (1)	PM <sub>2.5</sub> , PM <sub>10</sub> , mortality	Indirect	What is the association between pollution exposure and health? Is this a causal association (classical paradigm)?
A	Laden, 2006 (5)	PM <sub>2.5</sub> , mortality	Indirect	
A	Zeger, 2008 (6)	PM <sub>2.5</sub> , mortality	Indirect	
A	Pope, 2009 (7)	PM <sub>2.5</sub> , life expectancy	Indirect	
A	Correia, 2013 (8)	PM <sub>2.5</sub> , life expectancy	Indirect	
B	Pope, 1996 (2)	Utah Valley Steel Mill, PM <sub>10</sub> , various health indicators	Indirect	What is the causal effect of differential exposure to pollution on health (potential outcomes paradigm)?
B	Chay, 2003 (34)	1981–1982 recession, TSP, infant mortality	Indirect	
B	Pope, 2007 (35)	Copper Smelter Strike, SO <sub>2-4</sub> , mortality	Indirect	
B	Moore, 2010 (36)	O <sub>3</sub> , asthma	Indirect	
B	Currie, 2011 (37)	New Jersey E-Z Pass data, birth outcomes	Indirect	
B	Rich, 2012 (38)	Beijing Olympics, PM <sub>2.5</sub> , cardiovascular biomarkers	Indirect	What is the causal effect of the intervention on health (potential outcomes paradigm)?
B	Chen, 2013 (9)	Huai River Policy, TSP, life expectancy	Indirect	
C	Friedman, 2001 (3)	1996 Atlanta Olympics, traffic, O <sub>3</sub> , asthma	Direct	
C	Hedley, 2002 (39)	Hong Kong Sulfur Restriction, sulfur dioxide, mortality	Direct	
C	Clancy, 2002 (28)	Dublin Coal Ban, black smoke, mortality	Direct	
C	Tonne, 2008 (40)	London Traffic Charging, NO <sub>2</sub> , PM <sub>10</sub> , life expectancy	Direct	
C	Chay, 2003 (41)	1970 CAA, TSP, adult mortality	Direct	
C	Greenstone, 2004 (42)	1970 CAA, sulfur dioxide	Direct	
C	Zigler, 2012 (43)	1990 PM <sub>10</sub> nonattainment, PM <sub>10</sub> , mortality	Direct	
C	Deschenes, 2012 (44)	NO <sub>x</sub> Budget Program, O <sub>3</sub> , pharmaceutical expenditures, mortality	Direct	

Abbreviations: CAA, Clean Air Act; NO<sub>x</sub>, NO<sub>2</sub>, nitrogen oxides; O<sub>3</sub>, ozone; PM<sub>10</sub>, PM<sub>2.5</sub>, particulate matter; SO<sub>2</sub>, sulfur dioxide; SO<sub>4</sub>, sulfate; TSP, total suspended particles.

Although the 10-plus years after the initial report from the Health Effects Institute saw an increase in studies framed as accountability studies (Table 1) (19–22), these studies have been heterogeneous with regard to analytic perspective and specificity of evidence. Many share accountability objectives but are actually the type of exposure-response studies that have been common in air pollution epidemiology for decades, and as such are not the most direct means for evaluating the effectiveness of specific policies. Relatively few accountability studies are designed to directly evaluate policies in line with the initial recommendations of the Health Effects Institute Report (14), and consideration of complex long-term interventions of direct relevance to regulatory policy has been particularly sparse. The goal of the present commentary is to sharpen the distinctions initially raised in the Health Effects Institute report (14), with particular regard to analytic perspectives on causal inference using observational data. Ultimately, we argue for increased emphasis on perspectives rooted in a potential-outcomes paradigm for causal inference to directly evaluate air quality regulations, highlighting distinctions between this endeavor and estimation of exposure-response relationships. We contextualize existing accountability studies as either direct or indirect accountability assessment, discuss the role of causal inference in air pollution accountability, and highlight several salient challenges with illustrative examples.

#### EXISTING ACCOUNTABILITY STUDIES: DIRECT OR INDIRECT ASSESSMENT?

Table 1 lists a variety of studies that have been integral to the discussion of accountability assessment and the formation of existing air quality-control policies. Each study is classified according to the scientific question of interest. Studies in categories A and B, which we term indirect accountability studies, answer questions of the form, “What is the relationship between exposure to pollution and health outcomes?” This type of question has been at the center of air pollution epidemiology for decades, and answers typically come in the form of exposure-response relationships between (changes in) pollution exposure and (changes in) health outcomes. Importantly, these studies do not consider the effectiveness of any specific regulatory action, but rather provide valuable evidence for indirectly predicting the impact of policies. For example, the EPA routinely uses exposure-response estimates to estimate the expected benefits of current and future policies; if a policy reduces (or is expected to reduce) pollution by a certain amount, then the exposure-response relationship indirectly implies the health impact of the policy insofar as the relationship can be deemed causal (10, 13, 23). We defer discussion of causality to later in the article but note here that this approach assumes that any observed exposure-response relationship would persist amid the complex realities of actual regulatory implementation that will typically affect a variety of factors. As a consequence, the health impacts of regulatory interventions may not be accurately characterized by indirectly applying exposure-response estimates to accountability assessments.

In contrast, studies in category C in Table 1 target a different scientific question that is of more direct relevance to accountability assessment. Rather than investigate the rela-

tionship between pollution and health, these studies answer the question, “What is the relationship between a specific regulatory intervention and health?” These studies are direct accountability studies in that they directly evaluate the effectiveness of well-defined regulatory actions, which more definitively informs questions as to the actual health benefits of these actions. Although relatively less common to air pollution epidemiology than studies of exposure-response relationships, we argue that direct accountability assessments are best equipped to meet the demands of a shifting regulatory environment wrought with questions surrounding the effectiveness of specific policies. Of particular importance is the noted lack of direct evaluations of broad, complex regulatory interventions, which are of the utmost relevance to policy debates (20–22).

#### CAUSAL ASSOCIATIONS, CAUSAL EFFECTS, AND THE EXPERIMENTAL PARADIGM

The role of causality is of obvious import for informing policy decisions, and the causal validity (or lack thereof) of epidemiologic evidence has always been central to the integration of scientific evidence into policy recommendations (10). However, approaches to inferring causality from available observational data can vary depending on the scientific question of interest and the data available for analysis.

Causal inference in air pollution epidemiology has most commonly been undertaken within a “classical” paradigm, which construes causal validity on a continuum according to how likely it is that an observed association (e.g., between pollution and health) can be interpreted as causal (24). This continuum is explicitly considered in the approach to Integrated Science Assessments conducted by EPA, which classify evidence of the association between pollution exposure and health as a “causal relationship,” “likely to be a causal relationship,” “suggestive of a causal relationship,” “inadequate to infer a causal relationship,” or “not likely to be a causal relationship” (10). Even in the absence of the word “causal,” the bulk of air pollution epidemiology has been implicitly undertaken with this classical approach; an exposure-response relationship between pollution and health is estimated (e.g., in a cohort study), then a judgment is made as to whether this relationship can be reasonably interpreted as causal, and finally, hypothetical changes in exposure are input into the exposure-response function to infer the resulting “health effect” that would be caused by such a change in pollution. Indirect accountability studies undertaken with a classical approach to causality are classified as category A in Table 1, and indeed they represent the bulk of epidemiologic research on air pollution being conducted today.

As an alternative to the classical paradigm, the potential-outcomes paradigm for causal inference has the distinctive feature that causal effects are explicitly defined as consequences of specific actions (25). Rather than infer causality based on belief of whether an estimated association can be interpreted as causal, potential-outcomes methods entail definition of a clearly-defined action (a “cause”), the effects of which are of interest. This perspective can clarify many threats to validity that plague accountability studies. Both indirect and direct accountability assessments have been

undertaken within a potential-outcomes paradigm for causal inference, the common thread being application of the core tenets of experimentation to observational settings. We elaborate how framing accountability studies in this way can clarify scientific objectives and possible threats to causal validity later in the article. Studies in Table 1 classified in categories B and C represent studies that are (often implicitly) framed as hypothetical experiments within a potential-outcomes paradigm. Importantly, the distinction between categories B and C is not the approach to causal inference per se, but rather the type of causal question being asked. Studies in category B are framed as hypothetical experiments to estimate the causal effect of differential levels of pollution exposure on health, rendering them indirect accountability studies of exposure-response relationships. Studies in category C frame actual air quality-control interventions as hypothetical experiments to estimate causal effects of these interventions, rendering them direct accountability studies of the effectiveness of specific interventions.

### CLARIFYING ACCOUNTABILITY ASSESSMENT WITH POTENTIAL OUTCOMES

The purpose of the present commentary is not to review specific methodologies or studies, nor is it to extoll the advantages of “causal” versus “associational” evidence. Rather, we argue that the shifting regulatory environment would be better informed by evidence of the effectiveness of specific control policies and that traditional epidemiologic approaches tailored to exposure-response estimation are not the most direct means to provide this evidence. In an environment that brings skepticism and doubt about results drawn from observational data, analyzing specific interventions with approaches rooted in potential-outcomes thinking can clarify the basis for drawing causal inferences and bring a higher level of credibility to evidence used to support policy decisions (12). Here, we outline this perspective as it relates to direct accountability assessment while alluding to challenges that have arisen and highlighting distinctions with traditional exposure-response estimation.

#### Accountability studies framed as approximate experiments: defining “the cause”

The underlying features of randomized studies that make them the gold standard for generating causal evidence remain pertinent to causal accountability assessment, with potential-outcomes methods framing observational studies according to how well they can approximate randomized experiments (26, 27). The key idea is to define a (possibly hypothetical) experiment consisting of an “intervention condition” and a “control condition” such that if populations could be randomly assigned to these conditions, differences in observed health outcomes would be interpreted as causal effects of the intervention. Although defining the intervention condition in accountability studies can be straightforward (e.g., it will likely be a regulatory action that actually occurred), framing accountability as a hypothetical experiment forces the specification of some alternative action that might have otherwise occurred to serve as a relevant control condition. This exercise

formalizes the research question by explicitly defining a causal effect as a comparison between what would happen under well-defined competing conditions, hence the name of the potential-outcomes paradigm; a causal effect of action A relative to action B is defined as the comparison of the potential outcome if action A were taken with the potential outcome if action B were taken. Thus, the salient question for accountability is not “Did health outcomes change after the intervention?” but rather “Are health outcomes different after the intervention than they would have been under a specific alternative action?” Of utmost importance is that the causal effect of interest is defined without regard to any assumed statistical model. Different models could be used to actually estimate this effect, but the effect itself, along with its interpretation, remains consistent regardless of the modeling approach. This clarity is essential for producing policy-relevant evidence. Compare this to traditional studies of exposure-response relationships, which 1) do not necessarily explicate an action defining effects of interest and 2) define health effects with parameters (e.g., regression coefficients) in a statistical model; that is, estimated health effects from 2 different models may not even share the same interpretation.

#### Confounding and estimating counterfactual scenarios

Estimating causal effects with comparisons between potential outcomes under competing intervention and control conditions is met with a fundamental problem: If the intervention is enacted, then outcomes under the control condition are unobserved. For example, evaluating the effect of a past regulatory policy requires knowledge of what would have potentially happened if the policy had not been implemented. Hypothetical scenarios that never actually occurred are often referred to as counterfactual scenarios, and estimating what would have happened under such scenarios is perhaps the most important challenge for direct accountability assessment.

Counterfactual scenarios have been explicitly considered, for example, in EPA cost-benefit analyses of the CAA mandated by section 812 of the Act, which project 2 counterfactual pollution scenarios: one that assumes past exposure patterns would have continued without the 1990 CAA amendments and another that assumes an expected change in exposure patterns under full implementation of the 1990 amendments. These projections are coupled with exposure-response functions from the epidemiologic literature to project counterfactual health scenarios that form the basis of the health-benefits analyses (13, 23). However, these counterfactual projections are not validated against studies of actual interventions and thus are not sufficient for fully characterizing the relationships between regulatory strategies and health (14).

Rather than project counterfactual scenarios by combining assumed exposure patterns with exposure-response estimates, potential-outcomes approaches typically use actual data from the control group of the hypothetical experiment to learn about what would have happened without the intervention, rendering identification of a control population of vital importance. When assessing the impact of regulatory strategies, control populations could be defined based on time (e.g., a population before promulgation of a regulation) or space (e.g., if some areas are subject to an intervention and others

are not). Whether outcomes in the control population can actually characterize what would have occurred without the intervention boils down to the familiar concept of confounding, although what exactly constitutes a confounder is slightly different than in the exposure-response setting.

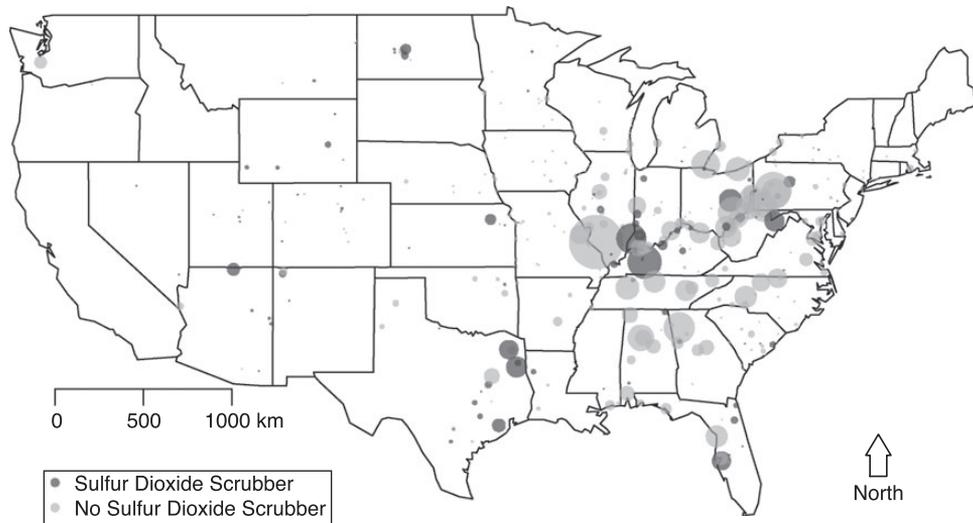
For direct accountability, a comparison between outcomes among the intervention and control conditions is not confounded if the 2 populations are comparable with regard to factors that relate to outcomes. A comparison of outcomes between the intervention and control conditions that is not confounded yields an estimate of the causal effect. If the 2 populations differ on important factors related to outcomes, such a comparison is a convolution of differences due to the intervention and differences due to other factors. Thus, if an important factor relating to health (e.g., smoking behavior) is comparable across the intervention and control populations, then that factor is not a confounder in the assessment of the intervention. Compare this to the typical setting of exposure-response studies, in which a confounder is generally regarded as a factor that is associated simultaneously with pollution exposure and health outcomes. In both settings, the definition of a confounder is a factor that is associated with exposure and outcome, the key difference being that in a direct accountability study, the exposure is actually the intervention, whereas in an indirect accountability study, the exposure is air pollution (Table 1).

There are a variety of analytic tools available to address confounding in nonrandomized accountability studies. Specialized study designs, often described as “quasi experiments,” circumvent the need to consider confounding directly because they support assumptions that an intervention was quasi-randomized in the sense that it is unrelated to health outcomes (12). Such studies have been primarily used for indirect accountability assessment (Table 1). Absent the availability of such specialized circumstances, methods for confounding adjustment (e.g., matching, weighting, stratification, or standardization) adjust for differences between intervention and control populations so that comparison groups can be regarded as similar on the basis of observed factors, thus mimicking the design of a randomized study. In either case, practical accommodation of confounding can be particularly challenging for air quality interventions, as we discuss in the context of the examples below.

### Two examples: localized action versus regulatory policy

We use 2 examples to illustrate specific features of framing direct accountability studies in a potential-outcomes paradigm. First, consider the accountability study of Clancy et al. (28) in which they investigated the health impacts of the ban on the sale and distribution of black coal in Dublin, Ireland. The coal ban represents a specific localized action that was followed by significant decreases in the concentration of black smoke immediately after the ban, with concurrent decreases in the number of deaths. As with many studies of abrupt, localized interventions, definition of the hypothetical experiment is straightforward; institution of the ban represents the intervention condition, with the control condition being no ban, and the causal effect of interest is that of instituting the ban versus the not instituting the ban. The counter-

factual scenario representing what would have happened without the ban is estimated using data from the time period immediately preceding the ban, that is, Dublin before the ban serves as a control group for Dublin after the ban. The key assumption permitting pre-ban conditions to represent what would have happened without the ban is that of temporal stability, which assumes that pre-ban health outcomes would not have changed (i.e., remained stable) if no ban had occurred (29). Localized interventions that result in immediate changes in pollution and health outcomes can often support assumptions such as temporal stability and obviate the need for sophisticated statistical methods to infer causality. However, even when studying an abrupt action, threats to causal validity can arise, as illustrated in extended analyses of the Dublin coal ban that revealed that long-term trends in cardiovascular health spanning implementation of the ban—not the coal ban itself—contributed to apparent effects on cardiovascular mortality (30). Thus, pre-ban Dublin was not an adequate control group because factors relating to cardiovascular health confounded the pre- versus post-ban comparisons. This violation of temporal stability was only determined after inclusion of other control areas that were not subject to coal bans. Similar threats to causal validity were illuminated through the inclusion of control populations in studies of the impact of transportation changes during the 1996 Olympic Games in Atlanta, Georgia (3, 31). These experiences speak to the importance of careful planning, possibly in the design stage of a prospective study, with regard to inclusion of appropriate control populations (17). Contrast the coal ban example with an accountability assessment of broad-scale regulatory policy measures, such as those emanating from Title IV of the 1990 CAA amendments that placed emissions limits on power-generating facilities, which bears relevance to the current debate over rules proposed by EPA to limit greenhouse gas emissions. Unlike a localized, abrupt action, measures to reduce power plant emissions represent a complex process comprised of a variety of actions targeting different pollutants at various time scales, which vastly complicates causal inference. Many links in the chain of accountability (14) could be of interest—causal effects on emissions (sulfur dioxide and others), on ambient air, and on health outcomes—but difficulties arise even in the definition of these effects, as the heterogeneity of actions taken does not point to a single clearly defined intervention. Defining the causal effect of instituting the emissions limits versus not instituting the limits is complicated by the fact that facilities were subject to different limits at various implementation phases, used different strategies to reduce emissions (e.g., air scrubbers, fuel shifts, low-sulfur coal), and were able to exceed limits by purchasing allowances on the open cap-and-trade market initiated as part of the Acid Rain Program. As one simplistic example to illustrate the specificity required to define causal effects in this setting, consider an accountability assessment of the extent to which installation of sulfur dioxide scrubbers on coal-burning plants during the first few years of the Acid Rain Program (1995–1997) impacted emissions, ambient air quality, and health outcomes. Figure 1 depicts the locations of 407 coal-burning power plants that participated in the Acid Rain Program during 1995–1997, distinguishing the 113 plants that installed sulfur dioxide scrubbers from the 294 plants that

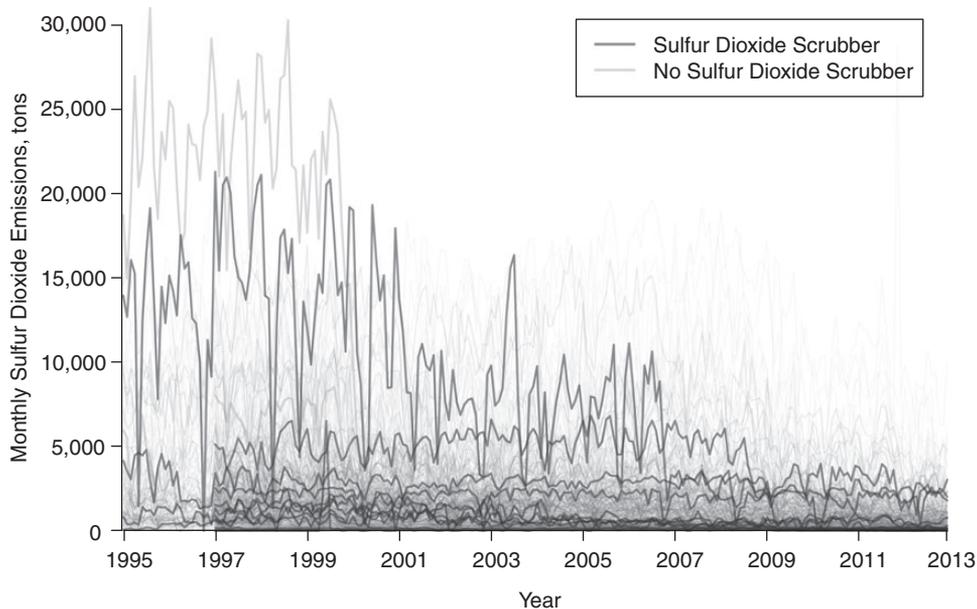


**Figure 1.** Locations of 407 coal-burning power plants participating in the Acid Rain Program in the United States, 1995–1997. The size of plotting symbol is proportional to the average number of tons of sulfur dioxide emitted at each location during 1995–1997.

did not. Figure 2 depicts monthly sulfur dioxide emissions in these plants from 1995 to 2012. The hypothetical experiment can be defined with an intervention condition comprising the pattern of scrubber installation that actually occurred during these years, and the control condition is the hypothetical setting in which no such scrubbers were installed during this time. This defines the causal effect of the scrubber installations on emissions, ambient air quality, and health outcomes

independently from other concurrent measures that may have been taken to control emissions.

To characterize the counterfactual scenario with no scrubbers during 1995–1997, the long time lag between scrubber installation and any measurable impact on health renders an analysis assuming temporal stability (e.g., pre- vs. post-scrubber comparisons) tenuous at best. Information about what would have happened without the scrubbers could be



**Figure 2.** Monthly sulfur dioxide emissions from 1995 to 2012 among coal-burning power plants participating in the Acid Rain Program during 1995–1997. Thick, bold lines correspond to facilities that had 1995–1997 sulfur dioxide emissions at each decile for the respective scrubber category.

gleaned during the same time frame from facilities that did not install scrubbers. Using facilities without sulfur dioxide scrubbers as a control group for those that did install scrubbers leads to at least 2 important complications. First is the reality that actions taken at a given plant could affect pollution and health outcomes in distant areas, including no-scrubber areas. This transport phenomenon, known in the statistical literature as interference, is an active area of current research in potential-outcomes methods (32, 33). Second, the success of using no-scrubber facilities to learn about what would have happened in and around facilities that did install scrubbers hinges on the ability to adjust for confounders to parse consequences of the scrubbers from inherent differences between types of facilities and their surroundings. Informally, confounding adjustment would ensure that emissions, ambient pollution, and health outcomes in and around facilities that installed a scrubber are only compared against those from a no-scrubber area that is comparable with respect to confounding factors (facility characteristics, controls for other pollutants, population demographics, historical pollution, etc.). Compare this perspective with one rooted in estimation of exposure-response associations, which would rely on estimates of the relationship between changes in sulfur dioxide emissions and changes in health outcomes, possibly comprised of separate estimates of the emissions-ambient air link and the ambient air-health link. Reliance on exposure-response functions in this setting would obscure the goal of accountability for specific, well-defined actions relative to a hypothetical experiment defining the causal effects of installing scrubbers (versus not installing scrubbers) on all outcomes of interest. Using a potential-outcomes approach for direct accountability assessment cannot escape the inherent difficulties of inferring causality with observational data but can serve to clarify the link between quantitative methods and the realities of evaluating broad, long-term regulatory policies. This clarity is essential for producing policy-relevant evidence.

## CONCLUSION

Over the past 10 years, important progress in accountability assessment has initiated a new dimension to the scientific evidence available for informing policy decisions. Important challenges remain, in particular for evaluating large-scale regulatory policies that are not characterized by a single action. We have attempted to sharpen the distinction between analytic perspectives for exposure-response estimation and for estimating causal effects of well-defined actions. Although the former has indirect relevance to accountability assessment, we argue that the latter perspective is necessary to advance accountability assessment beyond evaluation of localized, abrupt actions and toward informing policy debates with evidence of the effects of broad and complex regulations. Although no single analytic strategy can overcome all the challenges inherent to accountability, the best science should be generated from a variety of available approaches. We argue that rigorous efforts to directly evaluate causal effects of well-defined regulatory interventions constitute one such approach that, although distinct from traditional

epidemiologic tools, is essential to the current regulatory climate.

## ACKNOWLEDGMENTS

Author affiliation: Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts (Corwin Matthew Zigler, Francesca Dominici).

This work was supported by Health Effects Institute 4909 and US Environmental Protection Agency (EPA) grants RD83479801 and 83489401-0.

We thank Dr. Jonathan Samet for comments that helped improve this commentary and Dr. Christine Choirat for data processing to produce the power plant example.

This publication's contents are solely the responsibility of the grantee and do not necessarily represent the official views of the US EPA. Further, US EPA does not endorse the purchase of any commercial products or services mentioned in the publication.

Conflict of interest: none declared.

## REFERENCES

1. Dockery DW, Pope CA III, Xu X, et al. An association between air pollution and mortality in six U.S. cities. *N Engl J Med*. 1993;329(24):1753–1759.
2. Pope CA III. Particulate pollution and health: a review of the Utah valley experience. *J Expo Anal Environ Epidemiol*. 1996; 6(1):23–34.
3. Friedman MS, Powell KE, Hutwagner L, et al. Impact of changes in transportation and commuting behaviors during the 1996 Summer Olympic Games in Atlanta on air quality and childhood asthma. *JAMA*. 2001;285(7):897–905.
4. Krewski D, Burnett R, Goldberg M, et al. Overview of the reanalysis of the Harvard Six Cities Study and American Cancer Society Study of Particulate Air Pollution and Mortality. *J Toxicol Environ Health A*. 2003;66(16-19):1507–1551.
5. Laden F, Schwartz J, Speizer FE, et al. 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(6):667–672.
6. Zeger SL, Dominici F, McDermott A, et al. Mortality in the Medicare population and chronic exposure to fine particulate air pollution in urban centers (2000–2005). *Environ Health Perspect*. 2008;116(12):1614–1619.
7. Pope CA III, Ezzati M, Dockery DW. Fine-particulate air pollution and life expectancy in the United States. *N Engl J Med*. 2009;360(4):376–386.
8. Correia AW, Pope CA III, Dockery DW, et al. Effect of air pollution control on life expectancy in the United States: an analysis of 545 U.S. counties for the period from 2000 to 2007. *Epidemiology*. 2013;24(1):23–31.
9. Chen Y, Ebenstein A, Greenstone M, et al. Evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai River policy. *Proc Natl Acad Sci U S A*. 2013;110(32):12936–12941.
10. U.S. Environmental Protection Agency. *Integrated Science Assessment for Particulate Matter*. Research Triangle Park, NC: U.S. Environmental Protection Agency Office of Research and Development; 2009.

11. Samet JM. The Clean Air Act and health—a clearer view from 2011. *N Engl J Med*. 2011;365(3):198–201.
12. Dominici F, Greenstone M, Sunstein CR. Science and regulation. Particulate matter matters. *Science*. 2014;344(6181):257–259.
13. US EPA Office of Air and Radiation. *The Benefits and Costs of the Clean Air Act: 1990 to 2020*. Research Triangle Park, NC: US EPA Office of Air and Radiation; 2010. (Revised Draft Report).
14. HEI Accountability Working Group. *Assessing the Health Impact of Air Quality Regulations: Concepts and Methods for Accountability Research*. Boston, MA: Health Effects Institute; 2003. (Communication 11).
15. National Research Council. *Air Quality Management in the United States*. Washington, DC: National Academies Press; 2004.
16. Hidy GM, Brook JR, Demerjian KL, et al. *Technical Challenges of Multipollutant Air Quality Management*. New York, NY: Springer; 2011.
17. Hubbell B. Assessing the results of air quality management programs. *EM Magazine*. 2012;8–15.
18. US Environmental Protection Agency. *Designing Research to Assess Air Quality and Health Outcomes From Air Pollution Regulations*, Research Triangle Park, NC: 2013. <http://www.cleanaairinfo.com/finepmpolicy/index.htm>. Accessed October 29, 2014.
19. van Erp AM, Cohen AJ. *HEI's Research Program on the Impact of Actions to Improve Air Quality: Interim Evaluation and Future Directions*. Boston, MA: Health Effects Institute; 2009.
20. Health Effects Institute. *Proceedings of an HEI Workshop on Further Research to Assess the Health Impacts of Actions Taken to Improve Air Quality*. Boston, MA: Health Effects Institute; 2010.
21. van Erp AM, Cohen AJ, Shaikh R, et al. Recent progress and challenges in assessing the effectiveness of air quality interventions towards improving public health: the HEI experience. *EM Magazine*. 2012;22–28.
22. van Erp AM, Kelly FJ, Demerjian KL, et al. Progress in research to assess the effectiveness of air quality interventions towards improving public health. *Air Qual Atmos Health*. 2012; 5(2):217–230.
23. US EPA. *Environmental Benefits Mapping and Analysis Program (BenMAP) User's Manual*. Research Triangle Park, NC: US EPA; 2012. (Technical Report).
24. Glass TA, Goodman SN, Hernán MA, et al. Causal inference in public health. *Annu Rev Public Health*. 2013;34:61–75.
25. Rubin DB. Bayesian inference for causal effects: the role of randomization. *Ann Stat*. 1978;6(1):34–58.
26. Rubin DB. For objective causal inference, design trumps analysis. *Ann Appl Stat*. 2008;2(3):808–840.
27. Hernán MA, Alonso A, Logan R, et al. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. *Epidemiology*. 2008;19(6):766–779.
28. Clancy L, Goodman P, Sinclair H, et al. Effect of air-pollution control on death rates in Dublin, Ireland: an intervention study. *Lancet*. 2002;360(9341):1210–1214.
29. Holland PW. Statistics and causal inference. *J Am Stat Assoc*. 1986;81(396):945–960.
30. Dockery DW, Rich DQ, Goodman PG, et al. *Effect of Air Pollution Control on Mortality and Hospital Admissions in Ireland*. Boston, MA: Health Effects Institute; 2013.
31. Peel JL, Klein M, Flanders WD, et al. *Impact of Improved Air Quality During the 1996 Summer Olympic Games in Atlanta on Multiple Cardiovascular and Respiratory Outcomes*. Boston, MA: Health Effects Institute; 2010.
32. Hudgens MG, Halloran ME. Toward causal inference with interference. *J Am Stat Assoc*. 2008;103(482):832–842.
33. Tchetgen EJT, VanderWeele TJ. On causal inference in the presence of interference. *Stat Methods Med Res*. 2012;21(1): 55–75.
34. Chay KY, Greenstone M. The impact of air pollution on infant mortality: evidence from geographic variation in pollution shocks induced by a recession. *Q J Econ*. 2003;118(3): 1121–1167.
35. Pope CA III, Rodermund DL, Gee MM. Mortality effects of a copper smelter strike and reduced ambient sulfate particulate matter air pollution. *Environ Health Perspect*. 2007;115(5): 679–683.
36. Moore K, Neugebauer R, Lurmann F, et al. Ambient ozone concentrations and cardiac mortality in Southern California 1983–2000: application of a new marginal structural model approach. *Am J Epidemiol*. 2010;171(11):1233–1243.
37. Currie J, Walker R. Traffic congestion and infant health: evidence from E-ZPass. *Am Econ J Appl Econ*. 2011;3(1): 65–90.
38. Rich DQ, Kipen HM, Huang W, et al. Association between changes in air pollution levels during the Beijing Olympics and biomarkers of inflammation and thrombosis in healthy young adults. *JAMA*. 2012;307(19):2068–2078.
39. Hedley AJ, Wong C-M, Thach TQ, et al. Cardiorespiratory and all-cause mortality after restrictions on sulphur content of fuel in Hong Kong: an intervention study. *Lancet*. 2002;360(9346): 1646–1652.
40. Tonne C, Beevers S, Armstrong B, et al. Air pollution and mortality benefits of the London Congestion Charge: spatial and socioeconomic inequalities. *Occup Environ Med*. 2008; 65(9):620–627.
41. Chay K, Dobkin C, Greenstone M. The clean air act of 1970 and adult mortality. *J Risk Uncertain*. 2003;27(3):279–300.
42. Greenstone M. Did the clean air act cause the remarkable decline in sulfur dioxide concentrations? *J Environ Econ Manage*. 2004;47(3):585–611.
43. Zigler CM, Dominici F, Wang Y. Estimating causal effects of air quality regulations using principal stratification for spatially correlated multivariate intermediate outcomes. *Biostatistics*. 2012;13(2):289–302.
44. Deschenes O, Greenstone M, Shapiro JS. *Defensive Investments and the Demand for Air Quality: Evidence from the NO<sub>x</sub> Budget Program and Ozone Reductions*. Cambridge, MA: National Bureau of Economic Research; 2012 (Working Paper 18267).