Demystifying Causal Inference In Air Pollution Epidemiology

Corwin Zigler and Francesca Dominici
Department of Biostatistics, Harvard T.H. Chan School of Public Health

Background: The promise of identifying causal relationships between, for example, air pollution exposure and human health, has rightly generated ample enthusiasm in air pollution research. However, confusion remains as to what exactly makes a method a causal one, how exactly such methods differ from traditional approaches, and where evidence derived from causal methods fits with the large body of research on the relationships between air pollution and health.

Workshop Objective: To promote literacy among the general air pollution community on issues related to inferring causal relationships in observational air pollution research. To provide a framework for the general consumer of air pollution research to understand and assess the increasing amount of work being labeled as "'causal."

Target Audience: The general informed consumer of air pollution research. Expertise in statistical methods and/or econometrics is not required, but participants are expected to have familiarity with the epidemiological literature pertaining to air pollution and health.

Description of Workshop: This workshop will provide a broad overview of inferring causal relationships in air pollution studies. The focus is not instruction on specific statistical methods, but rather the description of a foundational perspective on methods for causal inference that underlie a large body of air pollution research. Frequent and interactive use of familiar examples will help illustrate how a potential-outcomes perspective on causal inference can shed light on different types of research studies and designs, regardless of whether such studies are explicitly labeled as "causal." Participants should leave this workshop with: a) improved ability to determine what makes a particular research study "causal," b) a framework with which to evaluate the assumptions underlying the causal validity of a particular result, c) improved ability to place the evidence from causal studies in proper context with the vast literature on the relationships between pollution exposure and human health.

Requested background reading:

- Glass et al. (2013). Causal inference in public health. *Annual Review of Public Health* 34.
- Zigler and Dominici (2014). Point: Clarifying Policy Evidence With Potential-Outcomes Thinking - Beyond Exposure-Response Estimation in Air Pollution Epidemiology. *American Journal of Epidemiology* 180(12).

Outline

I. Introduction: Causal Inference: What it is (and what it is not)

- **i.** Introduction of familiar "causal" questions that will be used to anchor workshop discussion
- **ii.** What causal inference is: A general analytic perspective
- **iii.** What causal inference is not: A specific method or a magic silver bullet
- **iv.** "Classical" paradigm for causal inference and biologic causality
- **v.** "Potential Outcomes" paradigm for causal inference and consequences of (possibly hypothetical) actions

II. Framing Observational Studies as Approximate Randomized Experiments

- **i.** Prelude: Why randomized experiments are the "gold standard" for causal inference
- **ii.** The experimental paradigm for observational data: "Designing" a hypothetical experiment that defines a causal effect
- **iii.** "Analyzing" the hypothetical experiment to estimate the causal effect
- **iv.** Key assumptions for causal validity
- **v.** An overview of some relevant methods
  - Methods for observed confounding adjustment
  - Methods for unmeasured confounding

———— Break ————

III. Rapid Fire Examples

Discussion (with audience participation via Q/A sheet distributed during the break) of several familiar studies from the literature, all from a causal inference perspective. Examples will involve studies that have been explicitly labeled as "causal" as well as those that have not, including (among others) a causal evaluation of the Six Cities Study.

IV. Putting it All Together: Illustrative Case Study: Causal effects of $PM_{2.5}$ Nonattainment Designations

V. Questions/Open Discussion

Demystifying Causal Inference in Air Pollution Epidemiology

Instructors

Francesca Dominici, *Senior Associate Dean for Research and Professor of Biostatistics, Harvard T.H. Chan School of Public Health*

Francesca Dominici is professor of biostatistics and senior associate dean for research at the Harvard T.H. Chan School of Public Health. She has extensive experience on the development of statistical methods for the analysis of big data with the ultimate goal of evaluating public health and clinical interventions. Dominici has led several national studies on the health effects of air pollution that have directly informed National Ambient Air Quality Standards. Recently she conducted a nationwide study on the effects of noise exposure near the 90 largest U.S. airports on cardiovascular disease. She has assembled an interdisciplinary group of scientists across the University to project the future public health impact of extreme weather events. She is working with clinicians at the Dana-Farber Cancer Institute to develop methods for the analysis of large administrative data for precision medicine—that is, to understand which patients benefit most from a given cancer treatment regimen. As senior associate dean for research, Dominici leads a course on grant writing for the faculty at the Harvard Chan School and leads several data science–centered initiatives at the School. She has served on a number of committees of the National Academies. She has received multiple awards recognizing her commitment to diversity. She has moderated and participated in multiple panel discussions on work-life balance across Harvard. She also co-chairs the University Committee for the Advancement of Women Faculty at the Harvard Chan School. Dominici received her PhD in statistics at the University of Padua in Italy.

Corwin Zigler, A*ssistant Professor of Biostatistics, Harvard T.H. Chan School of Public Health*

Corwin Zigler, Ph.D., is an assistant professor of biostatistics at the Harvard T.H. Chan School of Public Health. He specializes in the development of Bayesian statistical methods for causal inference with observational data. He has spent the last several years engaged in the development and deployment of causal inference methods to compare the effectiveness of complex public health interventions, with particular focus on air quality interventions. He was principal investigator of a recently completed HEI grant entitled "Causal Inference Methods for Estimating Long-Term Health Effects of Air Quality Regulations," and his work has appeared in journals such as *Journal of the American Statistical Association, Biometrics, Annals of Applied Statistics, Biostatistics, American Journal of Epidemiology,* and many other top peer-reviewed journals. He received his Ph.D from the UCLA School of Public Health in 2010.

Clarifying Policy Evidence with Potential-Outcomes Thinking: Beyond Exposure-Response Estimation in Air Pollution Epidemiology

Running Head: Beyond exposure-response

Corwin Matthew Zigler* Francesca Dominici

Department of Biostatistics

Harvard School of Public Health

Boston, MA

czigler@hsph.harvard.edu

Abstract

The regulatory environment surrounding air pollution control policies warrants a new type of epidemiological 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 impacts of air quality regulations. This commentary argues that directly evaluating specific control strategies is distinct from estimating exposure-response, and that increased emphasis on estimating effects of well-defined regulatory interventions would enhance the evidence supporting 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, 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"

*Corresponding author

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 impact 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 epidemiological evidence continues to support regulatory policies.

Key words: Accountability, Air pollution, Clean Air Act, Health outcomes, Particulate matter

Abbreviations: Clean Air Act (CAA), Environmental Protection Agency (EPA)

1 A New Regulatory Environment Invites a New Brand of Epidemiological Evidence

The claim that exposure to ambient air pollution is harmful to human health is hardly controversial in this day and age, due in large part to the evidence amassed through decades of air pollution epidemiological research. This body of research historically focused on hazard identification and more recently 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 US 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 motivating 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, 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 reaching $65 billion by 2020 [13]) are justified, and which existing strategies have provided the greatest health benefits. These considerations reflect a shifting demand towards 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 impact 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 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 epidemiological research continues to inform air quality control policies amid the current regulatory climate.

While the ten-plus years following HEI's initial report has seen an increase in studies framed as accountability (see Table 1 as well as [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 in [14], and consideration of complex long-term interventions of direct relevance to regulatory policy has been particularly sparse. The goal of this commentary is to sharpen the distinctions initially raised in [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 estimating exposure-response. Section 2 contextualizes existing accountability studies as either direct or indirect accountability. Section 3 discusses the role of causal inference in air pollution accountability. Section 4 highlights several salient challenges with illustrative examples.

2 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 provide valuable evidence for indirectly predicting the impact of policies. For example, 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 Section 3, 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 impact a variety of factors. As a consequence, health impacts of regulatory interventions may not be accurately characterized by indirectly applying exposure-response estimates to accountability assessments.

In contrast, studies labeled as Category C in Table 1 target a different scientific question of more direct relevance to accountability assessment. Rather than investigate the relationship 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. While relatively less common to air pollution epidemiology than studies of exposure-response, 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 utmost relevance to policy

debates [20–22].

3 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 epidemiological 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 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." 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 represent the bulk of air pollution epidemiological research 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. Section 4 elaborates how framing accountability studies in this way can clarify scientific objectives and possible threats to causal validity. 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 to 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. 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.

4 Clarifying Accountability Assessment with Potential Outcomes

The purpose of this 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 epidemiological 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.

4.1 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. While 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 definition of the causal effect of interest is conducted *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, which a) do not necessarily explicate an action defining effects of interest and b) define "health effects" with parameters (e.g., regression coefficients) in a statistical model, i.e., estimated health effects from two different models may not even share the same interpretation.

4.2 Confounding and Estimating "Counterfactual" Scenarios

Estimating causal effects with comparisons between potential outcomes under competing "intervention" and "control" conditions is met with the fundamental problem that if the "intervention" is enacted, then outcomes under "control" 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 two 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 epidemiological 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 not). Whether outcomes in the control population can actually characterize what would have occurred absent 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 unconfounded if the two populations are comparable with regard to factors that relate to outcomes. An unconfounded comparison of outcomes between the intervention and control conditions yields an estimate of the causal effect. If the two populations differ on important factors related to outcomes, then 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, for example, smoking behavior, is comparable across the intervention and control populations, then smoking behavior is not a confounder in the assessment of the intervention. Compare this to the typical setting of exposure-response studies, where a "confounder" is generally regarded as a factor that is simultane-

ously associated 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 (see 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, as 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 (see 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 of Section 4.3.

4.3 Two Examples: Localized Action vs. Regulatory Policy

We use two examples to illustrate specific features of framing direct accountability studies in a potential-outcomes paradigm. First, consider the accountability study of [28] that investigates 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 following the ban, with concurrent decreases in mortality. 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" of no ban, and the causal effect of interest is that of instituting the ban vs. the not instituting the ban. The counterfactual 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 revealing 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- vs. 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 [3, 31]. These experiences speak to the importance of careful planning, possibly in the design stage of a prospective study, about 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 placing 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 of 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 vs. not instituting the limits is complicated by the fact that facilities were subject to different limits at various implementation phases, employed different strategies to reduce emissions (e.g., scrubbers, fuel shifts, low-sulfur coal, etc.), 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 did not. Figure 2 depicts monthly sulfur dioxide emissions in these plants from 1995-2012. The hypothetical experiment can be defined with an "intervention condition" comprised of the pattern of scrubber installation that actually occurred during these years, and the "control condition" the hypothetical setting where 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 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 is met with at least two important complications. First is the reality that actions taken at a given plant could impact 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, 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.

5 Conclusion

Over the past ten 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. While the former has indirect relevance to accountability assessment, we argue that the latter perspective is necessary in order to advance accountability assessment beyond evaluation of localized, abrupt actions and towards informing policy debates with evidence of the effects of broad and complex regulations. While 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, while distinct from traditional epidemiological tools, is essential to the current regulatory climate.

Acknowledgments

This work was supported by HEI 4909 and US EPA RD83479801 and 83489401-0. 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.

The authors thanks Dr. Jonathan Samet and three anonymous reviewers for comments that helped improve this commentary, and Dr. Christine Choirat for data processing to produce the power plant example. No conflict of interests stated by either author.

References

[1] Dockery, D. W, Pope III, C. A, Xu, X., Spengler, J. D, Ware, J. H, Fay, M. E, Ferris, B. G, and Speizer, F. E. An association between air pollution and mortality in six US cities. *New England journal of medicine*. 1993; 329(24):1753—1759.

[2] Pope III, C. A. Particulate pollution and health: a review of the utah valley experience. *Journal of exposure analysis and environmental epidemiology*. 1996;6(1):23—34.

[3] Friedman, Michael S., Powell, Kenneth E., Hutwagner, Lori, Graham, LeRoy M., and Teague, W. Gerald. Impact of changes in transportation and commuting behaviors during the 1996 summer olympic games in atlanta on air quality and childhood asthma. *JAMA: the journal of the American Medical Association*. 2001; 285(7):897–905.

[4] Krewski, Daniel, Burnett, Richard, Goldberg, Mark, Hoover, B. Kristin, Siemiatycki, Jack, Jerrett, Michael, Abrahamowicz, Michal, and White, Warren. Overview of the reanalysis of the harvard six cities study and american cancer society study of particulate air pollution and mortality. *Journal of Toxicology and Environmental Health Part A*. 2003;66(16-19):15071552.

[5] Laden, Francine, Schwartz, Joel, Speizer, Frank E, and Dockery, Douglas W. Reduction in fine particulate air pollution and mortality: Extended follow-up of the harvard six cities study. *Am. J. Respir. Crit. Care Med.* 2006;pages 200503–443OC.

[6] Zeger, Scott L., Dominici, Francesca, McDermott, Aidan, and Samet, Jonathan M. Mortality in the medicare population and chronic exposure to fine particulate air pollution in urban centers {(2000-2005)}. *Environmental Health Perspectives*. 2008;116(12):1614–1619. ISSN 0091-6765.

[7] Pope III, C. A, Ezzati, M., and Dockery, D. W. Fine-particulate air pollution and life expectancy in the united states. *New England Journal of Medicine*. 2009;360(4):376—386.

[8] Correia, Andrew W., Pope, C. Arden, Dockery, Douglas W., Wang, Yun, Ezzati, Majid, and Dominici, Francesca. 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. ISSN 1044-3983.

[9] Chen, Yuyu, Ebenstein, Avraham, Greenstone, Michael, and Li, Hongbin. Evidence on the impact of sustained exposure to air pollution on life expectancy from chinas huai river policy. *Proceedings of the National Academy of Sciences*. 2013;110(32):12936—12941.

[10] {U.S. EPA}. *Integrated Science Assessment for Particulate Matter*. U.S. Environmental Protection Agency Office of Research and Development, Research Triangel Park, NC. 2009.

[11] Samet, Jonathan M. The clean air act and health a clearer view from 2011. *New England Journal of Medicine*. 2011;365(3):198–201. ISSN 0028-4793.

[12] Dominici, Francesca, Greenstone, Michael, and Sunstein, Cass R. Particulate matter matters. *Science*. 2014; 344(6181):257–259. ISSN 0036-8075, 1095-9203.

[13] {U.S. EPA}. The benefits and costs of the clean air act: 1990 to 2020. Revised draft report, US EPA Office of Air and Radiation, Research Triangel Park, NC. 2010.

[14] {HEI Accountability Working Group}. *Assessing the Health Impact of Air Quality Regulations: Concepts and Methods for Accountability Research. Communication 11.* Health Effects Institute, Boston, MA. 2003.

[15] {National Research Council}. *Air Quality Management in the United States*. National Academies Press, Washington, DC. 2004.

[16] Hidy, G.M., Brook, J.R., Demerjian, K.L., Pennell, W.T., and Scheffe, R.D. *Technical Challenges of Multipollutant Air Quality Management*. Springer, New York, NY. 2011.

[17] Hubbell, Bryan. Assessing the results of air quality management programs. *EM Magazine*. 2012;pages 8–15.

[18] {U.S. EPA}. Workshop on designing research to assess air quality and health outcomes from air pollution regulations. In *Designing Research to Assess Air QuAlity and Health Outcomes from Air Pollution Regulations*. Research Triangel Park, NC. 2013; .

[19] van Erp, Annemoon M. and Cohen, Aaron J. *HEI's Research Program on the Impact of Actions to Improve Air Quality: Interim Evaluation and Future Directions*. Health Effects Institute, Boston, MA. 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*. Health Effects Institute, Boston, MA. 2010.

[21] van Erp, Annemoon M, Cohen, Aaron J., Shaikh, Rashid, and O'Keefe, Robert. Recent progress and challenges in assessing the effectiveness of air quality interventions towards improving public health: the HEI experience. *EM Magazine*. 2012;pages 22–28.

[22] van Erp, Annemoon M., Kelly, Frank J., Demerjian, Kenneth L., III, C. Arden Pope, and Cohen, Aaron J. Progress in research to assess the effectiveness of air quality interventions towards improving public health. *Air Quality, Atmosphere & Health*. 2012;5(2):217–230. ISSN 1873-9318, 1873-9326.

[23] {U.S. EPA}. Environmental benefits mapping and analysis program (BenMAP) user's manual. Technical report, US EPA. 2012.

[24] Glass, Thomas A, Goodman, Steven N, Hernn, Miguel A, and Samet, Jonathan M. Causal inference in public health. *Annual review of public health*. 2013;34:61–75. ISSN 1545-2093.

[25] Rubin, Donald B. Bayesian inference for causal effects: The role of randomization. *The Annals of Statistics*. 1978;6(1):34–58. ISSN 00905364.

[26] Rubin, Donald B. For objective causal inference, design trumps analysis. *The Annals of Applied Statistics*. 2008;2(3):808–840. ISSN 1932-6157.

[27] Hernan, Miguel A., Alonso, Alvaro, Logan, Roger, Grodstein, Francine, Michels, Karin B., Willett, Walter C., Manson, JoAnn E., and Robins, James M. Observational studies analyzed like randomized experiments. *Epidemiology*. 2008;19(6):766–779. ISSN 1044-3983.

[28] Clancy, L., Goodman, P., Sinclair, H., and Dockery, D. W. Effect of air-pollution control on death rates in dublin, ireland: an intervention study. *The Lancet*. 2002;360(9341):12101214.

[29] Holland, Paul W. Statistics and causal inference. *Journal of the American Statistical Association*. 1986; 81(396):945–960. ISSN 01621459.

[30] Dockery, Douglas W., RIch, David Q., Goodman, Patrick G., Clancy, Luke, Ohman-Strickland, Pamela, George, Prethibha, and Kotlov, Tania. *Effect of Air Pollution Control on Mortality and Hospital Admissions in Ireland*. Health Effects Institute, Boston, MA. 2013.

[31] Peel, Jennifer L., Klein, Mitchell, Flanders, W. Dana, Mulholland, James A., and Tolbert, Paige E. *Impact of Improved Air Quality During the 1996 Summer Olympic Games in Atlanta on Multiple Cardiovascular and Respiratory Outcomes*. Health Effects Institute, Boston, MA. 2010.

[32] Hudgens, Michael and Halloran, Elizabeth. Toward causal inference with interference. *Journal of the American Statistical Association*. 2008;103(482):832—842.

[33] Tchetgen, Eric J. Tchetgen and VanderWeele, Tyler J. On causal inference in the presence of interference. *Statistical Methods in Medical Research*. 2012;21(1):55–75. ISSN 0962-2802, 1477-0334.

[34] Chay, Kenneth Y. and Greenstone, Michael. The impact of air pollution on infant mortality: evidence from geographic variation in pollution shocks induced by a recession. *The quarterly journal of economics*. 2003; 118(3):11211167.

[35] Pope III, C. Arden, Rodermund, Douglas L., and Gee, Matthew M. Mortality effects of a copper smelter strike and reduced ambient sulfate particulate matter air pollution. *Environmental Health Perspectives*. 2007; 115(5):679–683. ISSN 0091-6765.

[36] Moore, Kelly, Neugebauer, Romain, Lurmann, Frederick, Hall, Jane, Brajer, Victor, Alcorn, Sianna, and Tager, Ira. Ambient ozone concentrations and cardiac mortality in southern california 19832000: Application of a new marginal structural model approach. *American Journal of Epidemiology*. 2010;171(11):1233 –1243.

[37] Currie, Janet and Walker, Reed. Traffic congestion and infant health: Evidence from e-ZPass. *American Economic Journal: Applied Economics*. 2011;3(1):65–90. ISSN 1945-7782, 1945-7790.

[38] Rich, David Q., Kipen, Howard M., Huang, Wei, Wang, Guangfa, Wang, Yuedan, Zhu, Ping, Ohman-Strickland, Pamela, Hu, Min, Philipp, Claire, and Diehl, Scott R. 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):20682078.

[39] Hedley, Anthony Johnson, Wong, Chit-Ming, Thach, Thuan Quoc, Ma, Stefan, Lam, Tai-Hing, and Anderson, Hugh Ross. Cardiorespiratory and all-cause mortality after restrictions on sulphur content of fuel in hong kong: an intervention study. *The Lancet*. 2002;360(9346):1646–1652. ISSN 0140-6736.

[40] Tonne, C, Beevers, S, Armstrong, B, Kelly, F, and Wilkinson, P. Air pollution and mortality benefits of the london congestion charge: spatial and socioeconomic inequalities. *Occupational and Environmental Medicine*. 2008;65(9):620 –627.

[41] Chay, Kenneth, Dobkin, Carlos, and Greenstone, Michael. The clean air act of 1970 and adult mortality. *Journal of Risk and Uncertainty*. 2003;27(3):279–300.

[42] Greenstone, Michael. Did the clean air act cause the remarkable decline in sulfur dioxide concentrations? *Journal of Environmental Economics and Management*. 2004;47(3):585–611. ISSN 0095-0696.

[43] Zigler, Corwin M, Dominici, Francesca, and Wang, Yun. Estimating causal effects of air quality regulations using principal stratification for spatially correlated multivariate intermediate outcomes. *Biostatistics (Oxford, England)*. 2012;13(2):289–302. ISSN 1468-4357.

[44] Deschenes, Olivier, Greenstone, Michael, and Shapiro, Joseph S. Defensive investments and the demand for air quality: Evidence from the NOx budget program and ozone reductions. Working Paper 18267, National Bureau of Economic Research, Cambridge, MA. 2012.

Table 1: Existing accountability studies classified according to the causal question of interest, 1993–2013.

Category	Study	Study Name	Description	Direct or Indirect Accountability	Causal Analysis Questions
A	Dockery, 1993 [1]	Six Cities	$PM_{2.5}, PM_{10}$, Mortality	Indirect	What is the association between
A	Laden, 2006 [5]	Six Cities Follow Up	$PM_{2.5}$. Mortality	Indirect	**pollution** exposure and health?
A	Zeger, 2008 [6]	US	$PM_{2.5}$, Mortality	Indirect	Is this a causal association?
A	Pope III, 2009 [7]	US	$PM_{2.5}$, Life Expectancy	Indirect	(Classical Paradigm)
A	Correia, 2013 [8]	US	$PM_{2.5}$, Life Expectancy	Indirect	
B	Pope III, 1996[2]	Utah Valley Steel Mill	PM_{10}, Various Health Indicators	Indirect	
B	Chay, 2003[34]	1981-1982 Recession	TSP, Infant Mortality	Indirect	What is the causal effect
B	Pope III, 2007 [35]	Copper Smelter Strike	SO_{2-4}, Mortality	Indirect	of differential exposure to
B	Moore, 2010 [36]	Southern California	O_3, Asthma	Indirect	**pollution** on health?
B	Currie, 2011 [37]	New Jersey E-Z Pass	Birth outcomes	Indirect	(Potential Outcomes Paradigm)
B	RIch, 2012[38]	Beijing Olympics	$PM_{2.5}$, cardiovascular biomarkers	Indirect	
B	Chen, 2013 [9]	Huai River Policy	TSP, Life Expectancy	Indirect	
C	Friedman, 2001 [3]	Traffic, 1996 Atlanta Olympics	O_3, Asthma	Direct	
C	Hedley, 2002 [39]	Hong Kong Sulfur Restriction	Sulfur dioxide, Mortality	Direct	What is the causal effect
C	Clancy, 2002 [28]	Dublin Coal Ban	Black Smoke, Mortality	Direct	of the **intervention**
C	Tonne, 2008[40]	London Traffic Charging	NO_2, PM_{10}, Life Expectancy	Direct	on health?
C	Chay, 2003 [41]	1970 CAA	TSP, Adult Mortality	Direct	(Potential Outcomes Paradigm)
C	Greenstone, 2004 [42]	1970 CAA	Sulfur dioxide	Direct	
C	Zigler, 2012 [43]	1990 PM_{10} Nonattainment	PM_{10}, Mortality	Direct	
C	Deschenes, 2012 [44]	NO_x Budget Program	O_3, pharmaceutical expenditures, mortality	Direct	

Particulate Matter ($PM_{10}, PM_{2.5}$); Ozone (O_3); Total Suspended Particles (TSP); Nitrogen Oxides (NO_x, NO_2); Sulfur Dioxide (SO_2), Sulfate (SO_4); Clean Air Act (CAA)

Figure 1: Locations of 407 coal-burning power plants participating in the Acid Rain Program during 1995-1997. Size of plotting symbol is proportional to the average number of sulfur dioxide tons emitted at each location during 1995-1997.

Figure 2: Monthly sulfur dioxide emissions from 1995-2012 among coal-burning power plants participating in the Acid Rain Program during 1995-1997. Thick, bold lines are for facilities at deciles of sulfur dioxide emissions during 1995-1997.

Prepress

STATEMENT

Synopsis of Research Report 187

HEALTH EFFECTS INSTITUTE

Causal Inference Methods for Estimating Long-Term Health Effects of Air Quality Regulations

INTRODUCTION

The report by Dr. Corwin Zigler and colleagues, *Causal Inference Methods for Estimating Long-Term Health Effects of Air Quality Regulations*, is the latest in a series of reports funded as part of HEI's accountability research program. Established 15 years ago, this program has aimed at evaluating whether regulatory and other actions taken to improve air quality have resulted in the intended improvements in air quality, exposure, and health outcomes.

Zigler and his colleagues tackled a number of important questions that have remained unanswered by previous air pollution accountability research. A major goal of the study was to use both established methods and newly developed methods that would enable a "direct" accountability assessment of air pollution interventions — that is, to assess from a statistical standpoint whether the intervention had *caused* changes in pollutant levels or health outcomes. This "direct" approach contrasts with the "indirect" accountability approach in which the future health benefits of an intervention are estimated from the intervention's projected impact on future exposures combined with the exposure–response relationships derived from retrospective epidemiological studies.

As part of demonstrating their methods, they applied them in two well-developed case studies of interventions designed to have long-term impacts on health, not just the shorter term interventions that have been the focus of much previous accountability research. Longer term effects of air pollution interventions on health are important because they account for the majority of the estimated benefits from improving air quality. Another

What This Study Adds

- Zigler and colleagues have provided a well-written primer on how more systematic approaches to testing of causality (i.e., through use of causal inference frameworks and methods) could be adapted to the assessment of the effects of air pollution interventions on air quality and health.
- In a major undertaking, they successfully demonstrated the use of existing and newly developed methods in two case studies of regulatory actions: the designation of counties to be in nonattainment with the National Ambient Air Quality Standards for PM_{10} and the installation of SO_2 scrubbers on power plants.
- The scrubber case study provides both newly developed methods and a rare comparison of two different but analogous statistical approaches — principal stratification and causal mediation analysis — applied to the same complex multipollutant problem.
- Their work demonstrated the critical importance of involving multidisciplinary teams with detailed technical knowledge of the interventions to ensure appropriate study design and interpretation.
- The Committee concluded that these accountability methods are an important addition to the "toolkit" and should continue to be further explored, but cannot wholly substitute for accountability assessments that rely on evidence from other scientific methods, including more traditional epidemiology analyses.

This Statement, prepared by the Health Effects Institute, summarizes a research project funded by HEI and conducted by Dr. Corwin M. Zigler of the Harvard T.H. Chan School of Public Health, Boston, Massachusetts, and colleagues. The complete report, *Causal Inference Methods for Estimating Long-Term Health Effects of Air Quality Regulations* (© 2016 Health Effects Institute), can be obtained from HEI or our Web site (see last page).

important feature of this project is the investigators' development of new methods for evaluating the impacts of interventions on multiple pollutants and the pathways via which the interventions and pollutants may affect exposure and health outcomes. To provide expertise on the complexities of the air pollution interventions chosen, Dr. Zigler added to their team Mr. John Bachmann, former Associate Director for Science/Policy and New Programs for the U.S. EPA's Office of Air and Radiation. Finally, in a commitment to transparency and data access, they plan to make publicly available the statistical code necessary both to assemble and link their data sources and to implement their newly developed methods.

APPROACH

As in other published work on causal methods, the first important feature of their approach was the reframing of air pollution interventions as a hypothetical randomized experiment, analogous to a randomized clinical trial in which some subjects are randomly assigned to receive "treatment" and others receive none, the "controls." Randomized studies are considered the optimal study design for determining the efficacy, or causal influence, of treatment because randomization typically results in balance of potential confounders between the treatment and control groups.

The next important feature of their approach was to apply and extend two different but conceptually analogous methods, principal stratification and causal mediation, to investigate the importance of alternative causal pathways for the interventions. The causal pathways are the pathways through which an intervention may act to cause changes in the outcome of interest. The pathway may represent the direct effect of one factor on an outcome (e.g., air pollution on health outcomes) or may involve the mediation by some intermediate step or factor.

Principal stratification involves comparison of outcomes between key strata or groupings of the data (for example, the effects on health in areas where an intervention has caused a reduction in air pollution and those where it has not). Using this general example, it defines "associative" effects as those effects on health that occur when an intervention caused a meaningful reduction in air pollution and "dissociative" effects as the effects on health outcomes that occur when the intervention did not have a causal effect on air pollution. The size of the associative effects relative to the dissociative effects provides an indication of the relative importance of the two pathways, in this example an indication of the intermediate role of the reduction in air pollution. Causal mediation methods are also designed to evaluate the effect of mediators or intermediate steps on an outcome of interest but in a more formal way. Using our general air pollution example, causal mediation divides the effects of an intervention into two components: (1) the "natural direct" effect, defined as the direct effect of the intervention on the outcome, and (2) the "natural indirect" effect, defined as the causal effect mediated by changes in some intermediate factor like a specific air pollutant. However, unlike in principal stratification, these two effects sum to the total effect. The authors demonstrated the use of these methods in two case studies of different regulatory interventions.

In the first case study, the authors evaluated the effect on air quality and on health outcomes of designating areas of the Western United States to be in "non-attainment" with the 1987 National Ambient Air Quality Standards for PM_{10} in the period 1990–1995. Specifically, they examined the causal effects of these designations on ambient PM_{10} concentrations in 1999–2001 and on all-cause mortality and on hospitalizations for cardiovascular and respiratory diseases in 2001. In the framing of the analysis like a randomized controlled experiment, the areas designated as in nonattainment are considered to be assigned to "treatment" whereas attainment areas served as "controls." Because these two groups were not actually selected via a randomized process, the authors developed and used propensity scores, an aggregate measure of multiple potential confounding factors, to identify groups of nonattainment (219) and attainment areas (276) that appeared comparable. The first step was to estimate the causal effects of nonattainment designation on PM_{10} concentrations and on Medicare health outcomes, which they did using regression techniques.

The investigators next used principal stratification to examine whether causal effects of non-attainment designation on health outcomes were more likely than not to have occurred via causal reduction in ambient PM_{10} concentrations. For this case study, they defined "associative" effects as the effects on health when the nonattainment designation was found to cause a reduction in ambient PM_{10} by at least 5 $\mu g/m^3$, and "dissociative" effects as the effects on health outcomes that occurred when the designation did not have a causal effect on PM_{10}.

The second case study was designed to evaluate the causal impacts on emissions and ambient $PM_{2.5}$ of installing a range of scrubber technologies on coal-fired power plants pursuant to requirements to reduce emissions of multiple pollutants (SO_2, NO_x, and CO_2)

under the Acid Rain Program, a program created by the 1990 amendments to the Clean Air Act. The effects of scrubbers on pollutant emissions and ambient $PM_{2.5}$ concentrations have been well-studied and understood, so this intervention provided a good opportunity to test whether the new methods would perform as expected.

The investigators estimated the causal effect of scrubber installation on emissions by comparing the 2005 emissions of SO_2, NO_x, and CO_2 and levels of ambient $PM_{2.5}$ observed for 63 power plants that were equipped with scrubbers ("treated") with the emissions from those 195 power plants that were not ("controls"). Zigler and colleagues then applied both principal stratification and causal mediation methods to evaluate the extent to which the causal effect of a scrubber on ambient $PM_{2.5}$ was mediated through reduced emissions of SO_2, NO_x, and CO_2. This analysis is notable because it involves assessment of the roles of multiple pollutants whereas most accountability assessments consider only one. It is also a rare application of the two methods to the same complex problem.

The principal stratification analysis compared the "associative" effects of scrubbers on $PM_{2.5}$ — the causal effects of a scrubber on ambient $PM_{2.5}$ among power plants where emissions of SO_2, NO_x, and CO_2 were causally affected by the presence of a scrubber — with the "dissociative" effects — the causal effects of a scrubber on ambient $PM_{2.5}$ among power plants where the emissions were not meaningfully affected.

Zigler and colleagues developed new Bayesian nonparametric methods to conduct their multipollutant causal mediation analysis. In this case study, they defined the "natural direct" effect as the effect that the presence of scrubbers had on $PM_{2.5}$ and the "natural indirect" effects as the causal effects on $PM_{2.5}$ mediated by changes in the emissions of the three pollutants, either individually or in various combinations with each other.

RESULTS AND INTERPRETATION

In their evaluation of the effect of nonattainment designation, the authors concluded that there was some evidence that the intervention caused a small reduction, on average, in ambient PM_{10} levels, in all-cause mortality, and in hospitalizations for respiratory disease among Medicare beneficiaries. They did not find a reduction in hospitalizations for cardiovascular disease.

With their principal stratification analysis, Zigler and colleagues found differing results for the intermediary role of PM_{10} in causal effects on the three health outcomes. Contrary to expectations, their analysis suggested a reduction, on average, in mortality even in areas where their analyses reported that PM_{10} was not causally affected. The authors suggested that the observed causal effect of nonattainment designation on mortality, in the absence of a strong associative effect for PM_{10}, may be due to causal pathways other than the one involving reduction of PM_{10}. However, they suggested their results provide evidence that PM_{10} played a causal role in the reduction of hospitalization for respiratory disease, but again, not for cardiovascular disease.

As the authors noted, all of the estimates from these analyses were accompanied by substantial uncertainty, indicated by broad posterior 95% confidence intervals that included zero. As a result, the HEI Health Review Committee thought the investigators generally overstated the average causal effects of nonattainment designation and the role of PM_{10} in this study. The Committee agreed that a major contributor to the uncertainty in the results was the ambiguity of the intervention; that is, that nonattainment designation is not a discrete intervention, but is subject to a number of sources of heterogeneity in the actions implemented over space and time.

In their second case study, Zigler and colleagues found results that were consistent with what is known about scrubbers. They estimated that installation of scrubbers had, on average, caused reductions in SO_2, but not in NO_x and CO_2 emissions, and had also caused modest reductions in ambient $PM_{2.5}$ concentrations. Their multipollutant causal pathways analyses using principal stratification and causal mediation methods yielded broadly similar results. That is, both led the authors to conclude that the observed causal reductions in ambient $PM_{2.5}$ among power plants equipped with scrubbers were effected principally through the causal reduction of SO_2 emissions rather than through reductions in emissions of NO_x and CO_2. Their causal mediation analysis provided a somewhat clearer support for that conclusion because the reduction in $PM_{2.5}$ mediated by SO_2 (the natural indirect effect) was statistically significant and larger than those mediated either by NO_x and CO_2, which were all close to zero. The 95% posterior intervals for all the results in the principal stratification analysis were quite broad and included zero.

Although the scrubber case study was conceptually clearer for demonstrating the methods, the authors had made a number of simplifying assumptions that could have contributed to uncertainties in the results, a question that could be explored more fully in future analyses. The investigators' first iteration of the

analysis yielded results that ran counter to established knowledge (i.e., the results suggested SO_2 scrubbers' effects on ambient $PM_{2.5}$ were not causally mediated by changes in SO_2 emissions) that led them to identify and correct for additional important characteristics in their final analysis. It is still difficult to know if there were other regulation-related activities undertaken that blurred the distinctions between treated and untreated facilities and that could explain the high degree of uncertainty observed in the results.

CONCLUSIONS

The Committee concluded that Zigler and his colleagues provided a well-conducted study and a well-written report that makes a major contribution to the field of accountability research in the context of air pollution and health. The statistical framework described in this report provides a particularly clear and explicit approach to thinking about the health impacts of all kinds of interventions designed to reduce emissions and ambient air pollution. Although most of the causal inference methods Zigler and colleagues used were not new, their extensions to two substantive air pollution interventions and to multiple pollutants were a major undertaking in and of themselves. The advances they made in applying the methods in real applications have moved us further than other methodological studies and provided a clearer path toward further development and deployment of the methods in other settings.

What the considerable methodological work in this study indicates, however, is that the presence of a clear causal framework is not a substitute for detailed consideration of potentially important covariates and the testing of the sensitivity of results to key assumptions made in implementing the methods. Both these case studies demonstrated the critical importance of involving multidisciplinary teams with detailed technical knowledge of the interventions being studied. Even so, it is difficult to be sure to what extent the uncertainty in the causal effects estimated is attributable to weakness in the causal relationship or to the imprecision in the problem definition and underlying data. Finally, not all questions can necessarily be addressed in a causal framework, for example, situations in which suitable "controls" do not exist or in which analysts need to predict the potential impacts of some future intervention. The Committee concluded that these and other "direct" accountability methods are an important addition to the "toolkit" and should continue to be further explored, but cannot wholly substitute for "indirect" accountability methods.

HEALTH EFFECTS INSTITUTE

75 Federal Street, Suite 1400
Boston, MA 02110, USA
+1-617-488-2300 phone
+1-617-488-2335 fax

pubs@healtheffects.org
www.healtheffects.org

Recycled Paper