Pollution and Health

Joshua Graff Zivin, *University of California - San Diego*
Matthew Neidell, *Columbia University*

Available at: https://works.bepress.com/josh_graffzivin/62/
Introduction

A primary objective of environmental policies worldwide is to protect human health. Optimal policy design, however, is typically hampered by limited information regarding both the benefits and the costs associated with regulation. Benefits assessments frequently rely on translating laboratory findings to uncontrolled settings, extrapolating from high- to low-concentration exposures within and across societies, and drawing inferences from observational analyses that do not account for the endogeneity of pollution. Economic assessments have typically focused on the costs of compliance to firms. Efforts to improve societal welfare clearly depend on a strong understanding of both elements. Although the health-pollution relationship largely remains the pursuit of epidemiologists, the focus of economics on casual identification along with valuation techniques consistent with utility maximization has helped to reframe these relationships in a manner that facilitates policy choice and environmental rule setting.

Early epidemiological investigations of the impacts of extreme pollution events were some of the first compelling studies to suggest a causal relationship, with one of the most famous focused on the ‘killer fog’ in London, England in December, 1952. A temperature inversion combined with windless conditions led to a sudden and dramatic increase in air pollution. Because residents were used to winter fogs, there was little, if any, changes in behavior, leading to a rather clean measure of pollution impacts in this case. The dramatic rise in mortality that precisely coincided with the timing of the fog had been a driving force behind the federal regulations aimed at air pollution control.

The pollution levels experienced under this and similarly studied extreme events, however, had been dramatically higher than those that nearly all people in developed countries face today. Moreover, most exposures do not conveniently arrive as a ‘surprise’ under which causal impacts can be easily assessed, and it is on this front that economists have made their most significant contributions. In particular, economic studies have typically focused on quasi-experimental settings in order to synthesize the ‘surprise’ of pollution. Besides improving the causal understanding of these relationships by minimizing threats from confounding, it has also identified important compensatory behaviors undertaken by individuals to mitigate exposure.

These behavioral responses are often nontrivial because many pollutants are observable, and even those that are not easily detectable by the public, such as ozone and particulate matter, are forecast and publicized through a broad range of media outlets. If optimizing individuals compensate for changes in ambient pollution levels by reducing their exposure, estimates that do not account for these responses will underestimate the biologic relationship between ambient pollution levels and health. This problem is potentially severe because the more an individual is likely to suffer under pollution, the more they have to gain from reduced exposure. Indeed, emerging empirical evidence finds that behavioral responses are largest for more vulnerable individuals (Neidell, 2009; Graff Zivin and Neidell, 2009; Graff Zivin et al., 2011) and that individuals are more responsive to higher levels of pollution (Neidell, 2009; Mansfield et al., 2006). Equally important, these behavioral responses are costly and thus ignoring them will also understate the welfare effects of pollution (or the regulation thereof). Although the costs of spending additional time indoors, rescheduling activities, or even relocating to areas with better environmental quality are often difficult to enumerate, they can represent a substantial fraction of the total costs of pollution.

In the remainder of this article, a basic economic framework for evaluating environmental health impacts is presented, followed by a discussion of the core empirical challenges that researchers face in estimating the relationship between pollution and health. A selective review of significant contributions from the literature that focus on the effects of air pollution is then provided, concluding with some suggestions for fruitful lines of future research.

Conceptual Framework

Estimation of the relationship between pollution and health is typically focused on the following health production function:

$$h = f(P, A, E, S)$$  \[1\]

where $h$ is a measure of an individual’s health, $P$ is pollution levels assigned to the individual, and $A$ is avoidance behavior. $E$ are other environmental factors that directly affect health, such as weather and allergens, and $S$ are all other behavioral, socioeconomic, and genetic factors affecting health. Given that
meteorological elements can play an important role in pollution formation and can also affect health (e.g., cold weather increases asthma exacerbation). E is defined separately because it represents an important source of environmental confounding.

Two main approaches are taken to eqn [1], with the difference stemming from the treatment of avoidance behavior. The first, or ‘reduced-form’ approach does not directly control for avoidance behavior. As health impacts will depend on ambient pollution levels and avoidance behavior that determines exposure to those pollution levels, the health relationship can be expressed as the following total derivative: \[ \frac{dh}{dP} = \frac{dh}{dA} + \frac{dh}{dA} \cdot \frac{dA}{dP}. \] The second, or ‘production function’ approach directly controls for avoidance behavior to obtain the partial derivative: \[ \frac{dh}{dP}. \]

The importance in separating these two approaches is to relate each to the benefit calculation, or willingness to pay (WTP) for a reduction in pollution (Harrington and Portney, 1987; Cropper and Freeman, 1991; Deschene and Greenstone, 2011). In the reduced-form approach, welfare is typically expressed as: \[ WTP = \frac{dh}{dP} \cdot C_0 + P \cdot \frac{dA}{dP}, \] where \( C_0 \) is the ‘full’ cost associated with a change in health, and \( P \) is the price of avoidance behavior. In the production function approach, welfare is typically expressed as: \[ WTP = P \cdot \frac{[\frac{dh}{dP}]}{\frac{dA}{dP}}. \] Although the production function approach appears more data hungry because of the need to control for avoidance behavior when estimating eqn [1], the reduced-form approach must also control for avoidance behavior in order to estimate \( \frac{dA}{dP} \), although this can be done separately from estimating eqn [1]. Furthermore, as these expressions demonstrate, all forms of avoidance behavior must be accounted for at some point in order to obtain a proper estimate of WTP.

One advantage of the reduced-form approach is that the econometrician does not need to properly specify the functional form of eqn [1] with respect to P and A. This is particularly helpful because data limitations often necessitate the use of proxy measures for avoidance behavior, and economic theory provides little guidance on how these proxy measures should enter into eqn [1].

The value of the production function approach is that it provides estimates of the biological effect of pollution. Because avoidance behavior is likely to vary across socioeconomic and cultural environments, but the biology is considerably less context specific, it facilitates generalizations across settings. Moreover, focusing on the biological effect enables one to potentially identify important nonlinear effects, such as threshold effects, and heterogeneous effects based on individual susceptibility, both of which can play an important role in defining the feasible set of policy interventions. Interested readers should consult Graff Zivin and Neidell (2013) for more elaboration on this framework.

**Empirical Challenges**

In this section, three primary challenges confronted by empiricists when estimating the relationship between pollution and health are outlined. Although weather is a potential confounder, this is not discussed at length because it is directly observable (often at a finer scale than pollution data), so that any threat can be obviated through the careful control of relevant variables.

**Measurement of Health**

The measurement of health outcomes and how to place monetary values on them is a persistent challenge. A frequently used measure is mortality, which is objectively measured and can be readily monetized using estimates of the value of a statistical life (VSL). One concern with using mortality is that it is an extreme outcome that misses more subtle outcomes that may be more commonplace. Furthermore, using VSL to monetize these impacts may be misleading if the loss only represents short-term mortality displacement, commonly referred to as ‘harvesting.’

Measures of morbidity have also been examined using data on hospitalizations for various conditions, largely respiratory related. Although hospitalizations clearly capture events less severe than death, they may introduce sample selection. Those who have a relationship with a primary care physician (PCP) and receive regular care may never experience a hospitalization, and access to a PCP is clearly endogenous. Furthermore, the economic valuation of hospitalizations is particularly difficult as hospital charges (which are all that is typically available) do not capture the costs associated with the pain and suffering experienced by sickened individuals or their family members.

Birth outcomes are another metric that has some of the desirable properties of both mortality and morbidity endpoints, albeit for a select population. Like mortality and hospitalizations, birth outcomes are a census and not a sample, hence offering large sample sizes for analysis. Unlike mortality, birth outcomes can capture more subtle impacts, and unlike hospitalizations, they do not introduce sample selection because any birth that files for a birth certificate is reported. Valuation approaches can be used when the birth outcome studied has been linked to monetizable events – for example, birth weight has been linked with education and earnings (Black et al., 2007) – although these links may not capture all relevant costs.

An emerging area of focus is on indirect ‘health’ outcomes at school or the workplace, principally absenteeism and performance. Such outcomes offer terrific promise for capturing rather subtle health impacts that might be broadly disseminated throughout society. They are also generally straightforward to monetize, particularly for performance. Limited data availability, especially for representative samples, is a formidable obstacle to the conduct of credible empirical work in this area.

**Assignment of Local Pollution Levels**

Most studies focus on air pollution because of the availability of data from ambient air pollution monitors, which typically measure air concentrations at an hourly scale at a fixed location. Although this frequency of measurement generates data at a fine temporal scale, the limited number of monitor locations relative to the size of a country and the geographic distribution of the population leads to data that are rather
coarse on a spatial scale. As a result, studies often approximate contemporaneous pollution levels based on an individual's general location and the location of the monitor. This crude approach leads to measurement error that increases with an individual's distance from the monitor and the degree to which pollutants disperse nonuniformly. This measurement error will typically bias estimates downward, but with a large enough dataset, researchers can use data from multiple monitors and various weighting techniques to obtain more precise assignments of localized pollution levels. A finer level of geographic disaggregation for individuals, such as a residential address, also allows for better assignment of relevant pollution levels and hence is more likely to provide precise estimates.

The usual mobility of individuals in their everyday life (not in response to pollution, discussed below), both within a day and over time, can also present a measurement issue. Individuals spend their time not only at home, but at work, school, and other possible locations that are not typically recorded. Although the use of personal monitors is designed to overcome this, two issues remain: (1) the high costs of personal monitoring often result in the use of a small, unrepresentative sample without a clearly defined control group; and (2) the link to policy is less clear because indoor sources also contribute to pollution but are subject to different regulatory rules. Mobility over time also presents a significant measurement issue in assigning cumulative exposure over longer periods of time. Focusing on children, and in particular infants, whose parents are typically less mobile, can greatly limit this concern (Joyce et al. 1989; Chay and Greenstone, 2003).

**Behavioral Responses to Pollution**

Optimizing individuals may respond to pollution with permanent changes, such as relocating (i.e., sorting), and temporary changes, such as spending less time outside. As argued above, it is crucial to understand the role of these behavioral responses both to allow generalizations from one setting to another and to account for the full welfare costs of pollution. Although careful quasi-experimental designs can address permanent behavioral changes by exploiting exogenous shocks to pollution levels, short-run changes pose greater challenges because many of these responses involve nonmarket behaviors that are difficult to observe. For example, simply spending less time outside on a polluted day is a highly effective means for reducing exposure, but such an activity is rarely recorded. Clearly, the degree to which such short-run behavioral responses will be important depends on the 'visibility' of pollution, either literally, through information dissemination, or through health feedbacks that allow individuals to infer it on the basis of physiological responses.

**Evidence**

Rather than provide an exhaustive review of the economic literature examining the relationship between pollution and health, this section limits its attention to a selection of studies that offer key insights or introduce important methodological advances.

**Primary Impacts**

One of the earliest examples of a quasi-experimental approach to estimate an environmental health relationship in relatively recent times is found in a series of studies by Pope et al. (1992); Ransom and Pope (1992); Ransom and Pope (1995). The authors used changes in pollution that had resulted from the opening and closing of a steel mill, which was a major source of particulate matter, in the central Valley of Utah due to a labor strike. As the steel mill had closed due to a labor strike, the temporary changes in pollution were credibly exogenous and unlikely to lead to any immediate residential sorting. Furthermore, the authors selected a neighboring, unaffected community as a control group to account for time trends by estimating difference-in-differences models. When the steel mill was closed, the authors found significant declines in school absences, respiratory-related hospital admissions, and mortality. One potential concern with this study is that the steel mill closure has also led to a temporary change in income, which may affect one's use of time and services. This does not seem likely to be an issue for school absences, hence at least some of the findings are credibly causal. A more significant concern with the design is that, as an 'event study,' the pollution variable is common to all members in a group for a given time period (despite the availability of individual level health outcomes as dependent variables). As a result, their standard errors are likely to be nontrivially understated, making the appropriate statistical inference in this setting particularly challenging (Donald and Lang, 2007).

One important study by Chay and Greenstone (2003) overcame this problem by focusing on the recession of the early 1980s. The dramatic change in manufacturing that had resulted from this recession induced considerable spatial variation in total suspended particulates (TSPs) throughout the US in a short period of time, with some areas experiencing as large as a 35% decline in 3 years. These changes in TSPs are unlikely to be related to other factors affecting health. Importantly, although income changed considerably at the same time, it did not show comparable spatial patterns as with TSP. Using this exogenous variation in levels of pollution at the county-year level to identify environmental health effects, they estimate that a one-unit decline in TSPs associated with the recession yields benefits of roughly US$14 billion, recognizing that this captures only one health outcome and only for a specific group.

Although the Chay and Greenstone results are nontrivial, the continued improvements in air quality since then suggest that the results also apply to a time period when pollution levels in the US are considerably higher. Currie and Neidell (2005) turn their attention to infant mortality in California during the 1990s, a period that is much more reflective of contemporary pollution levels across much of the developed world. They use zip code fixed effects to account for residential sorting, thereby exploiting the strong temporal variations in pollution levels in the short-run due to changes in plausibly exogenous ambient conditions (rather than anthropogenic sources) to identify health impacts. They find that reductions in carbon monoxide over the 1990s saved approximately 1000 infant lives in California, which translates into benefits of roughly US$4.8 billion.
Currie et al. (2009), like Currie and Neidell (2005), focus on infant outcomes in a more recent time period, but use the exact address of the mother to improve pollution assignment and estimate sibling fixed effect models to control for differences in family background and genetics. They find that a one-unit change in mean carbon monoxide (CO) during the last trimester of pregnancy increases the risk of low birth weight by 8%, and a one-unit change in mean CO during the first two weeks after birth also increases the risk of infant mortality by 2.5% relative to baseline levels. The authors calculate that the 15-year decline in CO from 1989–2003 translates into US$720 million in lifetime earnings from improvements in birth weight and US$2.2 billion from the reduction in infant mortality for the 2003 birth cohort. The use of sibling fixed effects increases estimates, suggesting the importance of accounting for maternal characteristics within neighborhoods. And the better assignment of pollution by using the mother’s exact address rather than zip code also increases point estimates, consistent with measurement error inducing a downward bias.

In a novel design, Lleras-Muney (2010) uses the relocation of military personnel to estimate the effect of various pollutants on children’s health. The relocation of personnel is entirely based on ‘the needs of the army’, which explicitly rules out the possibility of sorting and offers a plausibly exogenous source of variation in pollution. Using this design, Lleras-Muney finds that a one standard deviation decrease in ground-level ozone exposure decreases the probability of a respiratory hospitalization for children by 8–23%. Her estimates suggest that lowering pollution levels nationwide to the levels experienced in ‘low’ pollution areas would save approximately US$928 million (US$1994) in direct medical expenditures alone.

All of the previously mentioned studies exploit ‘natural’ experiments that generate exogenous changes in ambient pollution in order to minimize concerns regarding residential sorting and other long-run behavioral responses to poor environmental quality. They generally ignore potential short-run adjustments that could also impact the environment-health relationship, and hence provide estimates of a reduced-form relationship between pollution and health. The key challenge in capturing these short-run behavioral responses is clearly the availability of data suited for the task, and researchers often follow creative paths for obtaining such data. One example is Neidell (2009), who uses attendance data from several outdoor facilities in Los Angeles to uncover significant behavioral responses to high ozone levels that are forecasted through smog alerts. As smog alerts are issued only when ozone is forecasted to exceed a particular threshold, he employs a regression discontinuity design to compare attendance on days just above the threshold to that just below. Although this paper does not provide estimates of the costs of avoidance behavior, in a closely related paper Graff Zivin and Neidell (2009) examine successive days of smog alerts to show that the costs of avoidance behavior, due to limited opportunities for intertemporal substitution, are increasing over time. Graff Zivin et al. (2011) identify substantial increases in the purchase of bottled water when local municipalities violate drinking water standards. As this type of avoidance behavioral is market-based, the authors have calculated the costs associated with it, and have found that water quality violations in 2005 induced roughly US$60 million worth of bottled water purchases nationwide.

Two notable studies attempt to produce estimates of the biological effect of ozone on health. In the paper discussed earlier, Neidell (2009) controls for smog alerts and ozone forecasts as a proxy for avoidance behavior when estimating the relationship between ozone and respiratory-related hospitalizations. Using zip code fixed effects and exploiting the strong daily temporal variation in ozone, he finds that including these proxies significantly increases the estimated impact of ozone on health. Moretti and Neidell (2011) use daily boat arrivals and departures into the port of Los Angeles as an instrumental variable (IV) for ozone levels, which deals with both avoidance behavior and measurement error in pollution assignment. Boat traffic represents a major source of pollution for the Los Angeles region and, because of the extended length of travel and unpredictable conditions at sea, daily variation in boat traffic is arguably uncorrelated with other short-run determinants of health and is not included in the ozone forecasts used to encourage avoidance behavior. Similar to Neidell (2009), they find that using boat traffic as an IV leads to significantly larger estimates for the impacts of pollution on health.

Although the short-run behavior literature has generally assessed the costs associated with avoiding exposure, this again represents a partial characterization of social welfare; a complete calculation requires an assessment of both avoidance costs as well as the costs of those adverse health effects that are not avoided. To our knowledge, the only attempt to bring both pieces together in a quasi-experimental setting is from Deschenes and Greenstone (2011), who focus on the health effects of extreme temperatures, which are forecast to increase under climate change. They construct a WTP estimate to avoid extreme heat that includes the costs due to excess mortality as well as expenditure on energy consumption as a proxy for air conditioning usage to buffer individuals from exposure to that heat. Using county fixed effects to exploit the plausibly exogenous variation in temperatures in an area within a given year, they find that the avoidance costs are roughly 25% of the mortality costs.

**Secondary Impacts**

Although most of the literature has focused on primary health endpoints, for example, mortality and hospitalizations, an emerging literature has begun to examine the manifestation of less visible health assaults on nonhealth outcomes. Although these impacts are referred to as secondary, it remains possible for them to exceed the costs of primary impacts depending on their prevalence. Almond et al. (2009) examine the impact from prenatal exposure to radioactive fallout from the 1986 Chernobyl accident on both birth and schooling outcomes for children in Sweden. Although Sweden is more than 500 miles away from Chernobyl, weather conditions forced some of the plume over Sweden, and local variation in rainfall levels led to stark geographic variation in the levels of fallout throughout the country. Their study reveals that radiation exposure exhibits latent effects that affect human capital development...
Pollution affects a wide range of health outcomes, and these effects are nontrivial even at current emissions levels in the developed world. The optimal level of these pollutants is highly contested. For example, a proposed ozone standard issued by the EPA in 1997 was finally upheld by the Supreme Court in 2002, but only after endless appeals and lengthy lawsuits initiated by states and industry (Bergman, 2004). Better estimates of the relationship between pollution and health and society's WTP for improvements in pollution through the use of quasi-experimental research designs offers an important tool for informing this debate. Additional work on the measurement of avoidance behavior and its costs remains a critical piece of the puzzle.

Despite the growth of quality evidence on this topic, one area in need of more evidence is on the long-run effects from cumulative exposure to various pollutants. Although it is clear that pollution has short-run impacts, the potential impacts from exposure over a lifetime may be considerably larger, as hinted at by the results from Almond et al. (2009). These impacts may also affect people's investment decisions throughout their life course, suggesting a wide range of potential economic outcomes that may be affected. The empirical issues are more daunting given the challenges in appropriately measuring health outcomes and pollution exposure, and the ability to isolate exogenous variation in pollution, but nonetheless deserve more attention.

The impact of pollution on human capital formation and its deployment in school as well as labor markets also represents a particularly fruitful area for additional exploration. The use of these indirect outcomes can capture a broader range of economic impacts, and they also have the ability to capture subtle, but likely more pervasive health impacts than those captured through standard measures of mortality and hospitalizations. As the improvement in biomedical understanding of the etiology of disease continues, this area of study is likely to rise in frequency and importance.

References