Invited Commentary

Invited Commentary: Epidemiologic Studies of the Health Associations of Environmental Exposures With Preterm Birth

Scott L. Zeger*

* Correspondence to Dr. Scott L. Zeger, Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, 615 North Wolfe Street, Baltimore, MD 21205 (e-mail: sz@jhu.edu).

Initially submitted July 12, 2011; accepted for publication September 22, 2011.

In this issue of the Journal, two different articles present epidemiologic evidence supporting the hypotheses that environmental exposures to particulate air pollution or higher temperatures modestly increase the risk of preterm birth. In this commentary, the author discusses environmental epidemiologic methods through the lens of these two papers with respect to the causal question, measurements, and quantification and interpretation of the evidence. Both groups of investigators present results from exploratory analyses that are at the hypothesis-generating end of the research spectrum as opposed to the confirmatory end. The present author describes in qualitative terms a method for decomposing evidence about the association of environmental exposures with prematurity into components representing different temporal and spatial scales. Finally, reproducible epidemiologic research methodology for studies like these is offered as one way to speed the transition from exploratory studies to confirmatory studies.

In this issue of the Journal, the authors of two articles present epidemiologic evidence supporting the hypotheses that environmental exposures to particulate air pollution or higher temperatures modestly increase the risk of preterm birth or stillbirth (1, 2). In the North Carolina study (1), the average level of particulate matter less than 2.5 μm in diameter (PM₂.₅) during different periods of gestation was the exposure of primary interest. The investigators crudely controlled for temperature by including seasonal indicator variables in a discrete-time survival regression model. In the Brisbane, Australia study (2), temperature and humidity were the primary exposures, with the investigators controlling for weekly mean levels of 5 air pollutants in the model. These two papers provide an opportunity to comment on the methods used in environmental epidemiology and to offer possibilities for sharpening inferences about environmental risks to human health.

The scientific method has been described from many perspectives (3–5). Most would agree that science is improved by these 4 components, which I will discuss in turn:

1. A clear definition of the question being asked or, equivalently, the competing hypotheses being evaluated.
2. Accurate and precise measurements relevant to the questions posed.
3. Objective quantification of the evidence generated in the study.
4. Careful interpretation of the evidence as it reflects upon the competing hypotheses.

SCIENTIFIC QUESTION

Despite more than a quarter-century of active discussion in the epidemiologic literature about causal analysis (6–9), the authors of both papers are vague about the question being addressed. In the North Carolina study, the investigators “examined the association,” “estimated the risks,” and concluded, “These findings suggest that exposure to ambient PM₂.₅ during pregnancy is associated with increased risk of

Abbreviation: PM₂.₅, particulate matter less than 2.5 μm in diameter.
preterm birth” (1, p. 000). In the Brisbane study, the investigators found that “higher ambient temperatures in the last 4 weeks of the pregnancy increased the risk of stillbirth” (2, p. 000). Based upon the evidence presented in these papers, should readers change their thinking about whether there is a mechanism by which particulate matter (or higher temperature) causes premature birth or how much the incidence of prematurity would decrease if particle concentrations were lowered? The causal question is discussed in these papers but not addressed head-on.

Each of these papers might benefit from the use of a causal diagram (8) as a road map to 1) identify the mechanisms that are potential causes of prematurity; 2) understand the specific hypotheses being tested; and 3) visualize how the study measurements relate to the underlying processes. For example, the viability of a fetus depends not only on the gestational period but also on the rate of uterine growth, a byproduct of maternal nutrition, infections, and other environmental exposures (10). What are the consequences to causal inference of using an imperfect surrogate for gestational period (see below) in the analyses presented in these papers?

MEASUREMENT

Valid scientific inference is commonly limited by a lack of availability and quality of measurements of the underlying processes. The most important and challenging measurement in these two studies was gestational age, the time variable for the analyses. Because both studies used birth registry data, gestational age was measured by clinical assessment at birth, based substantially on the estimated time of the last menstrual cycle. Measurement error in this key variable has the potential to be differential if the environmental exposures affect both the biologic processes leading to prematurity and the measurement of the degree of prematurity.

The second most important measurement was measurement of the exposure, and here both studies shine. The North Carolina study is distinguished by its use of measured and modeled values for particulate matter (1). The paper is a model for how to handle the inferential tradeoff of additional birth-exposure information that results from imputing particulate matter values against the bias caused by using model-based estimates rather than actual particulate matter measurements. Another strength of both studies is that they had sufficient exposure-data time to distinguish between acute and chronic exposures.

Finally, lack of control for potential confounders is an inherent limitation in this field. The North Carolina investigators used 3 seasonal dummy variables as surrogates for weather (1), but the Brisbane investigators showed that there may be nonnegligible weather effects that cannot be controlled for by 3 or even 12 season indicator variables (2). The Brisbane weather study had rich pollution data; however, because of high collinearity, the pollutants were entered into the model one at a time (2). It is unclear how sensitive the main weather findings were to model-based adjustment for air quality.

QUANTIFICATION OF THE EVIDENCE

Epidemiology is constrained by a lack of measurement quality for many key variables, a problem that requires longer-term solutions through research and a greater focus in graduate education on improved measurement. However, given current measurements, there are opportunities to improve how available information is used. In studies with space-time data, decomposition of the evidence into components representing different temporal and spatial scales, each with its own risk of confounding, can be a useful strategy for increasing confidence in causal inference.

In the North Carolina and Brisbane studies, the particulate matter and temperature data varied at multiple spatial scales (within communities, across communities) and time scales (within days, within weeks, within seasons, and across years). As a result, estimates of exposure relative risks from the statistical models were combinations of measures of association from these multiple scales. Each analytic model assigns different weights to the evidence at the different scales, some of which are more likely to be confounded than others.

It is possible to decompose model-based regression coefficients into their spatial and temporal scale components. When all of the component coefficients point in the same direction, the different parts of the evidence base are internally consistent. When they point in different directions, the pooled value from the model is not to be trusted without further exploration of potential explanations for the differences. Application of this decomposition to the relation between air pollution and mortality is illustrated, for example, by Janes et al. (11) and Greven et al. (12).

Short of decomposing epidemiologic associations into component parts, it is important for investigators to make clear which components of their data most strongly influence their findings. In the North Carolina model, a 2-stage analysis was used (1). In stage 1, a separate discrete-time hazard model was fitted to the data from each of the 25 counties (80 with the fused database); the county-specific log relative risk estimates were then combined in stage 2. With this approach, all of the spatial variation among counties is set aside and only within-county information is used. Expressed another way, if we imagined comparing the birth outcomes for all pairs of infants with different exposures, only pairs in which both infants were from the same county would be included. Two interesting questions are: 1) what fraction of the spatial variation in particulate matter is within counties versus among counties; and 2) what is the evidence for an association if the among-county information is also included?

In the time dimension, the particle data are averaged over different time scales, from week to trimester to the entire pregnancy. The information within a given scale is largely set aside when the exposure is an average over that scale. The model includes 3 seasonal dummy variables, which sets aside a small part of the seasonal variation, but information about the relative risk from year-to-year variation in the seasonal wave and from longer-term secular trends remains. Thus, if the rate of premature birth is decreasing because of better prenatal care across the state and if the level of particulate matter is slowly decreasing over the 5-year period due
to loss of economic activity, this spurious association will be included in the estimated relative risk.

In the North Carolina study, the estimated relative risks were derived entirely from comparing persons with higher average particulate matter exposures with those with lower average exposures over periods of weeks to months within counties (1). Because the level of particulate matter averaged over weeks or longer periods is likely to be smooth within counties, one might expect that most of the information about the relative risk comes from comparisons of individuals at different points in the 5-year follow-up period. Confounding from longer-term trends is therefore a reasonable concern.

**INTERPRETATION OF THE EVIDENCE**

That we seek answers to causal questions is apparent in both papers, despite the observational designs that lead the authors to report only on “associations.” How can epidemiologic studies like these move closer to achieving causal conclusions?

With this goal, it is useful to distinguish hypothesis-generating research from hypothesis-testing research. The statistician John Tukey distinguished what he called “exploratory” analyses from “confirmatory” analyses (13). A consistent association between an exposure and an outcome in exploratory analyses, like those reported in the two papers in this issue, ought to motivate the formation of hypotheses about the competing mechanisms that could give rise to the observed patterns. The postulated mechanisms lead to hypotheses to be tested in confirmatory analyses in subsequent experiments performed with the rigor that has become standard in multicenter clinical trials, whether or not the follow-up study is observational or experimental.

The transition from exploratory research to confirmatory research is an iterative process in which studies first generate ideas that are then refined over time. A challenge for epidemiology is the long cycle time inherent in population research. One way to accelerate the process is to require that epidemiologic papers be “reproducible” so that data and analytical methods with which to reproduce and critique the reported findings are publicly available (14). Following the lead of the genomics research community, reproducible research is gaining currency in many biomedical fields (15, 16). If the data and software from the papers published in this issue were easily accessed, even if only for the purpose of writing letters to the editors of the Journal, a number of the questions raised here could be addressed rapidly. New hypotheses might be generated at a faster pace than if new data sets had to be independently developed.

In summary, the authors of these papers are to be congratulated on their exploratory analyses, which have identified interesting associations that raise hypotheses worth pursuing about the possible effects of ambient temperature, humidity, and air quality on prematurity. A natural next step is to refine the hypotheses and measurements so that data analyses move closer to the confirmatory end of the spectrum, in an iterative process that is accelerated by reproducible research methods.

**ACKNOWLEDGMENTS**

Author affiliation: Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland.

Conflict of interest: none declared.

**REFERENCES**