HEALTH EFFECTS INSTITUTE

STATEMENT

Synopsis of Research Report 175

A Semiparametric Regression Approach for Air Pollution Research

BACKGROUND

The findings of a number of epidemiologic studies of air pollution and health have played a central role in setting air quality limits aimed at protecting public health. Since the mid-1990s, HEI has sponsored original research in this area, as well as research and review activities focused on the analytic methods used in such studies. These efforts include — among many others — the National Morbidity, Mortality, and Air Pollution Study (NMMAPS), the Reanalysis of the Harvard Six Cities Study and American Cancer Society Study of Particulate Air Pollution and Mortality, and the HEI Special Report on Revised Analyses of Time-Series Studies.

Time-series studies are commonly used to evaluate relationships between variations in short-term pollutant concentrations and acute human disease outcomes or mortality. Because time-series methods compare counts of disease events or deaths with pollutant concentrations on a specific day or other short time frame, the analyses do not need to account for subjects' smoking behavior or other risk factors that do not change from day to day. However, when evaluating the relationship between health outcomes and pollutant exposures, investigators do need to systematically adjust the data sets to control for time-dependent phenomena such as weather and seasonal trends that may influence observed disease patterns.

In 2003, HEI produced a Special Report on the Revised Analyses of Time-Series Studies of Air Pollution and Health after scientists discovered a problem with the commonly used S-Plus statistical software. The Special Report contained a number of wide-ranging recommendations for future time-series analyses, and specifically emphasized that the effect estimates derived from time-series data were shown to be sensitive to the statistical methods and parameters used to control for long-term time trends in the data.

Following publication of the Special Report, Dr. James Robins of the Harvard School of Public Health and his colleagues submitted a preliminary application to develop and apply statistical methods to address some of the issues raised by the report. They proposed to (1) develop methods that would improve the point estimates and confidence intervals for the parameters of a semiparametric regression model, (2) compare the new methods with standard methods in simulated studies, (3) develop efficient user-friendly software to implement the new methods, and (4) reanalyze critical data sets and compare the results with those from studies based on other methods.

What This Study Adds

- Robins and colleagues successfully developed semiparametric methods for epidemiologic investigations that are likely to produce risk estimates that are less biased than traditional Poisson time-series methods.
- When applied to the NMMAPS data set, the semiparametric methods produced estimates of the risk of health events relative to pollutant levels that were of similar magnitude to those obtained in HEI's Revised Analyses of Time-Series Studies, but with wider confidence intervals.
- Although the semiparametric methods are promising for future short-term studies of health events and air pollution levels, their utility and applicability could be enhanced by incorporating existing scientific understanding to control for confounding when relationships between covariates and mortality are well understood.

This Statement, prepared by the Health Effects Institute, summarizes a research project funded by HEI and conducted by Dr. James Robins at the Harvard School of Public Health, Boston, MA, and colleagues. The complete report, *New Statistical Approaches to Semiparametric Regression with Application to Air Pollution Research* (© 2013 Health Effects Institute), can be obtained from HEI or our Web site (see next page).

EVALUATION

Reviews of the Investigators' Report, from committee members and selected peers, were divergent on the overall importance and utility of this work for epidemiologic analyses. The Committee commented that the research team had performed high-quality work to develop statistical methods that are complex and represent a very significant effort on their part, and that the results are technically sound. They noted that the concurrence between the current investigators' results and HEI's results from the Revised Analyses of Time-Series Studies was reassuring.

The current study, although acknowledged to be highly innovative by the broader scientific community, can be most easily understood by experts who are immersed in this particular area of statistics. Therefore the Committee invited Dr. Sander Greenland of the University of California–Los Angeles to write a short editorial to be published with the report. Dr. Greenland's editorial reflects his understanding of and views about the methods developed. His comments are provided to assist the reader in understanding and interpreting this report and its contributions to epidemiologic methods for air pollution research.

In his invited editorial, Dr. Greenland agreed with the Review Committee that the research team largely achieved their major goal of developing highly flexible semiparametric regression analysis methods and successfully applying them to the analysis of a large data set. He noted that the similarity of the results from the current analysis and those from earlier analyses of the NMMAPS data was to be expected; the earlier analyses had already employed relatively flexible methods for control of time-based confounding variables, and the data did not contain many problematic departures from linear behavior that would have been better detected and controlled by these innovative methods.

Dr. Greenland also identified some important limitations of the new methods as they apply to air pollution and health research. One of his primary scientific concerns was that the development of methods did not include the type of dose–response modeling of the exposure's effect that is desirable for exploring different ambient concentrations that might be considered for regulatory standards. He also noted that the investigators chose to test their methods without incorporating established scientific understanding of how trends in certain variables may bias a time-series study. For example, the research team directly incorporated temperature and humidity data in the models as covariates, instead of using them to adjust the mortality data according to consensus assumptions about the effects of weather on daily mortality. Because the relationships between weather variables and mortality outcomes have been well explored and are well understood, relaxing such assumptions may unnecessarily reduce the precision of the results.

Dr. Greenland was also concerned about potential residual confounding from poorly defined or poorly measured exposure or from confounding variables in the models, since the team did not investigate how the commonly encountered types of measurement error might affect the results of this semiparametric analysis. Dr. Greenland noted, however, that exploration of the impact of these limitations was either beyond the scope of the defined research project or was not undertaken due to time and funding limitations.

CONCLUSIONS

The Review Committee agreed with most of the points that Dr. Greenland made. First, even in the largest and best-conducted observational studies, errors in the measurement of pollutant concentrations and major potential confounding factors create uncertainty in the magnitude, if not the direction, of estimated effects. Second, prior expert knowledge, to the extent that it exists, can be a valuable tool to inform the design and the interpretation of research. These are points that apply in general to observational epidemiology and have been recurring themes in Dr. Greenland's work. The Committee noted, however, that environmental epidemiologists conduct research in a world of imperfect data, and that no analytic method will ever be able to perfectly adjust for the shortcomings of such observational information as that found in the NMMAPS data set and others that have been established for large-scale air pollution and health outcomes research.

Overall, the Review Committee found that the semiparametric methods developed in this study are a promising addition to current practices for short-term studies of health events and air pollution levels. Although these methods do not address all of the potentially important sources of bias or confounding, the Committee agreed with the investigators that this research could be particularly useful in investigations where the relationships between time-varying confounders and health outcomes are not clearly understood or are difficult to characterize. The Committee also agreed with Dr. Greenland's suggestion that the applicability of these new methods and the precision of the estimates of risk that they produce could be improved in practice through the use of current methods to adjust for time-varying confounders when relationships between covariates and mortality are well understood.

HEALTH
EFFECTS
INSTITUTE

101 Federal Street, Suite 500
Boston, MA 02110, USA
+1-617-488-2300 phone
+1-617-488-2335 fax

pubs@healtheffects.org
www.healtheffects.org

Recycled Paper