STATEMENT

Synopsis of a Special Report

HEALTH EFFECTS INSTITUTE

Revised Analyses of Time-Series Studies of Air Pollution and Health

Over the past decade, time-series studies conducted in many cities have contributed information about the association between daily changes in concentrations of airborne particulate matter (PM) and daily morbidity and mortality. In 2002, however, investigators at Johns Hopkins University and at Health Canada identified issues in the statistical model used in the majority of time-series studies. This HEI Special Report details attempts to address several questions raised by these discoveries. The first section addresses the impact of the issues on the HEI-funded National Morbidity, Mortality, and Air Pollution Study (NMMAPS). The second section addresses the impact on additional studies selected by the US Environmental Protection Agency (EPA). Special Panels of the Health Effects Institute contributed Commentaries on the findings.

Analyses of the health effects of air pollution must account for other time-varying factors (such as weather and unmeasured risk factors) that may affect health outcomes. Otherwise, the effects of these factors could be counted as air pollution effects. Although many methods can be used for this purpose, generalized additive models (GAMs) have been the favored method in recent years. In May 2002, NMMAPS investigators at Johns Hopkins University discovered that part of the GAM programming in the S-Plus statistical software, which they and many others had used to fit GAMs to time-series data, was not entirely appropriate for this purpose. Specifically, the default convergence criteria were not appropriate and the iterative process required to obtain effect estimates was not likely to converge. After discovering these problems, the NMMAPS investigators quickly initiated alternative analyses of their data, including use of GAMs with appropriate convergence criteria, to see how the effect estimates might change. At about the same time, investigators at Health Canada found that, under certain conditions, programming to calculate standard errors of the regression coefficients in GAM software resulted in underestimates of the standard errors.

Concurrently, results of NMMAPS and other time-series studies were under review as part of the periodic review of the National Ambient Air Quality Standards (NAAQS) for PM. Thus understanding how these results might be changed by new analyses became a priority. As the funding sponsor for NMMAPS, HEI asked the NMMAPS investigators to prepare reports presenting their new analyses. Two NMMAPS reports were submitted to HEI: "Mortality Among Residents of 90 Cities" by Dominici and colleagues and "Morbidity and Mortality Among Elderly Residents of Cities with Daily PM Measurements" by Schwartz and colleagues. A Special Panel of the HEI Health Review Committee reviewed these reports.

In the summer of 2002, EPA identified additional key studies from the US, Canada, and Europe that were cited in the draft of the Air Quality Criteria for Particulate Matter and had used GAM in their analyses. The EPA requested that the investigators who had conducted those studies also carry out and report revised analyses. The agency asked that they (1) reanalyze the original data using the same nonparametric approach (GAMs) that was used originally, but with stricter convergence criteria; and (2) examine the sensitivity of the findings obtained with GAMs when using parametric models. The latter would also estimate more accurate standard errors.

EPA requested that HEI review the resulting short communication reports of the revised analyses and write a Commentary on the effect of different analytic approaches on the results. HEI agreed to take on this effort. A Special Panel of the HEI Health Review Committee, including members of the NMMAPS Review Panel and two additional methodologists, was formed to review the short communication reports. The Panel evaluated and interpreted changes in the original results due to the revised analyses but did not specifically evaluate the original study designs and methods.

Continued

This Statement, prepared by the Health Effects Institute, summarizes results from revised analyses of data from NMMAPS II and from selected time-series studies. The complete Special Report, *Revised Analyses of Time-Series Studies of Air Pollution and Health*, also contains Commentaries on each of these efforts written by special panels of the Institute's Health Review Committee. The complete Special Report can be requested from HEI or downloaded from our website (see reverse side). TIME-SERIES REV ANAL

METHODS

NMMAPS Revised Analyses

Reports of the revised NMMAPS analyses addressed both problems with the application of GAMs (settings for convergence criteria and maximum iterations, and standard error estimation) and left most other aspects of the analyses unchanged. Specifically, Dominici, Schwartz, and their colleagues carried out the following:

- *Replaced GAM functions with those using stricter convergence criteria*. These analyses were designed to correct the GAM convergence problem while acknowledging that the problem with standard error estimates was not addressed.
- *Replaced GAMs with generalized linear models (GLMs) with natural cubic splines*, using approximately the same degrees of freedom as were used in the original GAMs. These analyses were aimed at correcting problems with the standard errors and provided an alternative smoothing approach to GAMs.

Schwartz and colleagues used two additional alternatives to GAM and GLM for controlling temporal effects:

- *Penalized splines* with approximately the same degrees of freedom as in the original GAMs.
- *Case-crossover matching*. This approach was used as an alternative to GAM and GLM that might be conceptually more straightforward than the regression approaches for controlling temporal effects.

As in the original report, Dominici and colleagues applied the same model to each of the 90 cities included in the evaluation of daily mortality. The same variables and smoothing functions were used in each city to control for potential confounding, while parameter estimates and fitted smooth functions were allowed to vary from city to city. Schwartz and colleagues conducted the original and the revised analyses by fitting a city-specific model to each of the 14 cities included in the analysis of hospital admissions data and each of the 10 cities included in the evaluation of mortality. Both groups of investigators also reevaluated the effect of including copollutants in analytic models.

As in the original report, NMMAPS investigators calculated city-specific and overall estimates of mortality effects and investigated heterogeneity among cities. They also tested the sensitivity of the results to different degrees of control for unmeasured confounders.

Other Studies

In the revised analyses conducted at the request of EPA, the investigators sought to evaluate the sensitivity of effect estimates to choice of convergence criteria and maximum iterations in GAM and to use of parametric models that allow calculation of more accurate standard errors. EPA guidelines to authors suggested fitting a parametric model to the data with approximately the same degrees of freedom as for the original nonparametric model. Because of time limitations, investigators were encouraged to submit results of additional sensitivity analyses for publication elsewhere.

RESULTS

NMMAPS Revised Analyses

Overall, for the NMMAPS data, GAMs with stricter convergence criteria and GLMs with natural cubic splines resulted in lower estimates of effect than those from the original analyses conducted with GAM and default convergence criteria.

In individual cities, the revised effect estimates for mortality typically decreased and standard errors increased. Across the 90 cities, the revised mean effect on mortality decreased substantially from 0.41% (increase per 10 $\mu g/m^3$ increase in PM_{10} concentration at lag 1) to 0.27% when using GAM with stricter criteria and to 0.21% when using GLM with natural cubic splines: an overall decrease of nearly 50%. Lags 0 and 2 had corresponding decreases. Regional patterns of effect estimates remained across the 88 cities within the contiguous United States. Because the 90 city-specific estimates usually were smaller and generally had larger standard errors with the new analyses, tests for heterogeneity of effect across the cities indicated that heterogeneity was even less likely to be present than previously.

The overall decreases in effect estimates for hospitalizations for cardiovascular diseases and for chronic obstructive pulmonary disease were smaller (approximately 8% to 10%); a small but clear association continued to be found. The effect estimate on pneumonia hospitalizations was substantially reduced. As in the original studies, revised results for PM_{10} morbidity and mortality did not change substantially when copollutants were included in the models.

Other Studies

Nineteen primary authors submitted 21 short communication reports presenting results from analyses originally reported in 37 published original articles and reports. Differences between the original and

Continued

revised effect estimates varied substantially across and within studies. Overall, GAMs with stricter convergence criteria and GLMs with natural cubic splines yielded lower effect estimates but largely continued to identify an association of PM with mortality and morbidity, in particular for cardiovascular and respiratory diseases. A few investigators went beyond EPA's guidance and submitted additional sensitivity analyses. The impact of these analyses also differed across studies. No substantial impact was seen in some; in others, alternative modeling of time and weather factors resulted in substantial changes.

CONCLUSIONS

On the basis of their review, the Special Panels reached the following conclusions.

Study-Specific Conclusions

- In general, the estimates of effect in NMMAPS decreased substantially, but the qualitative conclusions did not change.
- Formal tests in NMMAPS for heterogeneity of PM effect across cities did not indicate heterogeneity. The Panel recognized, however, that the power to assess the presence of heterogeneity was low because of the generally larger city-specific standard errors. The possibility of heterogeneity therefore remains.

The overall impact of the other revised analyses included:

- While the number of studies showing an association of PM with mortality was slightly smaller, the PM association persisted in the majority of studies.
- In some of the large number of studies in which the PM association persisted, the estimates of PM effect were substantially smaller.
- In the few studies in which investigators performed further sensitivity analyses, some showed marked sensitivity of the PM effect estimate to the degree of smoothing and/or the specification of weather.

General Conclusions

- The impact of using more appropriate convergence criteria on the estimates of PM effect in the revised analyses varied greatly across the studies. In some studies, stricter convergence criteria had little impact, and in a few the impact was substantial. In no study were conclusions based on the original analyses changed in a meaningful way by the use of stricter criteria. Explanations for this variability considered by the Panel include the degree of temporal smoothing used in the original analyses, the number of smoothed terms in the models, and the degree of nonlinear collinearity (concurvity) among the smoothed terms. The relative importance of these and other explanations remains unclear.
- In general, the original PM effect estimates were more sensitive to the method used to account for temporal effects than to the convergence criteria used. Further, in the few studies in which temporal effects were extensively examined, some estimates of effect were more sensitive to the degree of smoothing of temporal effects than either the convergence criteria or the method used to account for temporal effects. In some studies the original effect estimates were largely insensitive to either the method or degree of smoothing. In several studies, however, the changes were substantial enough to result in meaningful changes in the study conclusions. In those few studies in which qualitative conclusions were changed as a result of the different approaches to smoothing, the revised results indicated no effect of PM.
- In most studies, parametric smoothing approaches used to obtain correct standard errors of PM effect estimates produced slightly larger standard errors than did GAM. The impact of these larger standard errors on level of statistical significance of the PM effect was minor.
- Alternative approaches used to model temporal effects in the revised analyses addressed the problems of obtaining incorrect effect estimates and standard errors when using GAMs. At this time, however, no approach can be strongly preferred over another for use in this context.
- These revised analyses have renewed the awareness of the uncertainties present in estimates of short-term air pollution effects based on time-series data. Neither the appropriate degree of control for time, nor the appropriate specification of the effects of weather, has been determined for time-series analyses. In the absence of adequate biological understanding of the time course of PM and weather effects, and their interactions, the Panel recommends exploration of the sensitivity of future time-series studies to a wider range of alternative degrees of smoothing and to alternative specifications of weather variables.

Continued

Impact

Air Pollution Time-Series Studies Compared with randomized experimental studies in which the investigator controls the intervention, findings from observational studies (such as time series) are always susceptible to uncontrolled biases and must therefore be interpreted cautiously. Observational air pollution and health studies are no exception. Uncovering inappropriate default convergence criteria in the GAM function again highlights the potential for confounding in air pollution time-series studies. As in many observational studies, avoiding confounding bias typically requires identification and specification of appropriate measures of the confounding factors as terms in a regression analysis. Determining the appropriate degree of smoothing time in air pollution time-series studies has become a central issue. Overly aggressive smoothing may allow residual confounding, whereas inadequate smoothing may allow some or all of the air pollution effect to be incorporated into the smooth term. The best method for selecting the appropriate degree of smoothing needed to control any confounding bias remains to be determined. Furthermore, as presented in the discussion of approaches to handling time, there is no gold standard for determining the appropriate degree for smoothing. The uncertainty that these issues introduce into time-series studies has motivated ongoing work to gain much needed insight. At this time, demonstration of sensitivity, or lack of it, to a range of sensible smoothing choices seems a reasonable approach.

Statistical Software The problem with applying GAMs has sent a cautionary note to investigators using statistical software. Clearly, the S-Plus GAM function underestimated standard errors in air pollution time-series studies, and until recently, the default convergence criteria were likely to lead to incorrect effect estimates. To their credit, investigators at Johns Hopkins continued to test their models and as a result brought the issue of default convergence criteria to light.

The nearly ubiquitous use of GAMs in time-series studies reflects one of the hazards of taking a standardized approach to analysis without verifying the detailed functioning of a given software. Clearly, as in this case, widespread use by applied biostatisticians and epidemiologists does not guarantee that a software or algorithm has no drawbacks. Looking ahead, analysts need to ensure that statistical software is appropriate for a given application. Again, the use of sensitivity analyses is included among these cautions (in this case addressing sensitivity, or the lack of it, in software tuning parameters and their defaults).

Impact Calculations Common practice has come to use effect estimates from observational air pollution studies to estimate the impact of air pollution on a large population such as an entire country. If effect estimates from the NMMAPS 90 cities mortality study were applied, the revised impact would be approximately half of the estimated impact derived using the original effect estimates. This example reinforces the need to qualify estimates of impact by specifying the assumptions and uncertainties on which the estimates are based.

Long-Term Effects Studies Some have noted that the calculated health impact of short-term air pollution based on time-series studies is substantially smaller than that of long-term air pollution based on cohort studies. Because of the vastly larger number of time-series studies performed, however, assessors of health risk from air pollution have often had more confidence in time-series results than in results from the few cohort studies. The problem with applying GAMs has involved primarily the time-series studies, however, and correction of the problem has generally decreased estimates of effect from these studies. Thus, more emphasis on cohort studies can be expected. Further, uncertainty regarding the estimates of effect obtained from time-series studies can also be expected to place additional emphasis on long-term air pollution studies, on studies of natural experiments (so-called quasi-experimental studies), and on human and animal experimental studies.

Closing

The Panels were impressed by the rapid reporting and comprehensive response to the discoveries by NMMAPS and other investigators regarding GAM software used in time-series studies. NMMAPS investigators conducted and reported results of additional analyses of virtually all of their previous NMMAPS research. Authors of the short communication reports were responsive to EPA's requests and completed a great deal of work in a short period of time. As with findings of the original analyses, all of the revised findings will continue to inform the regulatory process regarding PM. At the same time, these revised analyses have renewed the interest in important questions and uncertainties that should inform future time-series analyses of air pollution and health.

HEALTH
EFFECTS
INSTITUTE

Charlestown Navy Yard
120 Second Avenue
Boston MA 02129 USA

+1-617-886-9330
+1-617-886-9335 fax

pubs@healtheffects.org
www.healtheffects.org

Recycled Paper