Health Effects Institute
75 Federal Street
Suite 1400
Boston MA 02110 USA
+1-617-488-2300
FAX +1-617-488-2335
www.healtheffects.org

Comments of Daniel Greenbaum, President
Health Effects Institute (HEI)
July 17, 2018

HEI Comments on Proposed Rule EPA–HQ–OA–2018–0259; FRL–9977–40–ORD

HEI is pleased to have the opportunity to present these brief oral comments. We are preparing and will submit more detailed written comments

1. HEI has a longstanding commitment to the principles being addressed by this proposal: producing science of the highest integrity and quality, with special attention to issues of reproducibility and transparency. This includes:

    - Rigorous research and statistical design – Subject to competition, continuous oversight, data quality assurance audits, and more
    - Extensive efforts to test all findings against a wide range of different statistical techniques and assumptions
    - Intensive and independent peer review, with *all* results published
    - An active *Data Access Policy* for nearly 20 years to ensure access to underlying data for all HEI-funded studies.

2. Reproducibility is a critical challenge for science: can the results of an important study be reproduced? In HEI's view the most effective way to test the reproducibility and validity of scientific results is not necessarily to simply reproduce the same results in the same data sets -- because that also reproduces all the weaknesses and limitations of the original study. Rather, it is most important to answer the question: *Are the results consistent when tested in other independent studies*:

    - That use new and different data not affiliated with the original studies?
    - Have different investigators applying the same and/or alternative statistical techniques?
    - And test the sensitivity of the results against a wide range of possible other explanations, e.g. smoking behavior, socioeconomic status, access to medical care, and more.

3. In a limited number of cases, where there are not comparable studies in other datasets, it may be useful to gain access to the original study data and analytic codes to allow for independent evaluation: *Can the original results be replicated? And are they robust to a wide range of alternative assumptions, models and potential confounders?*

- This is the approach that HEI applied in its independent, rigorous reanalysis of the Harvard Six Cities and American Cancer Society Studies (see attached description of the Reanalysis):
- This approach can – and did – provide comprehensive assurance of the quality, integrity, and validity of the original results
- However, this is a highly cost-intensive and time-consuming endeavor which should only be applied in cases where there are one or just a few studies in a given area.

4. HEI also agrees with the continuing need to enhance transparency and data access, but would note that these issues are not new, and have been addressed now for over 15 years by administrations from both parties and by the scientific community:

    - This has included Guidelines for the Information Quality Act adopted by the Office of Information and Regulatory Affairs (OIRA) in 2002, numerous actions by the scientific community and journals to enhance access, and most recently the requirements for enhanced data access across the Federal Government promulgated by the Office of Science and Technology Policy (OSTP) in February 2013
    - We would strongly urge EPA to review the progress already made under these several major initiatives, and to carefully consider whether or not there are additional efforts that could further enhance transparency, *before* proceeding with a final rule.

5. Finally, access to private medical information is essential to conducting high quality and reproducible air quality and health research:

    - There are of course longstanding federal rules for protecting the privacy of individual medical information of the subjects of studies (HIPPA, Common Rule, etc.)

    - Gaining access to data from older studies may be difficult, given the privacy commitments that were made to study subjects in the past.

    - However, there *are* today several means to make such data available to investigators with appropriate privacy protections (e.g. Medicare, Federal Research Data Centers) and many investigators have been taking advantage of these.

    - Although it is possible, as some have suggested, to create a "depersonalized" data set by stripping all personal identifiers, such as address, date of birth, etc.
        - *It is not possible to conduct a high-quality air pollution and health study without knowing the locations of those being studied*, i.e. where they live, and what are the sources and levels of their air pollution exposure?

Thank you for this opportunity to testify – we look forward to submitting our detailed written comments and would welcome the opportunity to further assist EPA in these efforts to ensure the widest array of quality science is available for decisions.

ATTACHMENT: The HEI Reanalysis Statement

HEI

STATEMENT

HEALTH EFFECTS INSTITUTE

Synopsis of the Particle Epidemiology Reanalysis Project

BACKGROUND

Epidemiologic work conducted over several decades has suggested that long-term residence in cities with elevated ambient levels of air pollution from combustion sources is associated with increased mortality. Subsequently, two prospective cohort studies, the Six Cities Study (as reported in Dockery et al 1993) and the American Cancer Society (ACS) Study (as reported in Pope et al 1995) estimated that annual average all-cause mortality increased in association with an increase in fine particles (all particles less than 2.5 μm in median aerodynamic diameter [$PM_{2.5}$]).

As part of the Six Cities Study, Dockery and colleagues (1993) had prospectively followed a cohort of 8,111 adult subjects in northeast and midwest United States for 14 to 16 years beginning in the mid-1970s. The authors found that higher ambient levels of fine particles and sulfate (SO_4^{2-}) were associated with a 26% increase in mortality from all causes when comparing the most polluted to the least polluted city, and that an increase in fine particles was also associated with increased mortality from cardiopulmonary disease. The relative risks in all-cause mortality were associated with a difference (or range) in ambient fine particle concentrations of 18.6 μg/m^3 and a difference of ambient sulfate concentrations of 8.0 μg/m^3, comparing the least polluted city to the most polluted city.

In the much larger ACS Study, Pope and colleagues (1995) followed 552,138 adult subjects in 154 US cities beginning in 1982 and ending in 1989 (3 cities did not overlap between the 151 and 50 cities studied, resulting in a total of 154 cities). Again, higher ambient levels of fine particles were associated with increased mortality from all causes and from cardiopulmonary disease in the 50 cities for which fine particle data were available (sampled from 1979 to 1983). Higher ambient sulfate levels were associated with increased mortality from all causes, cardiopulmonary disease, and lung cancer in the 151 cities for which sulfate data were available (sampled from 1980 to 1982). The difference between all-cause mortality in the most-polluted city and the least-polluted city was 17% and 15% for fine particles and sulfate, respectively (with a range of 24.5 μg/m^3 for fine particles and of 19.9 μg/m^3 for sulfate).

Both of these studies came under intense scrutiny in 1997 when the EPA used the results to support new National Ambient Air Quality Standards for fine particles and to maintain the standards for particles less than 10 μm in median aerodynamic diameter (PM_{10}) already in effect. Members of Congress and industry, the scientific community and others interested in regulation of air quality scrutinized the studies' methods and their results. Some insisted that any data generated using federal funding should be made public. Others argued that these data had been gathered with assurances of confidentiality for the individuals who had agreed to participate and that the concept of public access to federally funded data did not take into account the intellectual property rights of the investigators and their supporting institutions. To address the public controversy, Harvard University and the ACS requested that the Health Effects Institute organize an independent reanalysis of the data from these studies. Both institutions agreed to provide access to their data to a team of analysts to be selected by HEI through a competitive process.

APPROACH

To conduct the reanalysis, the HEI Board of Directors, with support from the EPA, industry, Congress, and other stakeholders, appointed an Expert Panel chaired by Dr Arthur Upton from the University of Medicine and Dentistry of New Jersey and former Director of the National Cancer

This Statement, prepared by the Health Effects Institute, is a summary of a research project conducted by the Reanalysis Team, led by Dr Daniel Krewski at the University of Ottawa. The following Special Report contains the detailed Investigators' Report (Summary, Introduction, and Parts I and II), Commentary on the project prepared by a special panel of the Institute's Health Review Committee, and Comments on the Reanalysis Project by the Original Investigators (Drs Douglas W Dockery, C Arden Pope III et al).

Institute. The Expert Panel selected competitively a Reanalysis Team—led by Dr Daniel Krewski of the University of Ottawa—and oversaw all aspects of the team's work. They were assisted in their oversight efforts by a broad-based Advisory Board of knowledgeable stakeholders and scientists who, in the project's early stages, provided extensive advice to the Expert Panel on the key questions to be analyzed. The final results of the Reanalysis Team were intensively and independently peer reviewed by a Special Panel of the HEI Health Review Committee, which was chaired by Dr Millicent Higgins of the University of Michigan.

The overall objective of what became the Particle Epidemiology Reanalysis Project was to conduct a rigorous and independent assessment of the findings of the Six Cities and ACS Studies of air pollution and mortality. This objective was met in two parts. In *Part I: Replication and Validation*, the Reanalysis Team sought to replicate the original studies via a quality assurance audit of a sample of the original data and to validate the original numeric results. In *Part II: Sensitivity Analyses*, they tested the robustness of the original analyses to alternate risk models and analytic approaches.

RESULTS AND IMPLICATIONS

PART I: REPLICATION AND VALIDATION

- An extensive audit of the study population data for both the Six Cities and ACS Studies and of the air quality data in the Six Cities Study revealed the data to be of generally high quality with a few exceptions. In both studies, a few errors were found in the coding and inclusion of certain subjects; when those subjects were included in the analyses, they did not materially change the results as originally reported. Because the air quality data used in the ACS Study could not be audited, a separate air quality database was constructed for the sensitivity analyses described in Part II.
- The Reanalysis Team was able to replicate the original results in both studies using the same data and statistical methods as used by the Original Investigators. The Reanalysis Team confirmed the original point estimates: For the Six Cities Study, they reported the relative risk of mortality from all causes associated with an increase in fine particles of 18.6 $\mu g/m^3$ as 1.28, close to the 1.26 reported by the Original Investigators. For the ACS Study, the relative risk of mortality from all causes associated with an increase in fine particles of 24.5 $\mu g/m^3$ was 1.18 in the reanalysis, close to the 1.17 reported by the Original Investigators.

PART II: SENSITIVITY ANALYSES

Once the original results of the studies had been validated, the Reanalysis Team sought to test an array of different models and variables to determine whether the original results would remain robust to different analytic assumptions.

- First, the Reanalysis Team used the standard Cox model used by the Original Investigators and included variables in the model for which data were available from both original studies but had not been used in the published analyses (eg, physical activity, lung function, marital status). The Reanalysis Team also designed models to include interactions between variables. None of these alternative models produced results that materially altered the original findings.
- Next, for both the Six Cities and ACS Studies, the Reanalysis Team sought to test the possible effects of fine particles and sulfate on a range of potentially susceptible subgroups of the population. Although different subgroups did show some variation in their estimated effects, the results were not statistically significant with one exception. The estimated effects of fine particles did appear to vary with educational level; the association between an increase in fine particles and mortality tended to be higher for individuals without a high school education than for those who had completed high school or for those with more than a high school education.
- In the ACS study, the Reanalysis Team tested whether the relationship between ambient concentrations and mortality was linear. They found some indications of both linear and nonlinear relationships, depending upon the analytic technique used, suggesting that the

issue of concentration-response relationships deserves additional analysis.

- In the Six Cities Study where data were available, the Reanalysis Team tested whether effect estimates changed when certain key risk factors (smoking, body mass index, and air pollution) were allowed to vary over time. One of the criticisms of both original studies has been that neither analyzed the effects of change in pollutant levels over time. In general, the reanalysis results did not change when smoking and body mass index were allowed to vary over time. The Reanalysis Team did find for the Six Cities Study, however, that when the general decline in fine particle levels over the monitoring period was included as a time-dependent variable, the association between fine particles and all-cause mortality dropped substantially, but the effect continued to be positive and statistically significant.
- Using its own air quality dataset constructed from historical data to test the validity of the original ACS air quality data, the Reanalysis Team found essentially the same results.
- Any future analyses using the sulfate data should take into account the impact of artifactual sulfate. Sulfate levels with and without adjustment differed by about 10% for the Six Cities Study. Both the original ACS Study air quality data and the newly constructed dataset contained sulfate levels inflated by approximately 50% due to artifactual sulfate. For the Six Cities Study, the relative risks of mortality were essentially unchanged with adjusted or unadjusted sulfate. For the ACS Study, adjusting for artifactual sulfate resulted in slightly higher relative risks of mortality from all causes and cardiopulmonary disease compared with unadjusted data. The relative risk of mortality from lung cancer was lower after the data had been adjusted.
- Because of the limited statistical power to conduct most sensitivity analyses for the Six Cities Study, the Reanalysis Team conducted the majority of its sensitivity analyses using only the ACS Study dataset with 154 cities. In that dataset, when a range of city-level (ecologic) variables (eg, population change, measures of income, maximum temperature, number of hospital beds, water hardness) were included in the analyses, the results generally did not change. Two exceptions were that associations for both fine particles and sulfate were reduced when city-level measures of population change or sulfur dioxide were included in the model.
- A major contribution of the Reanalysis Project is the recognition that both pollutant variables and mortality appear to be spatially correlated in the ACS Study dataset. If not identified and modeled correctly, spatial correlation could cause substantial errors in both the regression coefficients and their standard errors. The Reanalysis Team identified several methods for dealing with this, all of which resulted in some reduction in the estimated regression coefficients. The full implications and interpretations of spatial correlations in these analyses have not been resolved and appear to be an important subject for future research.
- When the Reanalysis Team sought to take into account both the underlying variation from city to city (random effects) and the spatial correlation between cities, only sulfur dioxide as a city-level variable continued to decrease the originally reported associations between mortality and fine particles or sulfate. This effect was more pronounced for sulfate.
- When the Reanalysis Team conducted spatial analyses of sulfur dioxide, the association between sulfur dioxide and mortality persisted after adjusting for sulfate, fine particles, and other variables.
- As a result of these extensive analyses, the Reanalysis Team was able to explain much of the variation between cities, but some unexplained city-to-city variation remained.

CONCLUSIONS

The Reanalysis Team designed and implemented an extensive and sophisticated series of analyses that included a set of new variables, all the gaseous copollutants, and the first attempts to apply spatial analytic methods to test the validity of the data and the results from the Six Cities Study and the ACS Study. Overall, the reanalyses assured the quality of the original data, replicated

the original results, and tested those results against alternative risk models and analytic approaches without substantively altering the original findings of an association between indicators of particulate matter air pollution and mortality.

At the same time, the reanalyses did extend and challenge our understanding of the original results in several important ways.

- The Reanalysis Team identified a possible modifying effect of education on the relation between air quality and mortality in that estimated mortality effects increased in the subgroup with less than high school education.
- The use of spatial analytic methods suggested that, when the analyses controlled for correlations among cities located near one another, the associations between mortality and fine particles or sulfate remained but were diminished.
- An association between sulfur dioxide and mortality was observed and persisted when other possible confounding variables were included; furthermore, when sulfur dioxide was included in models with fine particles or sulfate, the associations between these pollutants (fine particles and sulfate) and mortality diminished.

In reviewing these results, the Special Panel of the HEI Health Review Committee identified the following factors to consider when interpreting the results from the Reanalysis Team.

- The inherent limitations of using only six cities, understood by the Original Investigators, should be taken into account when interpreting results of the Six Cities Study.
- The Reanalysis Team did not use data adjusted for artifactual sulfate for most alternative analyses. When they did use adjusted sulfate data, relative risks of mortality from all causes and cardiopulmonary disease increased. This result suggests that more analyses with adjusted sulfate might result in somewhat higher relative risks associated with sulfate.
- Findings from spatial analyses applied to the ACS Study data need to be interpreted with caution; the spatial adjustment may have overadjusted the estimated effect for regional pollutants such as fine particles and sulfate compared with the effect estimates for more local pollutants such as sulfur dioxide.
- After the Reanalysis Team completed its spatial analyses, residual spatial variation was still noticeable; this finding suggests that additional studies might further refine our understanding of the spatial patterns in both air pollution and mortality.
- No single epidemiologic study can be the basis for determining a causal relation between air pollution and mortality.

In conclusion, the Reanalysis Team interpreted their findings to suggest that increased relative risk of "mortality may be attributed to more than one component of the complex mix of ambient air pollutants in urban areas in the United States". The Review Panel concurs. In the alternative analyses of the ACS Study cohort data, the Reanalysis Team identified relatively robust associations of mortality with fine particles, sulfate, and sulfur dioxide, and they tested these associations in nearly every possible manner within the limitations of the datasets. Future investigations of these issues will enhance our understanding of the effect of combustion-source air pollutants (eg, fine particles, sulfate, and sulfur dioxide) on public health.