The Honorable Lamar Smith
Chairman
Committee on Science, Space, and Technology
2321 Rayburn House Office Building
Washington, DC 20515

Dear Chairman Smith,

It was with great interest that I read your July 22, 2013, letter to Environmental Protection Agency (EPA) Administrator Gina McCarthy. In that letter, you outline the several requests Majority Members of the Committee on Science, Space, and Technology have made to the EPA for disclosure to the Committee of research data used in certain studies which the EPA relied upon in various Clean Air Act related regulatory decisions. Your letter states that you haven’t received an adequate response from the EPA regarding these requests. Perhaps you are unaware that the EPA has responded on at least two occasions with detailed information pertaining to the studies you requested (attached). Since these detailed responses have not satisfied you, I feel compelled to highlight my own concerns about these ill-advised requests for vast quantities of American citizens’ personal health data.

Your July 22 letter contains material inaccuracies and conveys a general sense of misunderstanding of the underlying issues. Although the letter is mysteriously silent in specifying exactly which research data you are seeking, it is apparent from context that the information you are requesting stem from to two seminal studies on the health effects of long-term exposure to air pollution: “An Association between Air Pollution and Mortality in Six U.S. Cities” published in the New England Journal of Medicine and “Particulate Air Pollution as a Predictor of Mortality in a Prospective Study of U.S. Adults” which was published in the American Journal of Respiratory and Critical Care Medicine.¹ Both of these groundbreaking studies showed a causal relationship between

long-term exposure to air pollution and an increased likelihood of death. In addition, both studies were the subject of peer reviewed follow-up studies. ²

Despite the widely acknowledged seminal nature of these studies, in your July 22 letter you state:

“The National Academy of Sciences has stated that these analyses have “little use for decisionmaking.”

This is a gross mischaracterization of what the National Academy of Sciences (NAS) report says. What is most remarkable about this mischaracterization is that you actually provided the original NAS document in one of your previous letters which you attached to the July 22 letter. The relevant sentence you purport to paraphrase reads in full:

“Although these cohorts have provided critical evidence for long-term effects, evidence from further follow-up of these two U.S. cohorts alone will have little use for decisionmaking.” (emphasis added)

I assume that the mischaracterization in your July 22 letter is due to a misunderstanding of the difference between a research finding and a cohort. However, if this is the case of misunderstanding, then it gives me great pause in considering the prospect of you and your staff receiving the underlying data being sought.

The cohorts to which the NAS document you mischaracterized refer to are actually groups of American people. The Harvard Six Cities Study involved a cohort of random samples of adults collected from six different U.S. cities. The Pope study utilized the American Cancer Society’s Cancer Prevention Study II (CPS 2) cohort, which consisted of 1.2 million human subjects from across America. In both studies, the researchers tracked the cohorts over long periods of time to ultimately reach their conclusions. In other words, the cohorts are the groups of people being tracked (the research subjects), not the conclusions being reached by the researchers.

What the NAS was referring to in the above quote is that by 2004, the cohorts used in the Harvard Six Cities Study and the CPS 2 study were aging, dying, and otherwise losing touch with the researchers. For instance, of the CPS 2 cohort of 1.2 million people the American Cancer Society began tracking in 1982, almost 500,000 were deceased by 2006. The NAS document was saying that because of these inevitable issues with the existing aging cohorts, new cohorts were needed in order to conduct these types of long-term epidemiological studies in the future. This was not a criticism of the cohort data that had been collected up to that point, or any of the research results based upon that data. In fact, the American Cancer Society began enrollment for a new

cohort (CPS3) in 2006 to address this very issue. I would also note that the CPS 2 cohort data has been widely used in much groundbreaking health and cancer research, with scores of peer-reviewed papers based upon this data set. In addition to the groundbreaking air pollution research findings, scientists have also used CPS 2 data to show that cigarettes with reduced yield of tar and nicotine do not reduce the risk of lung cancer and that obesity is associated with increased mortality from at least ten cancer sites. These are just a few highlights of the significant contributions to public health associated with research using the CPS 2 cohort data.

Moreover, the same paragraph that you mischaracterize in support of your arguments criticizing these research papers actually refutes your own unsupported assertions. Not two sentences before your misquote, the NAS paper reads:

“The findings of the two studies were confirmed with an extensive reanalysis (Krewski et al. 2000) and on further follow-up of the CPS 2 cohort (Pope et al. 2002).”

Again, I do not understand how anyone could read that page of the NAS report (attached), and then describe it as it was described in your July 22 letter.

In further criticism of these two studies you go on to state that “these analyses are inconsistent with studies based on more recent information.” However, you fail to support that loose assertion with any relevant research citations. You also conveniently ignore the other confirmatory research which was cited in the NAS report your letter mischaracterized. That paragraph, which can be found in an attachment to your own letter, lists no fewer than five other studies which support the general findings of the Harvard Six Cities Study and the CPS 2 study.

This cavalier mischaracterization of the results of research that has a profound impact on the health of American citizens concerns me greatly. It concerns me because nowhere in your letter do you indicate what you propose to do with this cohort data once it is in your possession. I have reviewed your staff list, and there does not appear to be a single epidemiologist in your employment. I can’t identify a single person on your staff who would be qualified to use this data in any meaningful way. And to be clear, we are talking about a massive amount of information. As I mentioned before, the CPS 2 cohort consisted of 1.2 million subjects. It is absurd to consider that our small Committee staff, talented as they are at their jobs, could make any legitimate use of this massive amount of health and environmental data.

So, I am at a loss to understand why you are requesting this data. Some would see this request as simply an attempt to harass the EPA. I would obviously strongly object to the needless harassment of the EPA, which is already operating in difficult budgetary times. Others may suggest that these are efforts to supply outside parties with data they could otherwise not obtain. I would also strongly object to obtaining this data for outside parties, who would not otherwise be given access to this sensitive data. That is not our job. Much, if not all of this data is already available for legitimate research
purposes. For instance, a simple search through the American Cancer Society’s website will lead you to the policies and procedures they have for application to access the CPS data sets. Many of the policies on CPS data access relate to human subject confidentiality and research protocols. Since the data in question are from human research subjects, there are, understandably, ethical training requirements for researchers requesting this data. People requesting this data must also have a legitimate research aim, as the subject data was released by the participants for this purpose only.

I am forced to question the scientific legitimacy of groups which cannot already obtain this data. They must not be legitimate scientists or must be untrustworthy with human research subject data, or they could simply apply to the American Cancer Society directly. I certainly don’t think Congress should be obtaining confidential human research data to supply to outside groups who can’t pass ethical muster, and I sincerely hope that is not the goal of this endeavor.

In conclusion, I am increasingly skeptical of the legitimacy of this data request. The data you are requesting is data that American citizens allowed to be collected with the understanding that it would be for scientific research purposes only. You have requested the personal medical histories of literally hundreds of thousands of American citizens. And for what purpose? There is no conceivable way that you or your staff could meaningfully use this data to refute the seminal health studies you seem preoccupied with attacking. Not unless you hired a team of expert epidemiologists. With the Republican budget cuts to our Committee’s funding, your ability to do such hiring seems highly unlikely. The truth is that there is no legitimate reason to warrant violating the trust of hundreds of thousands of American citizens who volunteered their personal information to make everybody’s lives better. For all of the reasons I have noted in this letter, I strongly urge you to stop what you are doing.

If you persist in this effort, I want to let you know in the clearest terms possible that I will require the following information before I can even consider supporting authorizing a subpoena for this information:

1. For what purpose do you seek this human subject data?
2. If Committee staff are to review this data, who are they, what are their scientific qualifications, and have they completed ethical training for handling human subject data? What steps will the Committee take to ensure compliance with the Ethical Principles and Guidelines for the Protection of Human Subjects of Research (Belmont Report)?
3. If it is your intent to transmit this data to outside entities, who are those entities and for what purpose are they being supplied with this data?
Thank you for your attention to this matter.

Sincerely,

[Signature]

EDDIE BERNICE JOHNSON
Ranking Member
Committee on Science, Space, and Technology

3 Attachments

Cc: The Honorable Gina McCarthy
Administrator, U.S. Environmental Protection Agency

Members, Committee on Science, Space, and Technology
The Honorable Andy Harris, M.D.
U.S. House of Representatives
Chairman, Subcommittee on Energy and Environment
Washington, D.C. 20515

Dear Chairman Harris:

Thank you for your letter of September 22, 2011, in which you ask several questions regarding the Environmental Protection Agency’s (EPA) Cross-State Air Pollution Rule (CSAPR) as a follow-up to the Science, Space, and Technology Committee’s September 15 hearing on this topic. I appreciate the opportunity to provide the additional information you have requested.

With regard to EPA’s estimates of avoided premature mortality under CSAPR, EPA estimated the number of fine particle (PM$_{2.5}$)-related deaths avoided due to the implementation of this rule using two long-term prospective cohort studies. The first study is the extended analysis of the American Cancer Society cohort by Pope and colleagues published in 2002.\(^1\) The second is an extended analysis of the Harvard Six Cities cohort by Laden and colleagues published in 2006.\(^2\) There are strengths to each study that argue for using both as the basis for the PM mortality estimates in CSAPR. In particular, the Harvard Six Cities cohort is located in cities in the eastern United States, which is the geographic area covered by CSAPR, and the demographic characteristics of this cohort are representative of the broader U.S. population. The American Cancer Society cohort is larger than the Six Cities cohort and covers a broader number of urban areas across the United States. Using these studies, we reported two estimates of PM$_{2.5}$-related mortality: 13,000 (95% confidence intervals from 5,200 to 21,000) using the Pope et al. (2002) study and 34,000 (95% confidence intervals from 18,000 to 49,000) using the Laden et al. (2006) study. Thus the phrase "up to 34,000" refers to the higher of the two central estimates from the range of results while communicating that there is uncertainty in the estimates. We did not separately estimate premature mortality impacts for different diseases for the CSAPR. We quantify all-cause mortality rather than cardiopulmonary or lung cancer mortality specifically because it is the most comprehensive estimate of PM-related mortality as supported by the scientific literature.

The EPA’s Regulatory Impact Analysis\(^3\) for CSAPR describes in detail the methods and data we employed to quantify these impacts as well as a suite of sensitivity and uncertainty analyses. Our

---


approach to quantifying the benefits of air quality improvements in general, and our reliance on the two studies mentioned above in particular, has been thoroughly reviewed by independent scientific bodies including the National Research Council and Advisory Council on Clean Air Compliance Analysis. EPA did not perform a Quality Adjusted Life Years (QALY) analysis for the Cross State Rule due to continuing methodological concerns about the approach.

With regard to your questions about the number of avoided premature deaths and other health benefits estimated by EPA in analyses for CSAPR and other recent Clean Air Act Rules, EPA has prepared a summary table (below) with links to the Regulatory Impact Analysis (RIA) of all Clean Air Act Rules issued since January 20, 2009 that estimated PM$_{2.5}$-related premature deaths. These RIAs provide information on specific health effects (including avoided premature deaths) attributable to each rule as you requested in your letter.

<table>
<thead>
<tr>
<th>Rule</th>
<th>Link to Document on EPA’s Website</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Existing Stationary RICE NESHAP</strong></td>
<td></td>
</tr>
<tr>
<td>Final Compression Ignition RIA (2/22/2010)</td>
<td><a href="http://www.epa.gov/trt/ecaas/regdata/RIAs/CIRCENESHA21RI2-17-10cleanpublication.pdf">http://www.epa.gov/trt/ecaas/regdata/RIAs/CIRCENESHA21RI2-17-10cleanpublication.pdf</a></td>
</tr>
<tr>
<td><strong>Cement NESHAP and NSPS</strong></td>
<td></td>
</tr>
<tr>
<td><strong>C3 Marine Rule</strong></td>
<td></td>
</tr>
<tr>
<td><strong>NO$_2$ NAAQS</strong></td>
<td></td>
</tr>
<tr>
<td><strong>2012-16 Light Duty Vehicle Rule</strong></td>
<td></td>
</tr>
<tr>
<td><strong>SO$_2$ NAAQS</strong></td>
<td></td>
</tr>
</tbody>
</table>

---


Ozone NAAQS Reconsideration Proposal

Boiler NESHAP and Area Source Rule

Solid Waste Incineration Units
NSPS and Emission Guidelines

Cross-State Air Pollution Rule
Final RIA (7/12/2011)  http://www.epa.gov/airtransport/pdfs/FinalRIA.pdf

2014-18 Heavy Duty Vehicle Rule

Sewage Sludge Incineration Units
NSPS and Emission Guidelines

Mercury and Air Toxics Standards (Utility NESHAP and NSPS)

Chlor Alkali Plants NESHAP

Ferroatloys RTR

With regard to the final question in your letter about my testimony regarding the availability of the scientific support for EPA's estimates of avoided premature deaths under CSAPR, there are numerous relevant documents that are publicly available and that have been through public comment and review. As is the case for all our significant rules, the basis for our benefits estimates for the Cross State Rule is set forth in the Regulatory Impact Analysis (RIA). 6 Chapters 3 and 4 discuss our analysis of the Cross State Rule's projected effect on emissions and air quality. That information then feeds into the benefits assessment, which is contained in Chapter 5. The benefits chapter alone runs approximately 250 pages, including 25 pages of references. This RIA went through the usual thorough vetting to which all RIAs

6 http://www.epa.gov/airtransport/pdfs/FinalRIA.pdf
are subject. After undergoing inter-agency review under the auspices of the Office of Management and Budget (OMB), we release proposed RIAs for public review and comment at the same time that we release the proposed rule. We then review and respond on the record to any significant public comments on the RIA, including the benefits analysis. Draft final RIAs also undergo interagency review before EPA finalizes them.

The scientific studies used by EPA as the basis for estimating public health benefits are evaluated by EPA during the development of the Integrated Scientific Assessments and the Risk/Exposure Assessments that EPA periodically issues for ozone, fine particles and other criteria pollutants pursuant to Sections 108 and 109 of the Clean Air Act. An Integrated Science Assessment is a comprehensive review, synthesis, and evaluation of the most policy-relevant peer-reviewed science, including key science judgments that are important to inform the rest of the review process for setting national ambient air quality standards. The Risk/Exposure Assessment draws upon the corresponding Integrated Science Assessment to characterize exposures and associated risks to human health or the environment associated with recent air quality conditions. These documents are peer reviewed by the independent and statutorily-mandated Clean Air Scientific Advisory Committee (CASAC), in addition to undergoing extensive public review and comment. Although the documents have shorter overview sections, the most recent assessments for particulate matter, for example, contain thousands of pages of analysis based on peer-reviewed scientific studies.\(^7\)

The benefits estimates also rely on rigorous, peer-reviewed methodologies grounded firmly in a vast body of research related to the health effects of air pollution. Our benefits assessment methods have been extensively peer reviewed and supported by the National Academies of Science and several panels of the independent EPA Science Advisory Board.\(^8\)

In response to the new request in your letter regarding the availability of data and analyses from five epidemiological studies (two American Cancer Society studies, the Harvard Six Cities Study, and two Nurses Health studies), we will take action under 2 CFR 215.36 as soon as possible to provide you with any data and analyses produced with EPA funds to the extent that this information remains available.

Again, thank you for your letter. If you have further questions, please contact me or your staff may call Diann Frantz in EPA's Office of Congressional and Intergovernmental Relations at (202) 564-3668.

Sincerely,

[Signature]

Gina McCarthy
Assistant Administrator

---

\(^7\) http://www.epa.gov/ttn/naaqs/standards/pm/s_pm_index.html

\(^8\) See, e.g., reports cited n. 4 and 5 above.
The Honorable Andy Harris, MD
Chairman
Subcommittee on Energy and Environment
U.S. House of Representatives
Washington, D.C. 20515-6301

Dear Mr. Chairman:

On September 22, 2011, you requested data and analyses from five epidemiological studies used in the benefits analysis for the Cross-State Air Pollution Rule. As outlined in my November 30, 2011 letter, the U.S. Environmental Protection Agency agreed to take action under 2 CFR 215.36 to request the information produced with the EPA funds, to the extent that this information remains available. Only two of the five epidemiological studies identified in your request were used in the benefits analyses and were the focus of our data collection efforts – the Cancer Prevention Study II compiled by the American Cancer Society (Pope et al., 2002) and the Harvard Six Cities Study (Laden et al., 2006). The remaining three studies – the Cancer Prevention Study I compiled by the American Cancer Society and two Nurses Health studies – were not used in our benefits analyses and, therefore, EPA did not request data for these studies.

On January 9, 2012, the EPA sent letters to New York University and Harvard University requesting any research data produced with the EPA grant funds relating to the Pope et al. (2002) and Laden et al. (2006) studies, respectively. We provided copies of these letters to your staff. Enclosed are the data provided by these universities in response to these requests including additional information from the researchers specifying a key to clarify the data contained in the spreadsheets. The enclosed data represent the data that were developed with the EPA grant monies. We note that the American Cancer Society and Harvard Six Cities studies are large epidemiological research projects that have received funding from a number of sources.

For the Laden et al. (2006) study, the health event data (i.e., deaths) were obtained from the National Death Index (NDI), which is part of the National Center for Health Statistics (NCHS) at the Centers for Disease Control and Prevention (CDC). These data are available to researchers exclusively for medical and health research statistical analyses. To obtain data from the NDI, the Harvard University...
researchers signed a confidentiality agreement in accordance with section 308(d) of the Public Health Service Act (42 U.S.C. § 242m(d)) promising not to publish or release data in any form to any party if a particular individual was identifiable. The Public Health Services Act provides that the data collected by NCHS may be used only for the purpose for which they were obtained; any effort to determine the identity of any reported cases, or to use the information for any purpose other than for health statistical reporting and analysis, would violate the statutory restriction.

In the data set provided by Harvard University, the researchers indicated that they believe these data could not be used to identify individuals without additional analysis and information. Additionally, we coordinated extensively with CDC to determine whether there may be restrictions on disclosure of these data to your Committee. CDC advised that any NDI information about an individual that is more specific than what was provided by Harvard University may not be shared with anyone who has not already signed a confidentiality agreement with NCHS.

The enclosed data complete our response to the questions you raised in the September 22, 2011 letter. If you have further questions, please contact me or your staff may call Cheryl Mackay in EPA’s Office of Congressional and Intergovernmental Relations at (202) 564-2023.

Sincerely,

[Signature]

Gina McCarthy
Assistant Administrator

Enclosures
Attachment A:

Excerpt from the National Research Council’s 2004 report, *Research Priorities for Airborne Particulate Matter: IV. Continuing Research Progress*

**INVESTIGATING THE HEALTH EFFECTS OF LONG-TERM EXPOSURE TO AIR POLLUTION**

**Epidemiological Approaches**

The striking findings of the Harvard Six Cities Study (Dockery et al. 1993), which linked chronic exposure to increased mortality, provided a strong impetus for reevaluating the PM NAAQS, particularly after their confirmation in the 1995 publication based in the American Cancer Society’s Cancer Prevention Study 2 (CPS 2) (Pope et al. 1995). The findings on increased mortality associated with longer-term exposures to higher concentrations of particles suggested that the associations observed in the time-series studies did not reflect only a slight advancement of the timing of death for frail individuals. The findings of the two studies were confirmed with an extensive reanalysis (Krewski et al. 2000) and on further follow-up of the CPS 2 cohort (Pope et al. 2002). Findings from several other cohort studies have also been reported (Abbey et al. 1999; Lippert et al. 2000; Hoek et al. 2002). Although these cohorts have provided critical evidence for long-term effects, evidence from further follow-up of these two U.S. cohorts alone will have little use for decisionmaking. The cohorts were established decades ago, and some critical data items, including residence history and potential confounding and modifying factors, have not been comprehensively updated. Consequently, an increasing degree of exposure misclassification can be anticipated as the participants move from their original residences. And, most important, characterization of current air quality cannot recreate the complex air environments in which the individuals and populations lived and worked in the many years for which data are not available. Long-term studies are likely to remain central, however, in assessing the public health burden caused by air pollution. For quantitative risk assessment and cost-benefit analysis, estimates of the disease burden associated with exposure to particles are needed. These estimates could come from a new generation of studies with more complete information on short- and long-term exposures to PM, its components, and exposures to other pollutants.

Recognizing both the limitations of these studies and the need for ongoing information on long-term exposure to air pollution and health, the committee recommends that research approaches continue to be developed on the basis of existing and new cohorts. Mechanisms are needed for enrollment and tracking of cohorts over time to provide an ongoing characterization of any impact on health of long-term exposure to air pollution. Without substantial commitment of personnel and funds, studies, such as the Six Cities Study and the CPS 2 cohorts, cannot be readily and feasibly undertaken. Rather, such studies might be based on cohorts routinely enrolled for other purposes, for example, investigating cardiovascular diseases (Atherosclerosis Risk in Communities [ARIC 2004] and the Cardiovascular Health Study [CHS 2003]), Medicare participants, and cohorts assembled by the National Center for Health Statistics. However, even such studies will require substantial funding, and their value must be compared with data collection specifically designed as long-term studies of health effects of air pollution. Medicare has a large cohort under follow-up that is maintained with replacement sampling. The Veterans