American Thoracic Society

American Journal of Respiratory and Critical Care Medicine ®
American Journal of Respiratory Cell and Molecular Biology ®
Annals of the American Thoracic Society ®

JAMES M. BECK, MD, ATSF
President

JUAN C. CELEDÓN, MD, DrPH, ATSF
President-Elect

POLLY E. PARSONS, MD, ATSF
Immediate Past President

LYNN M. SCHNAPP, MD, ATSF
Vice President

GREGORY P. DOWNEY, MD, ATSF
Secretary-Treasurer

KAREN J. COLLISHAW, MPP, CAE
Executive Director

January 17, 2020

Dr. Michael Honeycutt
Chair, EPA Scientific Advisory Board
Environmental Protection Agency
1200 Pennsylvania Avenue, N.W.
Washington, DC 20460

RE: ATS comments on SAB review of EPA Transparency Rule

Dr. Honeycutt:

The American Thoracic Society appreciates the opportunity to comment on EPA's Scientific Advisory Board review of the EPA's proposed rule: Strengthening Transparency in Regulatory Science. We appreciate the thoughtful review and comments provided by members of SAB.

The ATS shares many of the concerns that SAB outlined in its draft letter and appreciates the recommendations SAB has provided to address these concerns. However, the ATS believes strongly that the rule is severely flawed and even with the recommendations offered by SAB, the underlying rule does not benefit public health. We continue to urge EPA to abandon this misguided policy.

While there are several flaws discussed in the SAB letter, the ATS would like to focus our comments on one specific issue – risks of data re-identification. Recent studies have shown that the risk for data re-identification for HIPAA compliant data are greater than previously thought, and the EPA appears to under-appreciate this risk. The ATS fears that, if implemented, the proposed Strengthening Transparency in Regulatory Science rule would severely jeopardize private health information of American citizens.

The proposed rule states that the EPA will apply the tools and methods developed by other Federal agencies to de-identify private information. The rule cites a Department of Health and Human Services (HHS) document as an example of how data can be de-identified to protect confidential patient information. (1) However, the very HHS guidance that EPA references notes that de-identification does not fully protect patient information, stating:

"Both methods, even when properly applied, yield de-identified data that retains some risk of identification. Although the risk is very small, it is not zero, and there is a possibility that de-identified data could be linked back to the identity of the patient to which it corresponds."

For environmental health research, which often involves information about location, the risk of data re-identification may be even greater. A recent (2017) study by Sweeney and colleagues took the HIPAA-compliant de-identified data from an air pollution epidemiology study, and using other publicly available data sets and commercially available computer programs, successfully re-identified 25% of study participants. (2) Even the Centers for Medicare and Medicaid Services requires researchers who use Medicare information to report only summary information across large numbers of people to ensure that individual people cannot be identified. A recent publication in *Nature Communications* strongly challenges the "de-identification release-and-forget" approach, finding that nearly 99.98% of Americans could be re-identified with just 15 demographic variables.(3) EPA's proposal to make the underlying data of policy-relevant studies fully available to the public in de-identified form risks disseminating the sensitive information about health problems, habits (such as smoking, drug use, etc.), deaths and addresses of study participants. Re-identification of a public de-identified dataset could harm those whose information is disclosed, and may also discourage others from participating in environmental health research about toxins in the air, water and food. In addition to the other concerns listed by SAB, we strongly urge SAB to consider likely unappreciated risk of data re-identification under the proposed EPA policy.

We would similarly note that it is impossible to justify the high risk of unintended harm caused by the proposed rule given that there has been no demonstration of potential benefit provided by the rule. Even the vague outlines of how this rule would be carried out lack any hint of benefits that might be expected to result once the rule is promulgated. If there is a demonstrable benefit of the proposed rule, it should be carefully examined in comparison to the likely harms created before the rule is promulgated. Even if revised according to the SAB comments, the rule would create complex and costly bureaucratic barriers to the use of peer-reviewed science in EPA rule-making, would exclude many environmental health studies from EPA consideration, and risks dissemination of private health data of American citizens, without any benefit to public health. The ATS therefore strongly urges EPA to abandon this misguided and harmful proposed rule.

Attached, please find the testimony of Mary B. Rice MD presented on behalf of the American Thoracic Society before the House Science Committee in November 2019 that further expand on the ATS's concerns with EPA proposed transparency policy.

We hope the SAB will consider our concerns about data re-identification and other serious issues in the attached science committee testimony as the committee moves forward with its deliberation.

Sincerely,

Mary B. Rice, MD
Chair, ATS Environmental Health Policy Committee
American Thoracic Society

1. US Dept Health and Human Services: Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule. https://www.hhs.gov/hipaa/forprofessionals/privacy/special-topics/de-identification/index.html#rationale.

2. Sweeney L, Yoo JS, Perovich L, Boronow KE, Brown P, Brody JG. Re-identification Risks in HIPAA Safe Harbor Data: A study of data from one environmental health study. Technol Sci. 2017;2017.

3. Rocher L, Hendrickx JM, de Montjoye Y-A. Estimating the success of re-identifications in incomplete datasets using generative models. Nat Commun. 2019;10(1):3069.

JAMES M. BECK, MD, ATSF
President

JUAN C. CELEDÓN, MD, DrPH, ATSF
President-Elect

POLLY E. PARSONS, MD, ATSF
Immediate Past President

LYNN M. SCHNAPP, MD, ATSF
Vice President

GREGORY P. DOWNEY, MD, ATSF
Secretary-Treasurer

KAREN J. COLLISHAW, MPP, CAE
Executive Director

Comments of the American Thoracic Society
Before the House Science Committee
Regarding EPA's Proposed Strengthening Transparency in Regulatory Science Rule
Presented by
Mary B. Rice MD MPH
November 13, 2019

Mr. Chairman, Ranking member and members of the House Science Committee, thank you for the opportunity to speak at today's hearing.. I am a pulmonary and critical care physician and assistant professor of medicine at Harvard Medical School, where I treat patients with lung disease, and investigate the effects of air pollution on lung health. I am speaking today on behalf of the American Thoracic Society (ATS). We are a 16,000-member medical professional organization of physicians, researchers, nurses, respiratory therapists and allied health professionals dedicated to the prevention, detection, treatment, cure and research of respiratory disease and critical illness. Our members treat patients whose illnesses were caused or worsened by air pollution, including patients with lung disease like asthma, cystic fibrosis or COPD, and critical illnesses like pneumonia. Our members are also engaged in basic, human, clinical and epidemiological research studies on the health effects of air pollution. We have serious concerns about the EPA's proposed rule called "Strengthening Transparency in Regulatory Science," and what it means for our patients who are especially susceptible to the harmful effects of air pollution.

The proposed rule requires that the EPA make publicly available the underlying data from "pivotal regulatory science" studies that the agency relies on to establish major regulatory policy - including standards, exposure thresholds, and dose-response relationships. **Our major objection to this rule is that by excluding studies whose underlying data cannot be shared in a public database (e.g. due to study participant privacy concerns), this rule would effectively block the use of most epidemiological research studies from EPA rule-making.** Instead of promoting transparency in regulatory science, this rule would decrease transparency, by giving the EPA administrator unchecked authority to pick and choose which research studies will inform policies that affect the

health of the U.S. population. The ATS strongly favors transparency in research, and I will outline some of the ways in which the NIH and scientific community are promoting transparency and replication of research while protecting the privacy of research participants. Our key take-away points about this obstructionist and potentially harmful EPA rule are summarized below:

1. **Under the rule, EPA would disregard studies of how pollution affects the health of children and adults**

This rule introduces a barrier that will exclude from EPA consideration any studies whose underlying data cannot be shared with the public. While studies that expose laboratory animals to pollution may be able to meet this demand, many studies of how pollution exposure affects risk of death and disease among real people (i.e. epidemiological studies) will not be able to meet the demand of public data-sharing. Sharing data about diagnoses, hospitalization or death, and address information about the home, school and work locations of study participants is not always feasible, because the privacy of study participants must be protected. Before a health study of children or adults can even begin, investigators must complete a rigorous review by an Institutional Review Board to ensure that the risks to study participants, including risks to privacy, are minimized. As part of its review, the Institutional Review Board carefully scrutinizes the consent form that study participants sign before joining a study, to ensure the form details how participants' private data might be shared, and what safeguards will remain in place to protect their privacy after study completion. This new rule would prevent most research about health effects of pollution in the real world from informing EPA policy, because the underlying data about the participants of these health studies cannot be shared with the public.

Ignoring medical research in regulatory decision-making is the opposite of progress, and is not in the interest of human health. As a doctor, I would do my patients a disservice if I ignored the best available evidence to guide my decisions. Medical guidelines are based on the weight of the evidence, which emerges from multiple peer-reviewed scientific studies, not just one study. It would be malpractice for a doctor to apply such a "transparency" standard, as proposed for the EPA, to the care of patients, because it would involve ignoring large portions of the scientific literature. Such a standard would lead to misinformed treatment decisions, like offering drugs that have been found to be unsafe, and may deny patients the best treatments that modern medicine offers today. Patients would suffer if the medical community ignored scientific evidence to guide therapy. The same would be true if the EPA ignored evidence in making decisions about air quality and other environmental standards that affect the health of children and adults living in the United States.

2. **The proposed rule could jeopardize confidential information about study participants**

The proposed rule states that the EPA will apply the tools and methods developed by other Federal agencies to de-identify private information. The rule cites a Department of Health and Human Services (HHS) document as an example of how data can be de-identified to protect

confidential patient information.[1] However, the very HHS guidance that EPA references notes that de-identification does not fully protect patient information, stating:

"Both methods, even when properly applied, yield de-identified data that retains some risk of identification. Although the risk is very small, it is not zero, and there is a possibility that de-identified data could be linked back to the identity of the patient to which it corresponds."

I have included the full print out of the HHS guidance – with the above text highlighted – with my statement. In environmental health research, which often involves information about location, it may be especially easy to re-identify study participants. A recent (2017) study by Sweeney and colleagues took the HIPAA-compliant de-identified data from an air pollution epidemiology study, and using other publicly available data sets and commercially available computer programs, successfully re-identified 25% of study participants.[2] Even the Centers for Medicare and Medicaid Services requires researchers who use Medicare information to report only summary information across large numbers of people to ensure that individual people cannot be identified. A recent publication in *Nature Communications* strongly challenges the "de-identification release-and-forget" approach, finding that nearly 99.98% of Americans could be re-identified with just 15 demographic variables.[3]

EPA's proposal to make the underlying data of policy-relevant studies fully available to the public in de-identified form risks disseminating the sensitive information about health problems, deaths and addresses of study participants. The long-term consequences of such a data breach could be devastating, not only for the study participants whose private health information is leaked, but also for the future of environmental health research. Imagine how such a breach would affect anyone's willingness to participate in an environmental health research study in the future.

3. **Multiple mechanisms and safeguards promote research transparency and data-sharing**

The ATS supports efforts by the NIH to fund and promote transparency in health research, including environmental health research. Major funding sources including the National Institutes of Health (NIH) and the U.S. EPA require scientists to establish a data sharing plan as part of the scientific granting process. Major journals that publish research on how pollution affects health, including *Lancet*, *the Journal of the American Medical Association*, the *New England Journal of Medicine*, and others require researchers to specify a mechanism for sharing data as part of their manuscript submission. To be clear, these data sharing plans are intended to facilitate data sharing within the scientific community – i.e. from one scientist to another—to facilitate replication of findings, or pooling of data from multiple studies. The receiving scientist must demonstrate that he/she has skills, resources and safeguards to appropriately use and protect the data. When research data is shared, a data use agreement is usually signed by both institutions to guarantee those safeguards are in place.

In addition, there are multiple data repositories for NIH-funded research in which de-identified data is deposited under NIH policy. For these NIH-designed data repositories, the informed

consent signed by study participants when they joined the study determines if the data is appropriate for the NIH repository, and whether the data should be available through unrestricted (public) or controlled access (e.g. for scientists with safeguards in place).

The proposed EPA rule does not fund a mechanism for improving scientific transparency. Rather, it creates an obstructive mechanism (*a process barrier)* by which environmental health research, especially epidemiologic research that cannot be fully de-identified or publicly shared, can be excluded by EPA in its rulemaking.

4. **An independent EPA-funded resource already exists to resolve scientific disputes**

The Health Effects Institute (HEI) is a research group funded equally by the motor vehicle industry and the U.S. EPA that has played a key role in resolving disputes about pivotal environmental health research that informs EPA regulation. For example, in July of 2000, the HEI conducted a re-analysis of two early air pollution studies: the Harvard Six Cities Study and the American Cancer Society Study, on the link between particulate matter pollution and mortality. The re-analysis was conducted by a team of independent scientists, and overseen by a broad board of stakeholders, and confirmed the findings of the original studies.[4] In addition, to assess differences in industry versus academic/NCI analyses regarding risk estimates for diesel exhaust as a carcinogen, HEI held a public workshop and convened an independent panel of scientists. HEI then issued a report verifying the original findings in that scientific controversy, too.[5] These past interventions by HEI to independently verify the data analysis for studies on controversial scientific issues provide a generalizable model to address the challenge of privacy vs. transparency in evaluating scientific research directly relevant to the regulatory process. Using the HEI to vet such key results is a practical and proven approach to address concerns raised about transparency, without compromising health data privacy of study participants.

5. **The rule follows a familiar script for how industry can discredit sound scientific findings**

What EPA is proposing comes straight out of the playbook of the tobacco industry and its attempts to discredit research findings that link environmental tobacco smoke exposure to health problems. I have included an internal tobacco industry memo,[6] authored by tobacco lawyer Chris Horner in 1996, that describes in detail the steps that can be taken to discredit scientific information. These steps included the construction of “explicit procedural hurdles the Agency must follow” and to address “process as opposed to scientific substance.” The memo used the same terms of “transparency,” “sound science” and calls for “reproducible” science - the language that the EPA is now using in its proposed rule. The goal of this strategy, as described by Mr. Horner, was to help RJ Reynolds avoid having to “undo the (EPA) Agency’s work” “after-the-fact.” Mr. Horner served on the EPA transition team.

Soon after the date of that memo to RJ Reynolds, congressional efforts were underway to mandate the release of environmental health research data to the public. I have enclosed a manuscript by ATS member Dr. George Thurston,[7] published in 1998 and still as relevant

today as it was 20 years ago, in which he articulates the risks of public release of environmental health data, and provides historical examples of what can happen when vested interests analyze health data to achieve corporate aims. In one example, consultants for the RJ Reynolds Company used a Georgia state law to access raw research data to discredit study findings by Dr. Paul Fischer, which concluded that the use of cartoon characters (such as "Joe Camel") appealed to children. RJ Reynolds even went as far as to request the telephone numbers of children who participated in the study. While Dr. Fischer's research was later validated by others (and RJ Reynolds later admitted targeting children in advertising), Dr. Fischer abandoned his research career as a result of the attacks he endured. The EPA rule would deliver sensitive environmental health research data straight to the potentially misleading manipulations of vested interests. Special interest groups, who may not like the conclusions of health studies about health risks of pollution and chemicals, will be free to report their alternative findings without peer review, and without having to demonstrate they have the skills and resources to appropriately analyze and interpret the data in an unbiased manner.

6. **The proposed rule gives unchecked discretion to the EPA Administrator**.

The proposed rule includes a provision allowing the Administrator to "exempt significant regulatory decisions on a case-by-case basis." The ATS is concerned that delegating the discretionary authority solely to the Administrator would grant excessive authority to one person without accountability to the public. Allowing the Administrator to pick research in this way is secretive, and flies in the face of any transparent ethical process.

Conclusion

In summary, this rule does not improve on existing measures to enhance the transparency of environmental health research, and instead would function as a roadblock against the use of epidemiologic research in EPA rule-making. This misguided rule, if implemented, would lead EPA to make decisions based on incomplete information. Our patients with lung disease, and all Americans, depend on the EPA to make well-informed decisions—based on the best available evidence—to set environmental standards that protect their health.

On behalf of the ATS, I greatly appreciate the Committee's attention to this important scientific issue. I would strongly urge this Committee and Congress to ensure that EPA uses the weight of the evidence in its policy-setting decisions, and to prevent EPA from adopting process rules that block peer-reviewed research from being considered. I look forward to answering your questions.

Mary B. Rice, MD MPH
Chair, ATS Environmental Health Policy Committee

References

1. US Dept Health and Human Services: Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule. https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html#rationale.
2. Sweeney L, Yoo JS, Perovich L, Boronow KE, Brown P, Brody JG. Re-identification Risks in HIPAA Safe Harbor Data: A study of data from one environmental health study. *Technol Sci*. 2017;2017.
3. Rocher L, Hendrickx JM, de Montjoye Y-A. Estimating the success of re-identifications in incomplete datasets using generative models. *Nat Commun*. 2019;10(1):3069.
4. Reanalysis of the Harvard Six Cities Study and the American Cancer Society Study of Particulate Air Pollution and Mortality | Health Effects Institute. https://www.healtheffects.org/publication/reanalysis-harvard-six-cities-study-and-american-cancer-society-study-particulate-air.
5. Diesel Emissions and Lung Cancer: Epidemiology and Quantitative Risk Assessment | Health Effects Institute. https://www.healtheffects.org/publication/diesel-emissions-and-lung-cancer-epidemiology-and-quantitative-risk-assessment. Accessed November 9, 2019.
6. Truth Tobacco Industry Documents, RJ Reynolds memo: Background and Proposal to Address Federal Agency Science, Dec 23 1996. https://www.industrydocumentslibrary.ucsf.edu/tobacco/docs/#id=jhxk0020 (Accessed 8/7/18).
7. Thurston GD. Mandating the release of health research data: issues and implications. *Tulane Environ Law Journal*. 1998;11(2):331-354.

HHS.gov U.S. Department of Health & Human Services

Health Information Privacy

Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule

This page provides guidance about methods and approaches to achieve de-identification in accordance with the Health Insurance Portability and Accountability Act of 1996 (HIPAA) Privacy Rule. The guidance explains and answers questions regarding the two methods that can be used to satisfy the Privacy Rule's de-identification standard: Expert Determination and Safe Harbor[1]. This guidance is intended to assist covered entities to understand what is de-identification, the general process by which de-identified information is created, and the options available for performing de-identification.

In developing this guidance, the Office for Civil Rights (OCR) solicited input from stakeholders with practical, technical and policy experience in de-identification. OCR convened stakeholders at a workshop consisting of multiple panel sessions held March 8-9, 2010, in Washington, DC. Each panel addressed a specific topic related to the Privacy Rule's de-identification methodologies and policies. The workshop was open to the public and each panel was followed by a question and answer period. Read more on the Workshop on the HIPAA Privacy Rule's De-Identification Standard. Read the Full Guidance - PDF.

General

1.1 Protected Health Information
1.2 Covered Entities, Business Associates, and PHI
1.3 De-identification and its Rationale
1.4 The De-identification Standard
1.5 Preparation for De-identification

Guidance on Satisfying the Expert Determination Method

2.1 Have expert determinations been applied outside of the health field?
2.2 Who is an "expert?"
2.3 What is an acceptable level of identification risk for an expert determination?
2.4 How long is an expert determination valid for a given data set?
2.5 Can an expert derive multiple solutions from the same data set for a recipient?
2.6 How do experts assess the risk of identification of information?
2.7 What are the approaches by which an expert assesses the risk that health information can be identified?
2.8 What are the approaches by which an expert mitigates the risk of identification of an individual in health information?
2.9 Can an Expert determine a code derived from PHI is de-identified?
2.10 Must a covered entity use a data use agreement when sharing de-identified data to satisfy the Expert Determination Method?

Guidance on Satisfying the Safe Harbor Method

3.1 When can ZIP codes be included in de-identified information?
3.2 May parts or derivatives of any of the listed identifiers be disclosed consistent with the Safe Harbor Method?
3.3 What are examples of dates that are not permitted according to the Safe Harbor Method?
3.4 Can dates associated with test measures for a patient be reported in accordance with Safe Harbor?
3.5 What constitutes "any other unique identifying number, characteristic, or code" with respect to the Safe Harbor method of the Privacy Rule?
3.6 What is "actual knowledge" that the remaining information could be used either alone or in combination with other information to identify an individual who is a subject of the information?
3.7 If a covered entity knows of specific studies about methods to re-identify health information or use de-identified health information alone or in combination with other information to identify an individual, does this necessarily mean a covered entity has actual knowledge under the Safe Harbor method?
3.8 Must a covered entity suppress all personal names, such as physician names, from health information for it to be designated as de-identified?
3.9 Must a covered entity use a data use agreement when sharing de-identified data to satisfy the Safe Harbor Method?
3.10 Must a covered entity remove protected health information from free text fields to satisfy the Safe Harbor Method?

Glossary of Terms

Protected Health Information

The HIPAA Privacy Rule protects most "individually identifiable health information" held or transmitted by a covered entity or its business associate, in any form or medium, whether electronic, on paper, or oral. The Privacy Rule calls this information *protected health information* (PHI)[2]. Protected health information is

information, including demographic information, which relates to:

- the individual's past, present, or future physical or mental health or condition,
- the provision of health care to the individual, or
- the past, present, or future payment for the provision of health care to the individual, and that identifies the individual or for which there is a reasonable basis to believe can be used to identify the individual. Protected health information includes many common identifiers (e.g., name, address, birth date, Social Security Number) when they can be associated with the health information listed above.

For example, a medical record, laboratory report, or hospital bill would be PHI because each document would contain a patient's name and/or other identifying information associated with the health data content.

By contrast, a health plan report that only noted the average age of health plan members was 45 years would not be PHI because that information, although developed by aggregating information from individual plan member records, does not identify any individual plan members and there is no reasonable basis to believe that it could be used to identify an individual.

The relationship with health information is fundamental. Identifying information alone, such as personal names, residential addresses, or phone numbers, would not necessarily be designated as PHI. For instance, if such information was reported as part of a publicly accessible data source, such as a phone book, then this information would not be PHI because it is not related to heath data (see above). If such information was listed with health condition, health care provision or payment data, such as an indication that the individual was treated at a certain clinic, then this information would be PHI.

Back to top

Covered Entities, Business Associates, and PHI

In general, the protections of the Privacy Rule apply to information held by covered entities and their business associates. HIPAA defines a covered entity as 1) a health care provider that conducts certain standard administrative and financial transactions in electronic form; 2) a health care clearinghouse; or 3) a health plan.[3] A business associate is a person or entity (other than a member of the covered entity's workforce) that performs certain functions or activities on behalf of, or provides certain services to, a covered entity that involve the use or disclosure of protected health information. A covered entity may use a business associate to de-identify PHI on its behalf only to the extent such activity is authorized by their business associate agreement.

See the OCR website http://www.hhs.gov/ocr/privacy/ for detailed information about the Privacy Rule and how it protects the privacy of health information.

Back to top

De-identification and its Rationale

The increasing adoption of health information technologies in the United States accelerates their potential to facilitate beneficial studies that combine large, complex data sets from multiple sources. The process of de-identification, by which identifiers are removed from the health information, mitigates privacy risks to individuals and thereby supports the secondary use of data for comparative effectiveness studies, policy assessment, life sciences research, and other endeavors.

The Privacy Rule was designed to protect individually identifiable health information through permitting only certain uses and disclosures of PHI provided by the Rule, or as authorized by the individual subject of the information. However, in recognition of the potential utility of health information even when it is not individually identifiable, §164.502(d) of the Privacy Rule permits a covered entity or its business associate to create information that is not individually identifiable by following the de-identification standard and implementation specifications in §164.514(a)-(b). These provisions allow the entity to use and disclose information that neither identifies nor provides a reasonable basis to identify an individual.[4] As discussed below, the Privacy Rule provides two de-identification methods: 1) a formal determination by a qualified expert; or 2) the removal of specified individual identifiers as well as absence of actual knowledge by the covered entity that the remaining information could be used alone or in combination with other information to identify the individual.

Both methods, even when properly applied, yield de-identified data that retains some risk of identification. Although the risk is very small, it is not zero, and there is a possibility that de-identified data could be linked back to the identity of the patient to which it corresponds.

Regardless of the method by which de-identification is achieved, the Privacy Rule does not restrict the use or disclosure of de-identified health information, as it is no longer considered protected health information.

Back to top

The De-identification Standard

Section 164.514(a) of the HIPAA Privacy Rule provides the standard for de-identification of protected health information. Under this standard, health information is not individually identifiable if it does not identify an individual and if the covered entity has no reasonable basis to believe it can be used to identify an individual.

§ 164.514 Other requirements relating to uses and disclosures of protected health information.
(a) *Standard: de-identification of protected health information.* Health information that does not identify an individual and with respect to which there is no reasonable basis to believe that the information can be used to identify an individual is not individually identifiable health information.

From: Christopher Horner
To: Hyde, Timothy N.; Tompson, Randy
CC:
BCC:
Subject: Federal Agency Science
Date: 12/23/1996 1:56:01 PM

Attachments:

Gentlemen: The following is the document we discussed. Have a happy holiday. CCH
M E M O R A N D U M

TO: Mr. Tim Hyde
Mr. Randy Johnson
RJ Reynolds Tobacco Company

FROM: Mr. Christopher C. Horner
Bracewell & Patterson, L.L.P.

DATE: December 23, 1996

RE: Background and Proposed Program to Address Federal Agency Science

Per our earlier conversations, the following sets forth what needs to be done to reform agency science, focusing on the need based upon your interests, and how you are positioned to take a behind the scenes leadership position. It provides an overview of the issues relevant to this goal, and details a program taking advantage of the increasingly flagrant way regulators have perverted the scientific process, hiding behind a wall of selected scientists to essentially cow industry and Congress into accepting fringe scientific conclusions.

Summary

We propose creating, beginning with congressional oversight and a goal of enacting legislation, required review procedures which EPA and other federal agencies must follow in developing "extra-judicial" documents (i.e., those documents produced as guidance, science or other government products issued by regulatory agencies which are not necessarily at time of publication ripe for judicial review). This is important to your organization because, at some point in the near future, EPA will most likely be ordered to re-examine ETS. The only way to do

so on a level playing field is to construct explicit procedural hurdles the Agency must follow in issuing scientific reports.

Because there is virtually no chance of affecting change on this issue if the focus is ETS, our approach is one of addressing process as opposed to scientific substance, and global applicability to industry rather than focusing on any single industrial sector. Thus the examples of questionable science, to justify these standards. Congress must require those examples serve as the test cases.

Background

On the surface, now appears an opportune moment for addressing agency science head on, tackling the substance. This would seem the case because the first run at legislative attempts to reform the regulatory process failed and concerned Members are searching for a new mechanism to control EPA and other regulatory bodies. The landscape of the past year is littered with examples of persistent or newly-promoted "bad science," including the Mercury Report to Congress, MACT Hazardous Waste Combustion Rule, Methylene Chloride and the Dioxin Reassessment. Regarding the latter example, as you are likely aware, for the next round of EPA Science Advisory Board (SAB) review of the Dioxin Reassessment the Agency has removed any SAB members who were too vocal in their disagreement with the Agency. There will still be SAB review, but it will be an already-transparent group of "agreeable" scientists. So, in addition EPA is flagrantly "stacking the deck" with those whose conclusions are predetermined and in the Agency's favor.

Irrespective of this pattern, it is clear the 104th Congress was singularly unsuccessful in managing the Agency on a chemical-by-chemical or industry-by-industry basis. EPA actions demonstrate the it has taken measure of its legislative and industry adversaries, and decided upon aggressive campaigns on several of these issues to impose its policy-driven will upon scientific conclusions. The Agency helps create, and responds, to, the political winds, so you should anticipate no relief on re-evaluating ETS. EPA has of late played its public relations card very well, avoiding long news cycles for its proposals -- even timing them around holidays when readership is at its nadir -- while engaging the environmental press for the coming conflicts. EPA, helped by the backlash of the generally "pro-environment" public to a poorly implemented reg-reform agenda, has fostered an atmosphere where "industry" are reluctant to match the Agency's hardball tactics out of fear either that Congress would duck/mismanage the issue, or of Agency retribution. Thus, through a lack of industry support and unfavorable press, Congress has to date lacked the requisite support to effectively use the oversight powers of the legislative branch.

It is in this climate you will face a chastened but at least as aggressive EPA on re-evaluating the ETS study.

Project Approach

To improve the climate, and process, under which ETS and others are reviewed, we recommend initiating reforms by playing a strong role in molding and guiding Congress's oversight of EPA's latest Clean Air Act initiative (on PM 2.5/ozone). Such an effort would work toward requiring EPA to institute certain procedural changes to the pre-regulatory process. These would serve as a set of checks and balances to ensure a fair and equitable development and publication of scientific findings (i.e., reform the scientific process). It is that process, which is beyond the reach of the Administrative Procedure Act, which sets the stage for the rulemaking process. These procedures could then be subject to judicial review without the courts becoming involved in specific scientific issues (i.e., discern if EPA followed the requisite steps, rather than if it achieved the "right" answer).

When EPA announced its proposal to regulate particulate matter and tropospheric ozone, despite their news cycle management, the set the predicate for procedural change. These proposed regulations, based on questionable science, are not focused on those industries that comprise EPA's "usual suspects", but rather all industries including small businesses. Congress is expected to conduct heavy oversight of this process, with most leaders expressing that the actions are unnecessary and unrealistic. EPA has already signaled a desire to compromise as the process moves forward, and will start airing its options in the January 14-15 initial public hearings. It is critical to our overall goal that EPA not be allowed to change the forum into an industry-by-industry examination. Equally important, the process should not devolve into "outdoor air" interest seeking to shift the focus to "indoor air" interests. Instead, the efforts we envision focus on the process by which EPA arrived at its scientific conclusions, avoiding to the extent possible specific scientific issues, contaminants, or industries.

While some will approach these hearings as regulation-specific, as you can appreciate, from our perspective the greater problem is EPA (and OSHA) "science," encompassing all the scientific reports, studies, guidance documents and procedures produced by the nonregulatory offices of these agencies. None of these products are subject to timely challenge. In some instances, industry must wait years before regulations are promulgated, thus allowing industry to sue. Then, when industry has that opportunity, the court is faced with the ramifications of overturning years of EPA actions and policies based on this scientific document. Moreover, industry face mindsets such as "how can a

document which has been around for so long be wrong?" (the "historical credibility" argument). Finally, once industry's hands are tied in Washington, EPA or OSHA has distributed the documents or guidance to the press or states, forcing industry to face a public relations nightmare.

Thus, as we seek to create a regime where this cycle is a thing of the past while highlighting problems with contemporary studies. These studies will be the first "test cases" for the reformed process. This requires developing (1) overall criteria for a "sound science" process, and (2) a record, through congressional oversight, on how the Agency typically does not meet those criteria.

To illustrate, criteria could be as follows:

"Sound Science" Criteria - any government scientific program must have four components:

Inclusive - The scientific community, the public, Congress, and other Executive Branch agencies are given fair and timely access to review and affect change in the development of the science/document.

Transparency - the public can follow the developmental process the steps followed to develop the final science/document.

Able to be reproduced - Can the answer be reproduced from the record?

Algorithm - Given the set of all available scientific knowledge on the subject would independent groups arrive at the same answer?

[a possible fifth component which could be included as a deal closer could be:
Not judicially reviewable - This may seem counterintuitive, but one of the aspects of reg-reform which its opponents exploited to bring it down was the belief that everything would be litigated. Thus, it may be possible to achieve reforms through the principle that the scientific portions of a successful program should not be easily placed before the courts. Instead, the courts should be able to easily look at procedures followed (e.g., did the Agency follow its own procedures).]

We envision these new steps being "field tested" on, e.g., the methylene chloride study, ETS, etc. which, having been used as justification for reform would be held and reviewed under the new procedures.
To ensure Agency compliance Congressional oversight is also required. This at worst builds a record for judicial review and at best sets in

motion a set of enforceable procedures. We intend to develop for the Hill a set of scientific and procedural questions on scientific issues which different committees could then use. This requires:

Written Record - Submit lengthy, detailed questions to the agency requiring written responses. This creates a written record which the Agency often seeks to avoid, because it otherwise is permitted to develop scientific documents without responding explicitly (unlike the proposal/promulgation process) to public concerns.

Followup Hearings - Once the Agency has responded use this record both within and across an issue in oversight on how the Agency develops science. (e.g., this is an ideal place to inquire into risk assessment default values and risk criteria, which seem to change from office to office).

We envision the end results of the oversight hearings to be: (1) EPA publication in the Federal Register of a formal process for handling "extra-judicial" documents; (2) new legislation; and/or (3) inclusion in environmental or regulatory reform legislation which appears moving in the 105th Congress.

This approach merely ensures a fair hearing, but that is typically all the situations require to avoid the skewed result the federal agency prescribes. Critically, this approach also circumvents the tenuous situation you otherwise likely will face, of seeking after-the-fact, RJR-specific congressional support to undo the Agency's work.

What makes the National Association of Manufacturers a strong base for the above work is NAM's broad, yet non-specific, business base. Its one of a small handful, at best, of broad based associations not associated with particular industries. Thus, their lead on this general issue will not bog the hearings down in "anti-environmental," industry-specific rhetoric, nor create an environment where specific industries can legitimately fear Agency retaliation.

Conclusion

We envision a program, using contemporary studies and reports to illustrate how the Agency skews its results in the pre-regulation stage, to create set, reviewable science procedures. That process and its criteria will first be tested on those current examples of Agency misfeasance, which obviously must be sent back to the Agency or otherwise placed on hold in the interim. We need to meet again with you to discuss this proposal and how to best implement it, specifically beginning with the audiences with NAM and NFIB we discussed. We need another meeting, to hammer out the presentation to the two

referenced audiences, and reach consensus with you on the issues and approach we intend to pursue. Until we speak with you on this further, Happy Holidays.

CCH

/cch

Tulane
Environmental
Law Journal

MANDATING THE RELEASE OF HEALTH RESEARCH DATA: ISSUES AND IMPLICATIONS

GEORGE D. THURSTON

Mandating the Release of Health Research Data: Issues and Implications

George D. Thurston*

I. INTRODUCTION 331
II. THE CASE AT HAND: AIR POLLUTION EPIDEMIOLOGY 334
III. ISSUES AND IMPLICATIONS 340
A. *Research Credibility* 340
B. *Confidentiality of Participant Medical Records* 345
C. *Intellectual Ownership Rights* 347
D. *Effect on Scientific Progress* 348
E. *Unfunded Mandates* 349
F. *Are Existing Mechanisms Sufficient?* 349
IV. DISCUSSION AND RECOMMENDATIONS 352

I. INTRODUCTION

"Show me the data!" sounds a lot like a soundbite from a Hollywood movie, but it accurately characterizes the demands that U.S. industry representatives and legislators on Capitol Hill have in recent years been making of researchers who study environmental and occupational health problems. Indeed, in July 1997, an amendment to a 1998 appropriations bill in the U.S. House of Representatives was proposed that, if passed, would have required researchers with government grants to make their raw medical and scientific data publicly available within ninety days after the first public reporting of any study results.[1] No hearings were held on the implications of such a step. Only defense-related research and cases in which "adverse economic harm to

---

* © 1998 George D. Thurston. Associate Professor of Environmental Medicine; Director, Community Outreach; Institute of Environmental Medicine, New York University School of Medicine. Dr. Thurston testified before both the U.S. Senate and the U.S. House of Representatives in 1997, regarding air quality regulations in the United States. He serves as an advisor to the State of New York on air quality matters, having served on the Department of Environmental Conservation's Air Management Advisory Committee since its inception in 1991. He presently serves on the National Academy of Science's Committee on Health Effects of Incineration, and was Chairman of Canada's Health and Environment Panel for the Health Canada Sulfur in Gasoline Study in 1997. A.B., Sc.B. Envtl. Engineering 1974, Brown University; M.S. Envtl. Health Science 1978, Harvard University; Sc.D. Envtl. Health Science 1983, Harvard University.

1. *See* Jocelyn Kaiser, *Data Release Proposal Raises Alarm*, 277 SCIENCE 627 (1997).

commercial proprietary interests ... would result" would have been exempted from this blanket data release mandate.[2] Is this proposal to mandate a blanket release of federally funded research data a necessary and worthwhile solution to a real problem that is impeding the advance of scientific knowledge? Or, alternatively, are the cries for Congress to take such an action merely a manifestation of vested interests' attempts at undermining the credibility of researchers who publish results that run counter to their financial interests?

According to the journal *Science*, the data release amendment proposed by Representative Robert B. Aderholt (R-AL) was, in part, a response by Congress to industry demands for data from Harvard School of Public Health air pollution studies, the results of which were at the center of proposed new air pollution regulations.[3] The studies' authors objected to making their raw research data publicly available because it would violate the crucial confidentiality agreements they had made with study subjects to protect their individual privacy.[4] Although these Harvard researchers were willing to share the data with other scientists when that confidentiality could be protected, they were not willing to capitulate to unrestricted release of the personal health records.[5] In the end, this particular congressional amendment was defeated by a vote of nineteen to thirty-four.[6] Discussion of such a measure, however, will no doubt surface on the Hill again in the near future, as demands for congressional action are likely to continue due to other regulatory measures being questioned by industry. It is therefore important to air both the issues involved in, and the implications of, such a mandate for the release of federally funded health effects research data.

At first glance, this proposal may seem to be a simple and straightforward idea. The basic logic behind the proposal, apparently, was that the data collection was paid for, at least in part, by the government, therefore it should be available to the public and to anyone else who wishes to evaluate it further. In a cover letter to his colleagues in the House, the sponsor of the amendment stated that "the federal government does not have a standardized government-wide process for making research data available for independent review. My amendment seeks to remedy this while still allowing for a limited number of

2. *Amendment to Treasury, Postal Service, and General Government Appropriations Bill, 1998* (offered by Rep. Robert B. Aderholt (R-AL), July 25, 1997) (rejected July 31, 1997) [hereinafter *Amendment to Treasury Bill*].

3. *See* Jocelyn Kaiser, *Government Grants: Academia Wins a Round on Raw Data*, 277 SCIENCE 758 (1997).

4. *See id.*

5. *See id.*

6. *See* H.R. REP. No. 105-240 (1997).

exceptions .... I strongly believe that sunshine is the best antiseptic."[7] In addition, the argument has been made that government regulations that are based upon federally funded health research might cost billions of dollars to affected businesses and industries.[8] Accordingly, it is important to make doubly sure that the research is right. Thus, there were some seemingly plausible rationales for such a measure, however, practicality and ethical concerns quickly arose.

In the days that followed the congressional proposal, numerous confidentiality, logistical, and fairness objections came to light from other legislators, the Clinton Administration, and the nation's research universities.[9] Representative George E. Brown, Jr. (D-CA), ranking minority member of the House's Science Committee, expressed his "deep concern" that "the amendment as drafted would create significant legal uncertainties and substantial and unnecessary costs for scientists, research universities, high tech industries, and federal agencies."[10] In addition, the White House Office of Management and Budget enumerated potential problems, including the impeding of commercial agreements and the risk of problems if the data were not analyzed correctly by others unfamiliar with the data collection process.[11]

This Article provides a detailed consideration of the ongoing data access debate in the context of the United States Environmental Protection Agency's (EPA or the Agency) recent air pollution regulations and the research upon which they are based, followed by a discussion of the key issues surrounding the data access debate in general. These key issues include the potential effects of a mandate requiring the release of health research data on: (1) the scientific credibility of the research involved, (2) the confidentiality of research participants' medical records, (3) the intellectual ownership of research ideas and their results, and (4) the speed of research progress in the medical and public health fields. Information from past cases of data release demands and their aftermath are supplied as examples. Consideration is then given to whether there are sufficient deficiencies in the current practices of scientific assessment and data sharing that warrant such government mandated intervention

7. Letter from Rep. Robert B. Aderholt (R-AL) to the House of Representatives (July 24, 1997) (on file with author).

8. *See* Madeleine Jacobs, *EPA Should Take a Breather*, CHEM. & ENG. NEWS, Apr. 14, 1997, at 5; *see also* Philip H. Abelson, *Proposed Air Pollutant Standards*, 277 SCIENCE 15 (1997); Jocelyn Kaiser, *Showdown Over Clean Air Standards*, 277 SCIENCE 466 (1997).

9. *See* Kaiser, *supra* note 3, at 758.

10. Letter from Rep. George E. Brown, Jr. (D-CA), Ranking Minority Member, House of Representatives Science Committee, to Rep. Bob Livingston (R-LA), Chairman, House Committee on Appropriations (July 28, 1997) (on file with author) [hereinafter Brown Letter].

11. *See* Kaiser, *supra* note 3, at 758.

into medical and public health research, or whether the side-effects of this proposed solution are worse than the initially perceived problems. Finally, alternative approaches to addressing the question of the validity of published scientific research are also proposed.

II. THE CASE AT HAND: AIR POLLUTION EPIDEMIOLOGY

The National Ambient Air Quality Standards (NAAQS), the cornerstone of the nation's air pollution control program, are aimed at establishing air quality requirements sufficient to protect public health and welfare.[12] The Clean Air Act (CAA)[13] and its Amendments[14] require that these national air quality standards be set at a level stringent enough to protect the health of the public, with an adequate margin of safety.[15] The CAA Amendment of 1977, as adopted by Congress, requires that each of the NAAQS be reviewed by the EPA at least every five years in order to determine whether the NAAQS are still appropriately protective of public health and welfare based on the most recent research information.[16] Revisions of the NAAQS by the EPA Administrator are based upon scientific air quality criteria documents that are prepared by the EPA for the air pollutant under review and subsequently reviewed by an independent scientific advisory panel, the Clean Air Scientific Advisory Committee (CASAC).[17]

In 1979, upon review of the nation's photochemical oxidants standard, the EPA relaxed the ozone (O_3) NAAQS from a once-per-year, one-hour maximum of 80 parts per billion (ppb) up to 120 ppb, due to a lack of published information supporting the then existing standard.[18] Ozone is a secondary pollutant, or one that is formed in the atmosphere in the presence of sunlight from precursor pollutants, most notably nitrogen oxides and hydrocarbons that are emitted by a variety of sources, including automobiles, coal-fired power plants, and industry.[19] This

12. *See* Clean Air Act § 109, 42 U.S.C. § 7409 (1994).
13. *See* CAA §§ 101-618, 42 U.S.C. §§ 7401-7671*q*.
14. *See* Clean Air Act Amendments of 1966, Pub. L. 89-675, 80 Stat. 954; Clean Air Amendments of 1970, Pub. L. No. 91-604, 84 Stat. 1676; Clean Air Act Amendments of 1977, Pub. L. No. 95-95, 91 Stat. 685; Clean Air Act Amendments of 1990, Pub. L. No. 101-549, 104 Stat. 2399 (codified as amended in scattered sections of 42 U.S.C.).
15. *See* CAA § 109(d), 42 U.S.C. § 7409(d).
16. *See* CAA Amendments of 1977, Pub. L. No. 95-95, 91 Stat. 385 (codified as amended in scattered sections of 42 U.S.C.).
17. *See* CAA § 109(d), 42 U.S.C. § 7409(d).
18. *See* National Ambient Air Quality Standards for Ozone, 44 Fed. Reg. 8201, 8204 (1979).
19. *See Review of EPA's Proposed Ozone and Particulate Matter NAAQS Revisions—Part 2 Before the Subcomm. on Health and Environment and the Subcomm. on Oversight and Investigations of the House Comm. on Commerce*, 105th Cong. 146 (1997) (prepared statement

standard remained in effect until 1997, when the EPA, after a long and extensive review of both new epidemiological and controlled pollutant exposure health studies, determined that the ozone NAAQS should be tightened back to a value of 80 ppb, but averaged over eight hours and allowing as many as three violations per year.[20] This new standard is therefore less protective than the once-per-year 80 ppb one-hour maximum standard in effect before 1979, but somewhat more protective than the pre-1997, 120 ppb one-hour maximum standard.

In 1997, the EPA also determined, after a similar extensive scientific review process, that the particulate matter (PM) NAAQS should also be modified to better protect the public health.[21] Fine PM (i.e., small particulate matter) is primarily composed of two components: carbonaceous primary particles, or soot, emitted directly from combustion sources such as diesel buses, coal and oil-fired power plants, and other industries; and, secondary particles formed in the atmosphere from gaseous pollutants such as sulfur dioxide and nitrogen oxides emitted from sources such as coal-fired power plants, automobiles, and industry.[22] In the case of PM, it was decided that a new standard was needed which focused on fine particles less than 2.5 micrometers in diameter ($PM_{2.5}$), which are particles small enough to reach deep into the human lung and most likely to have the highest concentrations of especially toxic PM components (e.g., acids, lead, arsenic, etc.).[23]

The implementation of these new air quality standards will require various businesses and industries to control their companies' air pollution emissions of gases and particles that some fear may cost large sums of money. This fear has caused those potentially affected parties to scrutinize the new standards intensely, and many of them have collectively or individually objected to the standards.[24] Partially in response to these industry concerns, Congress held numerous hearings on the new standards, including the consideration of bills to block the new standards.[25] However, no Congressional action has been taken to date to reverse the new air quality standards.

of Dr. George D. Thurston, Assoc. Prof., Dep't of Envtl. Medicine, New York University School of Medicine) [hereinafter *Thurston Testimony*].
20. *See* National Ambient Air Quality Standards for Ozone, 62 Fed. Reg. 38,856, 38,856-96 (1997) (to be codified at 40 C.F.R. pt. 50).
21. *See* National Ambient Air Quality Standards for Particulate Matter, 62 Fed. Reg. 38,652, 38,652-711 (1997) (to be codified at 40 C.F.R. pt. 51).
22. *See Thurston Testimony*, *supra* note 19, at 146.
23. *See Clean Air Act: Ozone and Particulate Matter Standards, Hearings Before the Subcomm. on Clean Air, Wetlands, Private Property, and Nuclear Safety and the Comm. of Environment and Public Works*, 105th Cong. 138 (1997) [hereinafter *Hearings*].
24. *See* Air Quality Standard Coalition, News Release (Nov. 25, 1996).
25. *See* S. 1084, 105th Cong. (1997); H.R. 1984, 105th Cong. (1997).

During the period when the EPA developed the new standards, demands surfaced for the release of the underlying health and scientific data upon which the key epidemiological pollution-health effects studies were based to set the new standards. In particular, in May 1994, Dr. George T. Wolff, a scientist for General Motors and the Chair of CASAC at that time, and Dr. Roger O. McClellan, the President of the Chemical Industry Institute of Toxicology (CIIT) and a former chair of CASAC, sent a letter to EPA Administrator Carol Browner asking that the EPA make demands for data and for data reanalyses.[26] However, neither Wolff nor McClellan indicated any scientific wrong-doing on the part of investigators in their letter.[27] Although this issue and the sending of a letter were discussed at a CASAC meeting, the Wolff and McClellan letter was not sent as a result of a consensus of the entire CASAC panel that the EPA should request such data, but at the initiative of these two specific CASAC panel members.[28] During congressional hearings in early 1997, these two scientists testified in opposition to the EPA's proposed $PM_{2.5}$ standard, with Wolff stating that "I can't endorse the present proposal,"[29] and McClellan stating that "the serious shortcomings in the scientific data on $PM_{2.5}$ and PM_{10} led me to not support the promulgation of either an annual or a 24-hour $PM_{2.5}$ standard."[30] The Wolff and McClellan letter to the EPA stated that:

> several recent published reports have indicated effects on both morbidity and mortality at about the level of the current PM_{10} standard. In some cases, the analyses are extremely complex because of the need to correct a wide range of potential confounders, such as temperature, cigarette smoking and other pollutants .... It is crucial that two or more groups analyze the same key data sets linking exposure and morbidity/mortality response to verify the adequacy of the complex analyses and that different analysts using the same data reach similar conclusions .... The EPA should take the lead in requesting that investigators make available the primary data sets being analyzed so that others can validate the analyses.[31]

In 1997, the Air Quality Standards Coalition (AQSC), in a submission to the EPA during the proposed O_3 and $PM_{2.5}$ NAAQS

26. *See* Letter from George T. Wolff, Ph.D., Chair, Clean Air Scientific Advisory Committee, and Roger O. McClellan, D.V.M., Past Chair, Clean Air Scientific Advisory Committee, to Carol Browner, Administrator, EPA (May 16, 1994) (on file with author) [hereinafter Wolff Letter].

27. *See id.*

28. EPA SCIENCE ADVISORY BOARD, CLEAN AIR SCIENTIFIC ADVISORY COMMITTEE, SO_2 CLOSURE & PARTICULATE MATTER, TRANSCRIPT OF PROCEEDINGS 169-70 (Apr. 12, 1994) (on file with author).

29. *Hearings*, *supra* note 23, at 38 (testimony of Dr. George T. Wolff).

30. *Id.* at 139 (testimony of Dr. Roger O. McClellan).

31. Wolff Letter, *supra* note 26.

comment period, cited the letter from Drs. Wolff and McClellan as a basis for requesting the data from the key Harvard "Six Cities" Studies. The AQSC submission requested that the studies be made "available in the rulemaking docket for assessment by other investigators and request EPA to reaffirm the existing PM standards until such time that these assessments are completed."[32] In its literature, the AQSC describes itself as "a broad-based coalition whose membership includes more than 500 corporations, associations and interest groups," whose goal is "to assure that the ... [EPA] makes scientifically ... sound decisions as it reviews the National Ambient Air Quality Standards for ozone and particulate matter."[33] However, the AQSC is described elsewhere as "a group of oil, steel, trucking, agricultural and auto companies, formed last July [1996] to fight the EPA's newly proposed air quality standards."[34] Thus, among the members of the AQSC are auto manufacturers, an industry group that includes General Motors, as well as oil companies and chemical manufacturers, two sectors well represented in the list of companies supporting the CIIT.[35] Indeed, CIIT's financial supporters include the Chemical Manufacturers Association, Chevron Corporation, Ethyl Corporation, Exxon Corporation, Texaco Inc., and Unocal Corporation.[36] In addition, the Mobil Corporation ran advertisements on the editorial pages of U.S. newspapers critical of the EPA proposal, including one ad stating that "data from a key study—the Harvard 'Six Cities'—has never been made public, despite repeated requests from scientists over a three-year period."[37] Thus, the most pointed demands for these studies' data have most often come from individuals and organizations either directly or indirectly supported by companies expected to be adversely affected by the new air standards based on those studies.

These recent demands for data release and reanalysis of the Harvard work have largely ignored the fact that these same Harvard researchers and their data have previously been reviewed for scientific integrity by the National Institutes of Health (NIH), Office of Scientific Integrity

32. Letter from Charles J. DiBona, President, American Petroleum Institute, and Andrew Card, President, American Automobile Manufacturers Association, Air Quality Standards Coalition Co-Chairs, to EPA Air Docket Section, Docket No. A-95-54(PM) (Mar. 12, 1997) <http://www.nam.org/bulletin/RER/rercmt.html>.

33. Air Quality Standard Coalition, News Release (Nov. 25, 1996).

34. Hanna Rosin, *Shades of Gray: St. Boyden's Unholy Money Trail*, 216 THE NEW REPUBLIC 21 (1997).

35. *See* National Association of Manufacturers Air Quality Standards Coalition (1998) (visited Apr. 29, 1998) <http://www.nam.org/air.html>.

36. *See* Chemical Industry Institute of Toxicology Supporting Companies (1998) (visited Apr. 29, 1998) <http://www.ciit.org/SUPPC/suppc.html>.

37. Mobil Corporation, *EPA: Let Science Do Its Job* (visited May 15, 1997) <http://www.mobil.com/this/news/opeds/970515_oped.html>.

(OSI), and the Health Effects Institute (HEI) and were cleared of any misconduct or scientifically inappropriate analyses.[38] The OSI investigation, which was the result of separate accusations raised in the mid-1980s, found that "there is no basis whatsoever for the allegations of serious errors and gaps in the database," and that "the quality control program of the Six Cities Studies considerably surpasses that of most continuously operating monitoring programs."[39] Furthermore, the HEI, which receives one-half of its fiscal support from the automotive industry and one-half from the U.S. government, subsequently commissioned an extensive reevaluation of the data and research methods of the Harvard team in conducting time-series analyses of various U.S. cities' daily records of mortality and PM pollution.[40] The HEI review found that the reanalysis results "agree closely with the earlier conclusions that particulate air pollution is tied to increased risk of death, even when weather and other pollutants are taken into account."[41] Thus, the Harvard researchers have in fact provided their data for evaluations in the past, and these previous evaluations have consistently confirmed the validity of their data and analytical methods.

However, in response to the continuing demands for the Harvard researchers' air pollution studies' data, Mary Nichols, then the EPA's Assistant Administrator for Air and Radiation, sent letters to Drs. Joel Schwartz and Douglas Dockery of the Harvard School of Public Health, as well as to Dr. Arden Pope, the lead author of another key PM study, stating that:

> there has been considerable interest in your research on the health effects of air pollution, including requests by members of Congress, governors of several states, and others for the raw data underlying your published research .... EPA is confident of the scientific integrity of your studies and their appropriateness for purposes of consideration in the Agency's present rulemaking on particulate matter without a separate or additional review of the underlying data. Nevertheless, given the strong interest in your research, EPA would encourage reasonable accommodations within the scientific and governmental community that would permit other interested scientists and agencies to understand fully the basis for your work. We therefore request that you make data associated with your published studies available to interested parties as rapidly as possible.[42]

38. Letter from Suzanne W. Hadley, Ph.D., Deputy Director, Office of Scientific Integrity, Department of Health & Human Services, to Michael W. Roberts, Esq., Office of the General Counsel, Harvard University (Nov. 15, 1990) (on file with author) [hereinafter Hadley Letter].

39. *Id.; see also NIH Clears Harvard Researchers*, WALL ST. J., Dec. 7, 1990, at A3.

40. Letter from James H. Ware, Dean for Academic Affairs, Harvard School of Public Health, to Mary D. Nichols, Assistant Administrator for Air and Radiation, EPA, (Apr. 8, 1997) (on file with author) [hereinafter Ware Letter].

41. Pamela Zurst, *Air Pollution Link to Rise in Deaths Confirmed*, CHEM. & ENG. NEWS Apr. 8, 1997, at 9.

Dr. James H. Ware, the Dean of the Harvard School of Public Health, subsequently recommended to Ms. Nichols that the Harvard "Six Cities" data be reviewed and tested by the HEI. Dr. Ware wrote "[w]e believe that HEI is well qualified to conduct a review process that will be thorough and fair, without jeopardizing confidentiality concerns."[43] This review is presently in progress. Thus, in this case, the concerns raised by industry and industry-funded groups concerning the results of this research are being addressed, without the need for a public release of the research health data.

In promulgating the new $PM_{2.5}$ air quality standards in the Federal Register, the EPA summarized the comments that it received during the NAAQS comment period regarding the issue of raw data availability.

> Several commenters questioned EPA's ability to rely on studies demonstrating an association between PM and excess mortality without obtaining and disclosing the raw "data" underlying these studies for public review and comment. In particular, a number of commenters cited Dockery, D.W., et al. 1993 and Pope, C.A. III, et al., 1995, as studies upon which EPA relied without obtaining and disclosing the underlying raw data .... A few commenters argued that section 307(d) of the [Clean Air] Act requires that EPA obtain the raw data underlying these studies and that a failure to do so contradicts the plain language of section 307(d)(3) of the Act, which requires EPA to place in the docket any "factual data on which the proposed rule is based." Other commenters argued that under section 307(d)(8) of the Act, a failure to obtain and disclose the underlying raw data used in the studies would constitute an error "so serious and related to matters of such central relevance to the rule that there is a substantial likelihood that the rule would have been significantly changed if such errors had not been made." According to one commenter, without the raw data and an opportunity for an analysis of it, "EPA has no legal alternative other than to conclude that no new air quality standard would be appropriate within the meaning of CAA section 109(a)(1)(B)." Finally, a number of commenters have argued that recent caselaw under the Clean Air Act and other statutes makes clear that EPA has a legal obligation to obtain and disclose the data used in these studies.[44]

In that same preamble, the EPA responded to those comments:

42. Letter from Mary D. Nichols, Assistant Administrator for Air and Radiation, EPA, to Dr. Douglas Dockery, Harvard School of Public Health (Jan. 31, 1997) (on file with author).

43. Ware Letter, *supra* note 40.

44. National Ambient Air Quality Standards for Particulate Matter, 62 Fed. Reg. 38,652, 38,689 (1997).

> In developing the proposed revisions to the PM NAAQS, the Administrator relied on the scientific studies cited in the rulemaking record, rather than on the raw data underlying them. In this case, the raw data consists of responses to health questionnaires based on information supplied by individual citizens, or computer tabulations of this information, which remains confidential, and air quality and monitoring data, most of which is now publicly available. EPA does not generally undertake evaluations of raw, unanalyzed scientific data as part of its public health standard setting process. Only in extreme cases—for example where there are credible allegations of fraud, abuse or misconduct—would a review of raw data be warranted. It would be impractical and unnecessary for EPA to review underlying data for every study upon which it relies as support for every proposed rule or standard. If EPA and other governmental agencies could not rely on published studies without conducting an independent analysis of the enormous volume of raw data underlying them, then much plainly relevant scientific information would become unavailable to EPA for use in setting standards to protect public health and the environment.[45]

Thus, while the EPA did request that the researchers in specific cases release their data for review, the Agency refused to require the release of such data as a requirement for a study's inclusion in the standard setting process.

III. Issues and Implications

A. *Research Credibility*

While the EPA ruled that there is no need for peer-reviewed, health study raw data to be released as a routine part of the NAAQS process, industry's public demands for the raw air pollution-health data in the case of the CAA standard setting process succeeded in generating skepticism in the press regarding the credibility of air pollution epidemiology results.[46] Thus, an unrestricted public release of such studies' subject health data would indeed provide one means for the researchers to allay any concerns that they are trying to hide something. Once the data were examined by all interest groups and reanalyzed by others, it would have the benefit of removing even the most remote possibility that the researchers are hiding anything, but at what cost?

The open and informed discussion of scientific issues and protections against biased analyses or reporting of scientific results are

45. *Id.* at 38,689 (citations omitted).

46. *See* Laura Johannes, *Pollution Study Sparks Debate Over Secret Data*, WALL ST. J., Apr. 7, 1997, at B1; Scott Allen, *Clean-Air Researchers Pressured to Show Data*, BOSTON GLOBE, Mar. 4, 1997, at A1.

indeed important to an informed debate regarding scientific issues. But, a key question about any policy mandating a blanket release of data as a means to advance debate on a scientific issue is: who is most likely to reexamine the publicly released data, and with what goal(s)? The three major groups that spring to mind are: (1) competing researchers wishing to capitalize on the more expensive and time-intensive work already done by the original researchers, by analyzing aspects of the data that the original authors have not yet had an opportunity to investigate;[47] (2) regulatory agencies wishing to verify the research results before relying on the studies for regulatory decision-making; and (3) vested interest groups that would be adversely affected by regulations, laws, or lawsuits based upon the published research.[48]

However, it is not necessary to speculate what might occur because past experience tells us much about what happens when health researchers allow open access to their data. The case of Dr. Herb Needleman and his research on the adverse effects of lead exposure on children provides one relevant case in point.[49] Dr. Needleman wrote:

> [H]aving satisfied myself that the tooth was a valid marker of past [lead] exposure . . . I studied a sample of children who were asymptomatic for lead, classifying them by dentine lead levels. The data showed that after controlling for a number of covariates, children with elevated lead in their teeth scored lower on tests of psychometric IQ, speech and language function, and on measures of attention . . . . The lead industry, in the form of the International Lead Zinc Research Organization . . . began to call for copies of my original data. I declined. I had seen what had happened to good data when massaged and distorted by industry technicians, and while I was happy to share my data with any bona fide scientist—and did—I was not willing to include the lead industry.[50]

As part of a lawsuit brought by the Department of Justice against three lead polluters, Dr. Needleman did ultimately have to make his records available for examination to witnesses on behalf of the lead industry, including a grantee of the International Lead Zinc Research Organization and someone who had appeared in testimony for Lead Industry Associates.[51] While the case was eventually settled out of court,

47. *See* Allen, *supra* note 46, at A1. Indeed, multiple analysis and publications often result from a single data set, and this step would deprive the original authors the opportunity to further "mine" their data set.

48. It might be well worth the expense to such vested interests to extensively investigate whether any conflicting conclusions could be derived from the same data.

49. *See* Herbert L. Needleman, *Salem Comes to the National Institutes of Health: Notes from Inside the Crucible of Scientific Integrity*, 90 PEDIATRICS 977 (1992).

50. *Id.* at 977-78.

51. *See id.* at 978.

Dr. Needleman indicated that these witnesses had written a lengthy document critiquing Needleman and his research that was forwarded to the National Institutes of Health by a law firm.[52]

As reported by Dr. Needleman:

> These kinds of issues are generally considered methodological disagreements and are fought out in the pages of journals; I could not understand why they were defined by my critics as scientific misconduct. Similar criticisms were raised before the EPA in 1982 and dismissed. These facts notwithstanding, in October of 1991, I was notified by the Dean of my medical school that an inquiry into charges of misconduct was being done at the instruction of NIH's Office of Scientific Integrity.[53]

Months after the hearing, Dr. Needleman was finally cleared, but he concluded that:

> If my case illuminates anything, it shows that the federal investigative process can be rather easily exploited by commercial interests to cloud the consensus about a toxicant's dangers, can slow the regulatory pace, can damage an investigator's credibility, and can keep him tied up almost to the exclusion of any scientific output for long stretches of time, while defending himself.[54]

Dr. Needleman's situation was also reported in an article in *The Chronicle of Higher Education* (*Chronicle*), along with that of a researcher who investigated the effects of tobacco company advertising on children, Dr. Paul Fischer.[55] Dr. Fischer's research was one of several studies published in the *Journal of the American Medical Association* (*JAMA*) that indicated children's attraction to the Camel cigarette "Joe Camel" advertising character.[56] R.J. Reynolds (RJR) responded by hiring consultants to analyze the studies and subpoenaed the research data supporting each of the studies.[57] The company's demands reportedly included that "the researchers supply the names and telephone numbers of all of the children who had participated in the studies."[58] As described by the *Chronicle*:

> Paul Fischer expected his college to back him. The request, he says, violated "the principles of confidentiality and academic freedom." Instead, the Medical College of Georgia sided with the tobacco company. Last

52. *See id.*
53. *Id.*
54. *Id.* at 980.
55. *See* Stephen Burd, *Scientists See Big Business on the Offensive*, THE CHRON. OF HIGHER EDUC., Dec. 14, 1994, at A26-A31.
56. Paul M. Fischer, M.D., et al., *Brand Logo Recognition by Children Aged 3 to 6 Years: Mickey Mouse and Old Joe the Camel*, 266 JAMA 3145 (1991).
57. *See* Burd, *supra* note 55, at A27.
58. *Id.* at A30.

> year, it turned over the documents .... Consultants to the cigarette industry then started criticizing his research. In disgust over the college's response, Dr. Fischer resigned and entered private practice in medicine.[59]

Since then, the substance of Dr. Fischer's research was subsequently verified by others,[60] including RJR itself in a memoranda that recently acknowledged that the company specifically targeted children in their advertising.[61] As reported by Dr. Fischer in a letter to *JAMA*:

> Our findings have been validated by other investigators. Henke studied 83 children aged 3 to 8 years using a similar board-game design and found a 54% recognition rate for Joe Camel, compared with 51% in our study. In a study funded by RJR, Mizerski looked at recognition rates among 790 children aged 3 to 6 years and found that 52% of all subjects could match Joe Camel with a cigarette and that an additional 8% associated him with a lit match, for an overall recognition rate of 60%. A third study, also funded by RJR and conducted by the Roper Group, surveyed 1,117 children aged 10 to 17 years and found a total awareness rate of the Joe Camel logo of 86%. The consistency of the findings across age groups, geographic populations, and various study designs validates the findings in our first report.
>
> Based on an estimated rate of 3,000 new teenage smokers per day, more than 5 million US teenagers have become regular smokers since the publication of our study. The most recent research not only confirms that advertising affects smoking rates, but also indicates that this effect is 3 times greater for teenagers than adults. Given the health consequences of cigarettes, tobacco industry advertising should be viewed as a major public health risk.[62]

More recently, Dr. John P. Pierce and colleagues have provided further confirmation, publishing the first longitudinal study (i.e., following subjects over time) indicating that tobacco company ads and promotional activities are indeed causally related to the initiation of smoking among adolescents.[63]

Ironically, on January 14, 1998, internal RJR memoranda were released that, according to the *Washington Post*, indicate that the company:

> sought for decades to reverse the declining sales of its brands by developing aggressive marketing proposals to reach adolescents as young

59. *Id.* at A26.
60. *See* Paul M. Fischer, M.D., *Recognition of Cigarette Advertisement Product Logos*, 277 JAMA 532 (1997) (citation omitted).
61. *See* John Mintz & Saundra Torry, *Internal R.J. Reynolds Documents Detail Cigarette Marketing Aimed at Children*, WASH. POST, Jan. 15, 1998, at A01.
62. Fischer, *supra* note 60 (citations omitted).
63. John P. Pierce et al., *Tobacco Industry Promotion of Cigarettes and Adolescent Smoking*, 279 JAMA 511 (1998).

as 14 years old.... The 81 documents contrast sharply with the company's repeated public declarations that it does not target young people, collectively sketching a picture of a company that seemed decades ago to determine that its financial future depended on recruiting a new generation of smokers. Many of the documents outline RJR's thinking that led up to the 1988 launch of its controversial Joe Camel cartoon advertising campaign.[64]

Thus, the criticized researcher was proven correct, and the vested interest company that attacked him was apparently seeking to discredit research findings that some individuals in that company must have known to have merit.

The Needleman and Fischer experiences are hardly unique, as the financial incentives to interest groups for such attacks on researchers are large. As recently noted by Dr. Richard A. Deyo in the *New England Journal of Medicine*:

> Attacks on health researchers are not new. Pierre Louis, for example, was vilified nearly two centuries ago for suggesting that bloodletting was an ineffectual therapy. In an open society such as ours, controversy is common and often socially useful. The fact that scientists are sometimes challenged by special-interest groups should be no surprise. However, with widening media coverage of health research, growing public interest in health hazards, and expanding research on the outcomes of clinical care, such attacks may become more frequent and acrimonious. The huge financial implications of many research studies invite vigorous attack.[65]

Dr. Deyo and colleagues go on to discuss three cases in other disciplines illustrating "how vituperative such attacks may be and the range of tactics employed," including: spinal-fusion surgery, multiple chemical sensitivity, and pharmaceuticals.[66] The authors conclude that:

> The common theme in these examples is an attack—through marketing, professional, media, legal, administrative, or political channels—on scientific results that ran counter to financial interests and strong beliefs. In each case, funding for the research involved peer review and the offending results were published in peer-reviewed journals. The interested parties had financial stakes in maintaining their market share or the legitimacy of a model of illness or a particular treatment. Their responses, which bypassed peer-reviewed scientific debate and further research, were nonscientific and aimed at discrediting the findings, investigators, or funding agencies. In each case, the attacks intimidated investigators, discouraged others from taking up the same lines of investigation, and took up the time of investigators and staff with legal, professional, and media responses.... The intent is to turn the tables on claimants, force them from a political to a judicial forum, and cast them as defendants.... In our cases, freedom-of-information requests, subpoenas, and complaints to the Office of Research Integrity were analogous to SLAPP [strategic lawsuits against public participation] suits.[67]

64. Mintz & Torry, *supra* note 61, at A01.
65. Richard A. Deyo et al., *The Messenger Under Attack—Intimidation of Researchers by Special-Interest Groups*, 336 NEW ENG. J. MED. 1176 (1997).
66. *Id.* at 1176-77.

Thus, policies as democratic and important as the Freedom of Information Act requirements can be subverted and employed as mechanisms for vested interests to "attack the messenger" when the message is financially or politically unwelcome to the interest group involved. It seems inevitable that the same things would have happened with Representative Aderholt's "Sunshine" amendment, despite its well intentioned goals.

Therefore, while there may be the initial benefit to researchers' credibility if they are willing to release all their underlying health data, past experience tells us that interest groups with a financial stake in the research outcome will likely be the primary user of that released data. These interest groups may use the data in order to further their own interests, irrespective of the merits of the original research, with little public health assessment benefit, and with the potential of significant public health disbenefit if appropriate public health measures are delayed by such tactics.

B. *Confidentiality of Participant Medical Records*

In March and April of 1997, as the pressure grew on the Harvard School of Public Health researchers to address the industry demands for their data, stories appeared in the *Wall Street Journal* and the *Boston Globe* on the topic.[68] In the *Wall Street Journal* article, one of the researchers pointed out that "giving up this data in violation of our agreements would completely cripple our ability to go out and do epidemiological studies of any type."[69]

Similarly, in the preamble of the Federal Register promulgation of the new PM standard, the EPA also pointed out that:

> such data are often the property of scientific investigators and are often not readily available because of ... arrangements made to maintain confidentiality regarding personal health status and lifestyle information of individuals included in such data. Without provisions of confidentiality,

67. *Id.* at 1177-78.
68. *See* sources cited *supra* note 46.
69. Johannes, *supra* note 46, at B1.

the possibility of conducting such studies could be severely compromised.[70]

Thus, the mandated release of health data collected in confidence during a research study, as proposed during the 105th Congress, would force researchers to violate the confidentiality agreements made with study participants at the start of the research years before. Retrospectively obtaining each subject's permission to release those data could be an onerous task, and may not be possible at all, in those cases where the subject has since died without designating responsible next-of-kin.

Moreover, when conducting new studies, investigators would have to tell subjects that their data would be publicly available at the end of the study, which could severely hamper researchers' ability to recruit new study populations. Thus, even if such data release mandates were to be applied only to new studies, one effect of the proposed data release mandate would be to stifle new research efforts funded by the federal government.

Ironically, these data release requirements would not apply to privately funded research, such as that funded by regulated industries, who have been among the most reticent in the past to make all of their private research data available to others. This bias in the data release requirement would be as unjustified as the present requirements in the House of Representatives that witnesses testifying before a committee must reveal their past government funding, but need not reveal past funding by interest groups that may have a vested interest in the outcome of the hearing.[71] Thus, under proposals such as Representative Aderholt's, vested interest groups will still be free to selectively publish research that supports their positions, while only government funded research will be encumbered by the data release requirements that, as will be shown below, will hamper its ability to expeditiously obtain research independent of special interest group influence upon which to base scientific assessments of health risks.

In light of these important concerns, and to at least partially offset the onerous effects of such a data release mandate, it seems possible that Congress might instead set up a new governmental agency, or assign an existing agency, with the task of collecting the data from researchers, and then releasing it to qualified parties on a limited basis, in order to at least partially protect the privacy rights of individuals. For example, this is presently done by the National Center for Health Statistics (NCHS) for

70. *See* National Ambient Air Quality Standards for Particulate Matter, 62 Fed. Reg. 38,652, 38,689 (1997) (to be codified at 40 C.F.R. pt. 51).

71. *See* Rules for the Comm. on Commerce: Rule 4(b)(2), 143 CONG. REC. H368-01, H369 (1997).

certain proprietary death certificate information, such as the date of death.[72] However, such a proposal for government control of data releases would raise the question of who is more appropriate to make decisions about sharing original research data: the individuals who collected it and were given permission to access the personal information by the subjects in question, or a government bureaucracy?

C. *Intellectual Ownership Rights*

A scientific data set often represents years of effort by a researcher and his or her colleagues, including: the conception of a research idea; the preparation of a research proposal for submission to a granting agency; obtaining institutional scientific Internal Review Board (IRB) approval to ethically collect the data; obtaining permission from each study participant; the collection, quality assurance, and statistical analysis of the data; and the preparation of reports documenting the work in the peer-reviewed literature. Usually, more than one publication results from a single data set, as there are multiple aspects of a data set that can be investigated. In the case of the Harvard Six Cities Studies, more than 100 research publications have resulted from this single data set. Oftentimes, further funding for support from agencies is obtained to investigate the many other scientific aspects of the data records. If the data were released after the first public use, then others could use the data to seek that funding to analyze and publish these further findings before the original researchers. In the case of the Six Cities Studies, the numerous publications and hundreds of thousands of dollars in research moneys that the researchers have accumulated for their institutions could have been lost to other competing researchers and institutions eager to get their hands on the Harvard data sets. Thus, a mandated "taking" of a data set from an original investigator shortly after the first public presentation of results from the study, as proposed in the 105th Congress,[73] and making it available to others for free, could represent a major loss, professionally and financially, to that investigator and his or her research institution.

If research is funded by a federal grant, does the government maintain any rights to demand access to those data beyond its rights to obtain data sets collected without federal funding? Congressman Brown, in his letter to the House Committee on Appropriations at the time of the Aderholt amendment, discussed this issue.

72. *See* National Center for Health Statistics, Centers for Disease Control and Prevention (last modified Mar. 3, 1998) <http://www.cdc.gov/nchswww/index.htm>.

73. *See Amendment to Treasury Bill*, *supra* note 2.

> [I]t is important to understand that the federal government usually supports research through grants, not through contracts. The distinction is significant. The purpose of a grant is to support or stimulate activity which serves the public good, such as the increase and diffusion of scientific knowledge. Unlike a contract, a grant does *not* purchase the product of the grantee's work. (See Government Accounting Office, *Principles of Federal Appropriation Law* (2d Ed.) pages 10-3 through 10-10.)
>
> For this reason, it has traditionally been understood that researchers receiving federal grants nevertheless retain significant property interests in their research. Congress has explicitly recognized and even recently expanded those property rights. For example, grantees have the right to copyright documents they produce with grant support, and can own and patent intellectual property created under the grant. (See Office of Management and Budget Circular A-110.) Just last year, Congress passed the National Technology Transfer and Advancement Act of 1995 (P.L. 104-113), which expanded the rights of persons receiving federal research support to own and license intellectual property created with federal support.[74]

Thus, the government apparently has no more right to insist that researchers who have collected scientific data as part of a federally funded grant release their data than it has to make the same demands of private industry funded research. As a result, any such mandates for the public release of data sets underlying published research results should apply equally to both industry and government funded research, and should be viewed as a "taking" of property from those investigators.

D. *Effect on Scientific Progress*

The proposed mandate to require raw data release upon first public use of results from those data, although aimed at advancing scientific knowledge, would undoubtedly have the reverse effect in many ways. As previously noted, a researcher in the Six Cities Studies, Dr. Douglas Dockery, pointed out in a *Wall Street Journal* article that violating their subject agreements would cripple their ability to do new epidemiological studies.[75] Potential subjects would be less likely to participate in research where their personal medical data will be made public. In addition, Dr. Joel Schwartz, another Six Cities data researcher, also noted in that same article that "[n]o epidemiologist can afford to be buried in so much time-consuming controversy for every study, yet that is what industry promises for every data set they get their hands on."[76] Dr. Needleman's experience

74. Brown Letter, *supra* note 10.
75. *See* Johannes, *supra* note 46, at B1.
76. *Id.*

is confirmation of the real-world potential for a realization of Dr. Schwartz's concerns. In addition, there would likely be a reluctance on the part of researchers to publicly release any research results from a study until all possible research opportunities are exhausted, if they must release their data after doing so. Financial considerations would likely ensure that the first completed results from a data set might well be used solely as justification in subsequent grant applications for further funding, rather than expeditiously published, and would therefore not be available to the public, the research community, or regulatory agencies until years later, when all further research avenues had been exhausted. In other words, the requirements for public release of data would have the overall effect of inhibiting, not enhancing, scientific progress and would thereby also have the effect of inhibiting governmental agencies from being fully informed about the most up-to-date state of scientific knowledge when making regulatory decisions.

E. *Unfunded Mandates*

Among the less politically popular things that Congress can do is to impose an "unfunded mandate," or a requirement for individuals to do things without providing any financial support to address these new requirements—which is exactly what these data release mandates represent. As noted in Representative Brown's letter to the Appropriations Committee:

> The Aderholt amendment would impose a significant unfunded mandate on individual researchers and universities—including state universities. To comply, universities would have to maintain a central repository of all of the raw data produced by all of its federally-supported researchers, respond to all public requests for documents at its own cost, and review all of the material before disclosure for potential legal liability for disclosure of sensitive personal or business information.[77]

F. *Are Existing Mechanisms Sufficient?*

Certainly important among the issues raised by data release mandates is the question as to whether the scope of the "solution" advanced is consistent with the "problem" it proposes to address. As stated by Representative Brown in his letter to the Appropriations Committee:

> Before we impose these costly burdens, we ought to ask ourselves what is the problem? As the ranking minority Member of the Science Committee,

77. Brown Letter, *supra* note 10.

> I am unaware that there is any general problem with federally-funded scientists failing to publish research results in public, peer-reviewed journals. I suspect that federally-funded scientists are no different than their colleagues in wanting to publish their work in respected scientific journals and to have a wide distribution of their research results.
>
> Nor am I aware that there is a general concern about the integrity of federally-funded research. The peer-review process, while not perfect, does a pretty good job of weeding out flawed research. In that regard, requiring the mandatory disclosure of raw research data would be overkill.[78]

Indeed, of the roughly 28,000 biomedical articles published each year by researchers in the United States,[79] only a small percentage have letters written to the journal editor about them, and only a handful of those are controversial enough to warrant requesting their data for reanalysis. Clearly, the requiring of tens of thousands of researchers to prepare their data in a form appropriate for public release and the setting up of a bureaucracy (or bureaucracies) to handle these data and their dissemination is regulatory overkill for a perceived problem involving such a very small percentage of these researchers.

Thus, there is no pervasive scientific credibility problem in federally-funded research that justifies the global mandates called for in Congress during 1997. A focused approach would seem much more commensurate with the scale of the perceived problem.

But what about those specific cases in which real scientific controversy does exist? Representative Brown, in his letter to the Appropriations Committee, goes on to address this point, stating:

> There may, of course, be isolated instances where there are problems . . . . Those instances need to be addressed on a case-by-case basis to ensure the careful consideration of all factors, including the confidentiality of patient and medical records. Agencies have adequate existing legal authority to obtain research results and data for federal purposes in such instances. There is no need for the sweeping across-the-board approach proposed in the Aderholt amendment.[80]

Available mechanisms used in the past to address specific concerns include an evaluation of the data integrity by a disinterested third party. In the case of the Harvard study data sets, even though there were no charges of any scientific misconduct, the HEI has again stepped in to

78. *Id.*

79. *See* National Science Board, *Science and Engineering Indicators* (1996) (visited May 28, 1998) <http://www.nsf.gov/sbe/srs/seind96/start.htm>.

80. Brown Letter, *supra* note 10.

address demands for a reexamination of the data and its analysis.[81] HEI will provide a neutral party to evaluate the scientific integrity of the data and the research that led to the important Six Cities Studies finding, without the need for the Harvard researchers to make their data publicly available.

In cases where scientific controversy surrounds a published research document that an Administrator has relied upon in making a regulatory ruling, the courts also provide an existing avenue to address concerns. A comprehensive discussion of the legal precedents surrounding the issue of research data availability is presented by the EPA in the preamble to the recent PM standard revision. One example where the courts interceded in the process is provided in *Endangered Species Committee v. Babbitt*[82] (*Gnatcatcher*), which involved the range of the coastal California gnatcatcher. In its Final Rule of the Particulate Matter NAAQS, the EPA stated that:

> the *Gnatcatcher* opinion itself notes, "courts have generally allowed agencies to rely on scientific reports." Thus, the question at issue in *Gnatcatcher* was whether specific circumstances exist in which an agency may not be entitled to rely on studies alone. In the *Gnatcatcher* case, a single author had published two directly contradictory studies on the same issue, while relying on the same data. In light of this clear contradiction, commenters in that rulemaking argued that without the underlying data it was impossible to determine whether the conclusions in either study were correct. The district court noted that:
>
> "The Secretary had before him a report by an author who, two years before had analyzed the same data and come to an opposite conclusion. It is the disputed nature of this report that distinguishes this from other cases where a scientific report alone has been considered sufficient for ESA purposes."
>
> . . . Thus, according to the court: "While courts have generally allowed agencies to rely on scientific reports * * * this is not sufficient in this case because the report itself is under serious question."[83]

In this case, the court concluded that, in the specific situation in which the author published conflicting results, the data should be made public, and this was required of the Department of the Interior.[84] This opinion appears to support the EPA's position in issuing the new $PM_{2.5}$

81. *See generally* Hadley Letter, *supra* note 38.

82. 852 F. Supp. 32 (D.D.C. 1994) [hereinafter *Gnatcatcher*] (cited in National Ambient Air Quality Standards for Particulate Matter, 62 Fed. Reg. 38,652, 38,692 (1997)).

83. *See* National Ambient Air Quality Standards for Particulate Matter, 62 Fed. Reg. 38,652, 38,692 (1997) (quoting *Gnatcatcher*, 852 F. Supp. at 37).

84. *See Gnatcatcher*, 852 F. Supp. at 43.

standard that only in extreme cases is the review of raw scientific data warranted.[85] Clearly, the courts provide a working and viable mechanism to address concerns about scientific research in those situations.

Thus, there does not appear to be a pervasive problem with the integrity of peer review literature results that calls out for the type of regulatory intervention being proposed on Capitol Hill. Moreover, in the rare cases in which the integrity of peer reviewed published research is credibly questioned, not just because the results are undesirable to vested interests, there are existing mechanisms in place to address and resolve those concerns.

IV. Discussion and Recommendations

Overall, it should be apparent from the considerations presented that the recent proposal to mandate the immediate and unrestricted release of raw health research data underlying federally-funded medical and public health research is an overly heavy-handed and burdensome solution to the infrequent problems that arise regarding limitations in access to published research data. Moreover, such an unrestricted data-release policy has the major drawback that it will undoubtedly worsen the very real and serious present-day problem of unwarranted attacks on scientists and physicians who publish research with conclusions that run counter to vested interests.

Qualified researchers who have published research results potentially damaging to vested interests have come under intensive attacks in the media through the initiation of scientific misconduct charges, via legal actions, and by the influencing of government agencies to demand specific studies' data release. Many of these attacks have come even when no scientific misconduct is suspected. These researchers have generally been ill-prepared to defend themselves. The attacks cause them to spend a great deal of time and money in defense of charges initiated or encouraged by vested interest groups having far "deeper pockets" and significant financial incentives to relentlessly pursue the attacks. The result is extremely detrimental to the scientists involved, both financially and professionally, and in one case documented here, has actually caused a researcher to leave the field of health research, despite the fact that the substance of his research results were later confirmed by others.[86] It may also have slowed the speed at which regulators took action in the cases where scientific integrity was questioned. A data release mandate would provide vested interest groups with even more "fodder" with which to attack the research upon which federal regulations

85. *See supra* text accompanying note 45.
86. *See supra* text accompanying notes 59-61.

unfavorable to their financial interests are based. Thus, in addition to slowing scientific progress, the legal and financial burdens on research institutions, and the undermining of research subject privacy, it seems very clear that a mandate to release the underlying data behind all published, federally-funded research would greatly exacerbate the problem of unwarranted attacks on researchers.

However, in the face of inevitable, future, contentious public policy debates, how can we best ensure that the important processes of information exchange, data-sharing, and validation of results are carried out without unwittingly making the affected researchers the target of unfair criticism and harassment by vested interests? Clearly, to avoid being onerous, any solution involving data release by researchers must be focused specifically on the critical issues and results, rather than a global release of all raw data. The solution will also need to provide a structured framework for the conscientious handling of data transfer, protection, and evaluation. This might involve the designation of rules and funding for the establishment of a deliberative entity to serve the role played so well by the HEI in the case of the Harvard air pollution research results. Perhaps the National Academy of Sciences could be funded to provide a forum for the design and implementation of such a deliberative body. The key interested parties will need to be involved, or at least considered, in designing such a mechanism, including: the scientists and/or physicians conducting the research; the editors of the journals that publish such research; the potentially affected vested interest groups and industries; and the governmental agencies involved in promulgating regulations based upon the research.

The editors of the various scientific journals that publish this research have an especially important responsibility to play a larger role in setting up a mechanism to address this issue. To date, the role of these journals has largely been limited to having scientific papers carefully reviewed before publication, rejecting inadequate papers, and/or passing along major and minor revisions suggested by scientific reviewers. After that, the journals basically "wash their hands" of any subsequent problems, merely publishing any substantive letters sent in to the journal criticizing a published paper. This seems an inadequate role in today's world of scientific debate in which the stakes can be so high, and in which researchers largely are left to fend for themselves, many times not even being supported by their own research institutions. Once a journal publishes an article, it must shoulder a responsibility for that work that goes beyond the mere publishing of letters to the editor and their responses. The *New England Journal of Medicine* (*NEJM*) has taken an aggressive stand on the issue of editorial writers and potential financial

conflicts of interest.[87] However, the *NEJM* has not yet "weighed in" on the issue of the independent evaluation of the scientific merits of already published research, even though it published a controversial air pollution study by Dr. Douglas Dockery et al.[88] Prominent journals, such as the *NEJM*, should consider setting up a review panel comprised of representatives, such as the editors from each journal, that would organize a second, more extensive, peer-review of especially controversial papers. This might be analogous to the Committee on Publication Ethics recently set up by editors of prominent British journals such as the *British Medical Journal* and *Lancet*.[89] Through a scientific journal "court of appeals," expeditious and fair re-reviews of contentious results might be conducted.

Whether these suggestions are followed, or some alternative mechanism is adopted, it seems imperative that the scientific journals and the scientific community "face-up" to the issues of peer-reviewed and published research method evaluation and data access. Otherwise, Congress may in fact take it upon itself to impose a remedy that will likely be far worse for science and policy-making than the perceived problem it proposes to cure.

---

87. Marcia Angell & Jerome P. Kassirer, *Editorials and Conflicts of Interest*, 335 NEW ENG. J. MED. 1055 (1996).

88. Douglas W. Dockery, Sc.D., et al., *An Association Between Air Pollution and Mortality in Six U.S. Cities*, 329 NEW ENG. J. MED. 1753 (1993).

89. Jocelyn Kaiser, *British Editors Form Misconduct Panel*, 277 SCIENCE 627 (1997).