UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
WASHINGTON D.C. 20460

OFFICE OF THE ADMINISTRATOR
SCIENCE ADVISORY BOARD

April 24, 2020

EPA-SAB-20-005

The Honorable Andrew Wheeler
Administrator
U.S. Environmental Protection Agency
1200 Pennsylvania Avenue, N.W.
Washington, D.C. 20460

Subject: Science Advisory Board (SAB) Consideration of the Scientific and Technical Basis of EPA's Proposed Rule Titled *Strengthening Transparency in Regulatory Science*

Dear Administrator Wheeler:

As part of its statutory duties, the EPA Science Advisory Board (SAB) may provide advice and comment on the scientific and technical basis of certain planned EPA actions. The Environmental Research, Development, and Demonstration Authorization Act of 1978 (ERDDAA) requires the EPA to make available to the SAB proposed criteria documents, standards, limitations, or regulations provided to any other federal agency for formal review and comment, together with relevant scientific and technical information on which the proposed action is based. The SAB may then provide advice and comments on the adequacy of the scientific and technical basis of the proposed action. At its May 31, 2018 public meeting, the chartered SAB identified the proposed rule titled *Strengthening Transparency in Regulatory Science* (Proposed Rule) as a planned action that merited review. In April 2019, the SAB Work Group on Planned Actions for SAB Consideration of the Underlying Science recommended that the SAB review the Proposed Rule and at its public meeting on June 5-6, 2019, the SAB elected to review the scientific and technical basis of the Proposed Rule. Subsequent to the June meeting, a work group of chartered SAB members was formed to carry out the review. Members of this work group then took the lead in SAB deliberations on this topic at a public teleconference held on January 21, 2020. On March 3, 2020, EPA released a Supplemental Notice of Proposed Rulemaking (SNPRM), which contains additional information and clarification of certain terms and provisions of the Proposed Rule. The SAB's advice and comments on the Proposed Rule and aspects of the supplemental notice are provided in the enclosed report.[1]

The Proposed Rule is intended to strengthen the transparency of EPA regulatory science by providing that, for the science pivotal to its significant regulatory actions, EPA will ensure that the data and models underlying the science are publicly available in a manner sufficient for validation and

[1] Drs. John Graham and Donald van der Vaart indicated that they did not concur with the enclosed report. Their dissenting opinions are included in Appendix A.

analysis. The SAB recognizes the importance of this rule and its purpose, establishing transparency of the influential scientific information used for significant regulations and enhancing public access to scientific data and analytical methods to help ensure scientific integrity, consistency and robust analysis.[2] Strengthening transparency by improving access to data can lead to an increase in the quantity and the quality of evidence that informs important regulatory science and policy decisions. The scientific community is moving toward adopting the precept of sharing accurate data and information to increase credibility, high-quality outcomes and public confidence in science. The SAB supports the adoption of this precept. However, the SAB finds that key considerations that could inform the Proposed Rule are not present in the proposal or presented without analysis and explanation of scope. In addition, certain key terms and implementation issues have not been adequately defined or described. To provide clarity on the procedures for conducting the proposed efforts, the SAB strongly encourages the development of additional policy and/or guidance documents. In addition, the SAB has concerns about the scientific and technical challenges of implementing some requirements of the Proposed Rule. The SAB's major comments and recommendations are as follows:

- The Proposed Rule requires the EPA to clearly identify all studies (or other regulatory science) relied upon when it takes any significant final agency regulatory action and to make all such studies available to the public to the extent practicable. The EPA's Supplemental Notice of Proposed Rulemaking expands the scope of this requirement to include studies relied upon in influential scientific information (i.e., scientific information that will or does have a clear and substantial impact on important public policies or private sector decisions). In some cases, this requirement could be complex and/or impractical because studies could be considered when integrating the evidence but not directly used to determine specific regulatory standards or levels. The lack of specific criteria for what might satisfy the requirement makes it difficult for the SAB to understand the implications. The SAB recommends that the Final Rule describe in greater detail and clarity how the requirement can be met. To effectively implement this requirement, at the minimum, pivotal regulatory science or other regulatory science (as defined in the Proposed Rule), as well as other types of data necessary to make regulatory determinations, should be publicly available with a scientific justification explaining why they were used for regulatory decisions. As stated in the Proposed Rule, EPA should indicate whether an independent peer review of all pivotal regulatory science has been conducted. As discussed below, the proposed rule provides for exceptions to be granted if it is not feasible to implement these requirements. The SAB suggests that the EPA consider establishing an office (or virtual office) on data sharing, and a peer review panel or workgroup to assist EPA in this process.

- The Proposed Rule and the supplemental proposal indicate that when promulgating significant regulatory actions or finalizing influential scientific information, the Agency shall ensure that data and models underlying pivotal regulatory science are publicly available in a manner sufficient for independent validation. EPA's Supplemental Notice of Proposed Rulemaking has indicated that the requirement to make data and models underlying pivotal science available in a manner sufficient for independent validation would apply to all data and models, not just dose-response models. Greater clarity is needed in the definitions of those terms. The SAB recommends that the definitions provided in the Proposed Rule and supplemental proposal be expanded and supported in the context of a guidance document.

[2] One SAB member does not agree with this statement, particularly in the context of international studies and those developed before the acceptance of current standards on science transparency that are discussed in this report

- The Proposed Rule indicates that the Administrator may grant exceptions to the requirements on a case-by-case basis if it is determined that compliance is impracticable because: (1) it is not feasible to ensure that all pivotal regulatory science is publicly available for independent validation in a manner that is: consistent with law; protects privacy, confidentiality and confidential business information; and is sensitive to national and homeland security; or (2) it is not feasible to conduct independent peer review on all pivotal regulatory science used to justify regulatory decisions. The SAB notes that in the supplemental proposal, EPA has provided additional information about exceptions to the requirements. However, the SAB recommends that EPA develop specific criteria for such exceptions as part of the Final Rule. Although it will be difficult to develop criteria for exceptions, outlining such criteria would benefit EPA and help ensure that the principles of transparency outlined in the Proposed Rule are accomplished. Case-by-case exceptions without such criteria may create public concerns about inappropriate exclusion of scientifically important studies. A framework and/or guidance document could also help EPA clarify how current scientific review procedures will be affected by this rule. It might be useful for the EPA to consider recommendations from a scientific advisory committee when making decisions to waive requirements.

- To assess the feasibility of making data and models available in a manner sufficient for independent validation, a number of questions must be answered (such as how to treat studies that are formatted in a manner that makes the data difficult to share, how to move forward if laboratories refuse to collate and release data, how to handle sensitive information such as individual participants' addresses, how to manage international studies, and how to manage conclusions drawn from meta-analysis). The SAB notes that historical data or international datasets may be unavailable or may have been discarded if deemed not necessary to maintain. A possible way to address this problem is to apply rule requirements only to information developed after the effective date of a Final Rule. Experimental considerations (such as the appropriateness of controls, protocols employed, limits of quantification, and other considerations) must be made known to determine whether data are valid. The SAB recommends that the EPA consider these questions and more specifically define "independent validation" in the Final Rule because this definition drives the feasibility of whether the EPA can make data and models available in a manner sufficient for validation.

- The SAB recommends that EPA consider seeking input from experts in library science, data curation management, and data retention to identify best practices and tools to ensure efficiency and utility of data that are made available. There will be costs associated with assessing and disseminating data as required in the Proposed Rule.

- The requirement in the Proposed Rule that "data" be made publicly available is vague and, as a result, can be interpreted in different ways. Extensive work is required, across a diversity of fields, data types and data of different ages, to understand the implications of adopting different definitions of data and more clarity is needed to define the nature of the "data" that are being required. The SAB notes that EPA has defined data in the supplemental proposal. The SAB finds that EPA could benefit from using the term "analysis dataset" as it refers to data that have been collected and processed (e.g., cleaned and transformed) for analysis.

- As stated in the Proposed Rule, when the Agency is making data or models publicly available, it shall do so consistent with the law to protect privacy and confidentiality. Therefore, this regulation should build on techniques and practices to protect the confidentiality of human data when data and models underlying pivotal regulatory science are made available to the public. Some individual data used in epidemiological studies are held by federal agencies such as the Centers for Disease Control

and Prevention (CDC) or the Department of Health and Human Services (Medicare data), while other data have been developed by state, local or tribal governments, academic institutions or private organizations, among others. Some federal agencies have efficiently developed methods for making data available to the public (e.g., Census Bureau, CDC). Currently, no comparable system exists for datasets that are owned by non-federal governments (e.g., states, tribes), and/or owned by private societies/organizations or academic institutions which are themselves protected by strong privacy and confidentiality requirements through their Institutional Review Boards (IRBs). The SAB recommends that EPA collaborate with federal agencies to make individual-level data (i.e., data associated with individuals in a sample) available through the system of Federal Statistical Research Data Centers, which are already widely used by the Census Bureau to allow researchers to gain access to individual data while protecting the data from public dissemination. The SAB also notes that there are techniques and practices to protect sensitive data that have been well-developed by researchers involved in studies with human subjects. The Proposed Rule should clearly address the issue of obtaining public access to datasets while maintaining the privacy of the study participants and confidentiality of the data. If this issue is not clearly addressed, there is a risk of entirely excluding datasets containing personally identifiable information from being considered as pivotal regulatory science.

- If the EPA wants to encourage reanalysis to validate datasets that are critically important for regulation, the Agency should consider providing funds to conduct such reanalysis. A model for this was established by the Health Effects Institute (HEI) in its 2000 reanalysis of datasets from the Six Cities Study and the American Cancer Society.

- The Proposed Rule requires the EPA to describe and document any assumptions and methods that pertain to the use of data and models underlying pivotal regulatory science and to describe variability and uncertainty. It is unlikely that uniform standards will be used across laboratories to report this information; therefore, the SAB strongly suggests that, before the implementation of the Final Rule, the EPA develop a guidance document pertaining to documentation of assumptions, methods, variability, and the definition of data and uncertainty.

The SAB appreciates the opportunity to provide the EPA with advice and comment on the Proposed Rule. We look forward to receiving the Agency's response.

Sincerely,

/s/

Dr. Michael Honeycutt, Chair
Science Advisory Board

Enclosure

NOTICE

This report has been written as part of the activities of the EPA Science Advisory Board (SAB), a public advisory group providing extramural scientific information and advice to the Administrator and other officials of the Environmental Protection Agency. The SAB is structured to provide balanced, expert assessment of scientific matters related to problems facing the Agency. This report has not been reviewed for approval by the Agency and, hence, the contents of this report do not necessarily represent the views and policies of the Environmental Protection Agency, nor of other agencies in the Executive Branch of the Federal government, nor does mention of trade names of commercial products constitute a recommendation for use. Reports of the SAB are posted on the EPA Web site at http://www.epa.gov/sab.

U.S. Environmental Protection Agency
Science Advisory Board

CHAIR

Dr. Michael Honeycutt, Division Director, Toxicology Division, Texas Commission on Environmental Quality, Austin, TX

MEMBERS

Dr. Rodney Andrews, Director, Center for Applied Energy Research, University of Kentucky, Lexington, KY

Dr. Hugh A. Barton, Independent Consultant, Mystic, CT

Dr. Barbara Beck, Principal, Gradient Corp., Cambridge, MA

Dr. Deborah Hall Bennett, Professor, Environmental and Occupational Health Division, Department of Public Health Sciences, School of Medicine, University of California, Davis, Davis, CA

Dr. Frederick Bernthal, President Emeritus and Senior Advisor to the Board of Trustees, Universities Research Association, Washington, DC

Dr. Bob Blanz, Associate Director, Office of Water Quality, Division of Environmental Quality, Arkansas Department of Environmental Quality, North Little Rock, AR

Dr. Todd Brewer, Senior Manager, Grants, Education, and Utility Programs, American Water Works Association, Denver, CO

Dr. Joel G. Burken, Curator's Professor and Chair, Civil, Architectural, and Environmental Engineering, College of Engineering and Computing, Missouri University of Science and Technology, Rolla, MO

Dr. Janice E. Chambers, William L. Giles Distinguished Professor and Director, Center for Environmental Health Sciences, College of Veterinary Medicine, Mississippi State University, Mississippi State, MS

Dr. John R. Christy, Distinguished Professor of Atmospheric Science and Director of Earth System Science Center, University of Alabama in Huntsville, Huntsville, AL

Dr. Samuel Cohen, Professor, Pathology and Microbiology, University of Nebraska Medical Center, Omaha, NE

Dr. Louis Anthony (Tony) Cox, Jr., President, Cox Associates, Denver, CO

Dr. Alison C. Cullen, Interim Dean and Professor, Daniel J. Evans School of Public Policy and Governance, University of Washington, Seattle, WA

Dr. Otto C. Doering III, Emeritus Professor, Department of Agricultural Economics, Purdue University, W. Lafayette, IN

Dr. Susan P. Felter, Research Fellow, Global Product Stewardship, Procter & Gamble, Mason, OH

Dr. Joseph A. Gardella, SUNY Distinguished Professor of Chemistry, Department of Chemistry, College of Arts and Sciences, University at Buffalo, Buffalo, NY

Dr. John D. Graham, Dean, O'Neill School of Public and Environmental Affairs, Indiana University, Bloomington, IN

Dr. John Guckenheimer, Professor Emeritus and Interim Director, Center for Applied Mathematics, Cornell University, Ithaca, NY

Dr. Margaret MacDonell, Department Head, Argonne National Laboratory, Lemont, IL

Dr. Robert E. Mace, The Meadows Center for Water and the Environment, Texas State University, San Marcos, TX

Dr. Clyde F. Martin, Horn Professor of Mathematics, Emeritus, Department of Mathematics and Statistics, Texas Tech University, Crofton, MD

Dr. Sue Marty, Senior Toxicology Leader, Toxicology & Environmental Research, The Dow Chemical Company, Midland, MI

Mr. Robert W. Merritt, Independent Consultant, Houston, TX

Dr. Larry Monroe, Independent Consultant, Braselton, GA

Dr. Thomas F. Parkerton, Senior Environmental Scientist, Toxicology & Environmental Science Division, ExxonMobil Biomedical Science, Spring, TX

Dr. Robert Phalen, Professor, Air Pollution Health Effects Laboratory, Department of Medicine, University of California-Irvine, Irvine, CA

Dr. Kenneth M. Portier, Independent Consultant, Athens, GA

Dr. Robert Puls, Owner/Principal, Robert Puls Environmental Consulting, Bluffton, SC

Dr. Kenneth Ramos, Executive Director, Institute of Biosciences and Technology, Texas A&M University, Houston, TX

Dr. Tara L. Sabo-Attwood, Associate Professor and Chair, Department of Environmental and Global Health, College of Public Health and Health Professionals, University of Florida, Gainesville, FL

Dr. Mara Seeley, Unit Chief – Exposure Assessment, Environmental Toxicology Program, Bureau of Environmental Health, Massachusetts Department of Public Health, Boston, MA

Dr. Anne Smith, Managing Director, NERA Economic Consulting, Washington, DC

Dr. Richard Smith, Professor, Department of Statistics and Operations Research, University of North Carolina, Chapel Hill, NC

Dr. Jay Turner, Professor and Vice Dean for Education, Department of Energy, Environmental and Chemical Engineering, McKelvey School of Engineering, Washington University, St. Louis, MO

Dr. Brant Ulsh, Principal Health Physicist, M.H. Chew & Associates, Cincinnati, OH

Dr. Donald van der Vaart, Senior Fellow, John Locke Foundation, Raleigh, NC

Ms. Carrie Vollmer-Sanders, Director, Agriculture Engagement Strategy, Efroymson Conservation Center, The Nature Conservancy, Indianapolis, IN

Dr. Kimberly White, Senior Director, Chemical Products and Technology Division, American Chemistry Council, Washington, DC

Dr. Mark Wiesner, Professor, Department of Civil and Environmental Engineering, Director, Center for the Environmental Implications of NanoTechnology (CEINT), Pratt School of Engineering, Nicholas School of the Environment, Duke University, Durham, NC

Dr. Peter J. Wilcoxen, Laura J. and L. Douglas Meredith Professor for Teaching Excellence, Director, Center for Environmental Policy and Administration, The Maxwell School, Syracuse University, Syracuse, NY

Dr. Richard A. Williams, Retired Economist, U.S. Food and Drug Administration, Independent Consultant, McLean, VA

Dr. S. Stanley Young, Chief Executive Officer, CGStat, Raleigh, NC

Dr. Matthew Zwiernik, Professor, Department of Animal Science, Institute for Integrative Toxicology, Michigan State University, East Lansing, MI

SCIENCE ADVISORY BOARD STAFF

Dr. Thomas Armitage, Designated Federal Officer, U.S. Environmental Protection Agency, Washington, DC

TABLE OF CONTENTS

1. EXECUTIVE SUMMARY

As part of its statutory duties, the EPA Science Advisory Board (SAB) may provide advice and comment on the scientific and technical basis of certain planned EPA actions. The Environmental Research, Development, and Demonstration Authorization Act of 1978 (ERDDAA) requires the EPA to make available to the SAB proposed criteria documents, standards, limitations, or regulations provided to any other federal agency for formal review and comment, together with relevant scientific and technical information on which the proposed action is based. The SAB may then provide advice and comments on the scientific and technical basis of the proposed action.

At its May 31, 2018 public meeting, the Chartered SAB identified the proposed rule titled *Strengthening Transparency in Regulatory Science* (Proposed Rule) as a planned action that merited review. In April 2019, the SAB Work Group on Planned Actions for SAB Consideration of the Underlying Science recommended that the SAB review the Proposed Rule and at its public meeting on June 5-6, 2019, the SAB elected to review the scientific and technical basis of the Proposed Rule. Subsequent to the June meeting, a work group of chartered SAB members was formed to carry out the review. Members of this work group then took the lead in SAB deliberations on this topic at a public teleconference held on January 21, 2020. On March 3, 2020, EPA announced a Supplemental Notice of Proposed Rulemaking (SNPRM),[3] which contains additional information and clarification of certain terms and provisions of the Proposed Rule. The SAB's advice and comments on the Proposed Rule and aspects of the supplemental notice are provided in this report.[4]

The Proposed Rule is intended to strengthen the transparency of EPA regulatory science by providing that, for the science pivotal to its significant regulatory actions, EPA will ensure that the data and models underlying the science are publicly available in a manner sufficient for validation and analysis. The SAB recognizes the importance of this rule. It can enhance public access to scientific data and analytical methods, and help ensure scientific integrity, consistency and robust analysis.[5] Strengthening transparency by improving access to data can lead to an increase in both the quantity and the quality of evidence that informs important regulatory science and policy decisions. The scientific community is moving toward adopting the precept of sharing accurate data and information to foster credibility, high-quality outcomes and public confidence in science. The SAB supports the adoption of this precept.

The SAB finds, however, that key considerations that could inform the Proposed Rule are not present in the proposal, or presented without analysis, and certain key terms and implementation issues have not been adequately defined or described. In addition, the SAB has concerns about the scientific and technical challenges and feasibility of implementing some requirements of the

[3] Strengthening Transparency in Regulatory Science; Supplemental Notice of Proposed Rulemaking, 25 Federal Register (18 March, 2020), pp. 15396-15406.
[Available at: https://www.federalregister.gov/documents/2020/03/18/2020-05012/strengthening-transparency-in-regulatory-science]

[4] Drs. John Graham and Donald van der Vaart indicated that they do not concur with this report. Their dissenting opinions are included in Appendix A.

[5] One SAB member does not agree with this statement, particularly in the context of international studies and those developed before the acceptance of current standards on science transparency that are discussed in this report.

Proposed Rule. Given the relatively skeletal nature of the Proposed Rule, it is not possible to define the implications of the rule with confidence. To ensure that the rule is evidence-based EPA must provide greater clarity regarding details of the rule and how it will be implemented, as well as example analyses of how it would be deployed. The development of additional policy and/or guidance documents is strongly recommended to provide clarity on the procedures for conducting the proposed efforts.

The Proposed Rule requires the EPA to clearly identify all studies (or other regulatory science) relied upon when it takes any significant final agency regulatory action and to make all such studies available to the public to the extent practicable. The EPA's Supplemental Notice of Proposed Rulemaking expands the scope of this requirement to include studies relied upon in influential scientific information (i.e., scientific information that will or does have a clear and substantial impact on important public policies or private sector decisions). In some cases, this requirement could be complex and/or impractical because studies could be used and considered when integrating evidence but not directly used to determine specific regulatory standards or levels. The lack of specific criteria for what might satisfy the requirement makes it difficult for the SAB to understand the implications. The SAB recommends that the Final Rule describe in greater detail and clarity how the requirement can be met. To effectively implement this requirement, at the minimum, pivotal regulatory science and regulatory science (as defined in the Proposed Rule), as well as other types of data necessary to make regulatory determinations, should be publicly available with a scientific justification explaining why they were used for regulatory decisions. As stated in the Proposed Rule, EPA should indicate whether an independent peer review of all pivotal regulatory science has been conducted. As discussed below, the proposed rule provides for exceptions to be granted if it is not feasible to implement these requirements.

If the intent of the rule is to identify and make available the pivotal regulatory science that was relied upon, scientific and technical challenges of implementing this requirement will consist of: (1) having EPA be explicit about which studies are pivotal to the recommended regulatory action, and (2) making the data and models for the underlying pivotal studies publicly available. Given the lack of clarity in the Proposed Rule, it is difficult to understand how this regulatory action could be accomplished in a standardized and consistent manner. The Proposed Rule and supplemental proposal acknowledge the importance of protecting personally identifiable information (PII) and confidential business information (CBI). This SAB report contains additional suggestions regarding the protection of PII and CBI. The EPA must make certain that PII and CBI are not available to persons and groups who are not approved to have access to this information. Without adequate protection, industry data generated by one company can be used by other companies to fulfill regulatory requirements in other geographies. It would be beneficial for the EPA to develop specific policies to address: the protection of PII and CBI, exceptions that would be appropriate where PII and CBI cannot be released, and whether data compensation should be considered. The identification and release of individual data to the public in epidemiological studies that arise from small datasets or targeted geographic areas is problematic, but there are approaches that have been used to protect PII, (e.g., conducting independent analysis by a third party such as Health Effects Institute). However, the lack of criteria for what data might satisfy the requirements of the Proposed Rule makes it difficult to understand the implications for protection of PII. The SAB recommends that the EPA develop specific definitions of terms and methods for meeting the requirements.

The Proposed Rule should build on techniques and practices to protect human data when data and models underlying pivotal regulatory science are made available to the public. Some individual data (i.e., data associated with individuals in a sample) used in epidemiological studies are held by federal agencies such as the Centers for Disease Control and Prevention or the Department of Health and Human Services (Medicare data), while other data have been developed by state, local or tribal governments, academic institutions or private organizations, among others. Some federal agencies have efficiently developed methods for making data available to the public (e.g., Census Bureau, CDC). Currently, no comparable system exists for datasets that are owned by non-federal governments (e.g., states, tribes), and/or owned by private societies/organizations or academic institutions, which are themselves protected by strong privacy and confidentiality requirements through their Institutional Review Boards (IRBs).

It seems reasonable that the standards applied by the EPA to protect sensitive data and copyrighted or confidential business information should be the same as the standards applied by editors of reputable scientific journals (e.g., guidance from the International Committee of Medical Journal Editors). Techniques and practices such as microaggregation to protect sensitive data have been developed by researchers involved in studies with human subjects, but such an approach may make the data unsuitable for modeling. The SAB recommends that EPA collaborate with other federal agencies to make individual-level data available through the system of Federal Statistical Research Data Centers, which are already widely used by the Census Bureau to allow researchers to gain access to individual data while protecting the data from public dissemination. The SAB also notes that there are techniques and practices to protect sensitive data that have been well-developed by researchers involved in studies with human subjects. The proposed regulation should clearly address the issue of obtaining public access to datasets while maintaining the privacy of the participants and confidentiality of the data, because without such access, sensitive data and confidential business information could be excluded entirely from consideration as pivotal regulatory science.

The Proposed Rule states that when promulgating significant regulatory actions, the Agency shall ensure that data and models underlying pivotal regulatory science are publicly available in a manner sufficient for independent validation. The SAB finds that greater clarity is needed in definitions of "data and models" and "pivotal regulatory science." EPA's Supplemental Notice of Proposed Rulemaking has indicated that the requirement to make data and models underlying pivotal science available in a manner sufficient for independent validation would apply to all data and models, not just dose-response models. The definitions provided in the Proposed Rule are not adequate and may be better supported in the context of a guidance document that includes realistic examples of the types of data and models of interest and the requirements for reporting this information. The SAB notes that this regulation could benefit from use of the term "analysis dataset" to define data that should be made publicly available.[6] This term refers to data that have been collected and processed (e.g., cleaned and transformed) for analysis. A technical issue to be considered is how to separate datasets and models that were the basis of calculations used to

[6] The SAB notes that protecting privacy and confidentiality must be taken into consideration when data and models underlying pivotal regulatory science are made available to the public. Requirements for protection of privacy have been established under the Health Insurance Portability and Accountability Act (HIPAA). Although the proposed Rule suggests that privacy and confidentiality can be addressed through anonymization or de-identification, even de-identified datasets present risks of re-identification (Rocher et al. 2019).

drive the quantitative assessment from ancillary data and models that were part of the weight of evidence.

The Proposed Rule indicates that the Administrator may grant exceptions to the requirements of the Proposed Rule on a case-by-case basis if it is determined that compliance is impracticable because: (1) it is not feasible to ensure that all pivotal regulatory science is publicly available for independent validation in a manner that is consistent with law; protects privacy, confidentiality, and confidential business information; and is sensitive to national and homeland security; or (2) it is not feasible to conduct independent peer review of all pivotal regulatory science used to justify regulatory decisions. The SAB recommends that the EPA develop specific criteria for such exceptions as part of the Final Rule. Although it will be difficult to develop criteria for exceptions, outlining such criteria would benefit EPA and help ensure that the principles of transparency outlined in the Proposed Rule are accomplished. Case-by-case exceptions without criteria may create public concerns about inappropriate exclusion of scientifically important studies. A framework and/or guidance document could also help EPA clarify how the rule will affect current scientific review procedures. It might be useful for the EPA to consider recommendations from a scientific and advisory committee when making waiver decisions.

In order to assess the feasibility of making data and models available in a manner sufficient for independent validation, some critical elements should be considered (e.g., how to treat studies that are formatted in a manner that make the data difficult to share, how to move forward if laboratories refuse to collate and release data, how to handle sensitive information such as individual participants' addresses, how to manage international studies, and how to manage conclusions drawn from meta-analysis). Experimental considerations must be made known to determine whether data are valid (e.g., the appropriateness of controls, protocols employed, limits of quantification). Assessing the validity of epidemiological studies for the purposes of the Proposed Rule could pose scientific and technical challenges. Important issues to be addressed include understanding bias, confounding factors, measurement errors and exposure characterization. All these factors play a role in defining what would be appropriate for data access and study validation purposes. The SAB encourages EPA to consider these questions and define "independent validation" in the Proposed Rule because this definition drives the feasibility of whether the EPA can make data and models available for validation. The SAB recommends that EPA develop a guidance document to clarify these issues and how the requirement would be managed.

The requirement in the Proposed Rule that "data" be made publicly available is vague and, as a result, can be interpreted in different ways. If "data" includes all machine output or individual data sheets on study participants associated with analysis the requirement would create demands on researchers that could impact the science-based decision-making process. The Proposed Rule should follow evolving norms developed by the scientific community as well as federal agencies (e.g., National Science Foundation, National Institutes of Health, National Oceanic and Atmospheric Administration, Department of Energy). The EPA should consider seeking input from experts in library science, data curation management and data retention to identify best practices and tools to ensure efficiency and utility of data that are made available.

There will be costs associated with assessing and disseminating data as required by the Proposed Rule. Funding agencies may have different time limits for retaining data. Historical datasets might not be available at the level of detail needed for recalculation. Some of the data or

computational methods may have been discarded if they were deemed not necessary to maintain. The SAB suggests that the EPA consider establishing an office (or virtual office) on data sharing, and a peer review panel or workgroup to assist EPA in this process (e.g., American Society for Testing and Materials (ASTM) and the International Organization for Standardization (ISO) have workgroups and approaches to establish valuable consensus standards). This group could identify: standard data formats (data templates); how to report methods/procedures used; uncertainty; and when and how to implement greater data protections for PII/CBI. The SAB notes that processing and documenting data and models developed prior to the effective date of the rule will pose challenges. A possible way to address this is to apply rule requirements only to information developed after the effective date of a Final Rule. Standards on transparency are evolving, and modern expectations do not apply to studies completed 10 or 20 years ago. It is reasonable to apply modern standards of transparency and public availability to current and future studies, but it will not always be possible to apply these same standards retrospectively.

It is difficult to develop a definition of "data" that would meet EPA's objectives in proposing this rule. The definitions of data would likely differ based on the available dataset and the types of data accumulated. However, the SAB recommends the development of definitions to clarify the requirement to make data available. "Data" should not be confused with personally identifiable data. Extensive work would be required, across a diversity of fields, data types and data of different ages, to understand the implications of adopting different definitions of data. More clarity is also needed to define the nature of the "data" that must be publicly available. As previously noted, this regulation could benefit from using the term "analysis dataset" to define data that must be made publicly available.

The Proposed Rule requires the EPA to describe and document any assumptions and methods that pertain to the use of data and models underlying pivotal regulatory science and to describe variability and uncertainty. High quality scientific studies identify the assumptions used in models, methods used, the variability of the replications, and any other confounders that add to the uncertainty of the final dataset, so these are not unusual or inappropriate factors to address. However, certain scientific and technical challenges must be surmounted. One would anticipate variability in the reporting across laboratories; therefore, the SAB strongly suggests that, before the implementation of the Final Rule, the EPA develop a guidance document pertaining to assumptions, methods, variability, the definition of data and uncertainty.

2. INTRODUCTION

As part of its statutory duties, the EPA Science Advisory Board (SAB) may provide advice and comment on the scientific and technical basis of certain planned EPA actions. The Environmental Research, Development, and Demonstration Authorization Act of 1978 (ERDDAA) requires the EPA to make available to the SAB proposed criteria documents, standards, limitations, or regulations provided to any other federal agency for formal review and comment, together with relevant scientific and technical information on which the proposed action is based. The SAB may then provide advice and comments on the scientific and technical basis of the proposed action.

At its May 31, 2018 public meeting, the chartered SAB identified the Proposed Rule titled *Strengthening Transparency in Regulatory Science* as a planned action that merited review. In April 2019, the SAB Work Group on Planned Actions for SAB Consideration of the Underlying Science recommended that the SAB review the Proposed Rule and at its public meeting on June 5-6, 2019, the SAB elected to review the scientific and technical basis of the Proposed Rule. Subsequent to the June meeting, a work group of chartered SAB members was formed to carry out the review. Members of this work group then took the lead in SAB deliberations on this topic at a public teleconference held on January 21, 2020. On March 3, 2020, EPA announced a Supplemental Notice of Proposed Rulemaking (SNPRM),[7] which contains additional information and clarification of certain terms and provisions of the Proposed Rule. This report provides the SAB findings and recommendations related to the Proposed Rule and aspects of the supplemental notice.[8]

[7] Strengthening Transparency in Regulatory Science; Supplemental Notice of Proposed Rulemaking, 25 Federal Register (18 March, 2020), pp. 15396-15406.
[Available at: https://www.federalregister.gov/documents/2020/03/18/2020-05012/strengthening-transparency-in-regulatory-science]

[8] Drs. John Graham and Donald van der Vaart indicated that they do not concur with this report. Their dissenting opinions are included in Appendix A.

3. SAB ADVICE AND COMMENT ON THE PROPOSED RULE

The SAB has reviewed EPA's proposed rule titled *Strengthening Transparency in Regulatory Science* (Proposed Rule) and provides the following comments on the scientific and technical basis of the proposed action. The SAB also provides recommendations to strengthen the science informing the Proposed Rule.

3.1. General Comments

The Proposed Rule is intended to strengthen the transparency of EPA regulatory science by providing that, for the science pivotal to significant regulatory actions, EPA will ensure that the data and models underlying the science are publicly available in a manner sufficient for validation and analysis. The SAB recognizes the importance of this rule and its purpose, establishing transparency of the scientific information used for significant regulations and enhancing public access to scientific data and analytic methods to help ensure scientific integrity, consistency and robust analysis.[9] Strengthening transparency by improving access to data can lead to an increase in both the quantity and the quality of evidence that informs important regulatory science and policy decisions. The scientific community is moving toward adopting the precept of sharing accurate data and information to increase credibility, high-quality outcomes and public confidence in science. The SAB supports the adoption of this precept.

The SAB recognizes that the long-term trend in most scientific fields is for authors to supply public access to data and analytic methods after scientific findings are published. Such transparency helps to ensure scientific integrity and facilitate robust analysis, as well as allow supplementary lines of knowledge to be developed from the same data. Enhancing the transparency and validity of the scientific information relied upon by EPA and increasing public access to data are worthy goals. However, the SAB finds that key considerations are not present in the Proposed Rule or presented without analysis and explanation of scope. In addition, certain key terms and implementation issues have not been adequately defined or described. The development of additional policy and/or guidance documents is strongly encouraged to provide clarity on the procedures for conducting the proposed efforts. The SAB also has concerns about the scientific and technical challenges of implementing some requirements of the Proposed Rule. The SAB has provided recommendations to facilitate implementation of the Proposed Rule.

3.2. Requirement to Identify All Studies and Regulatory Science Supporting Final Agency Actions

The Proposed Rule requires the EPA to clearly identify all studies (or other regulatory science) relied upon when it takes any significant final agency regulatory action and make such studies available to the public to the extent practicable. The EPA's Supplemental Notice of Proposed Rulemaking expands the scope of this requirement to include studies relied upon in influential scientific information (i.e., scientific information that will or does have a clear and substantial impact on important public policies or private sector decisions). This requirement could be complex and/or impractical because some studies could be considered when integrating evidence but not directly used to determine specific regulatory standards or levels.

[9] One SAB member does not agree with this statement, particularly in the context of international studies and those developed before the acceptance of current standards on science transparency that are discussed in this report.

The lack of criteria for satisfying the requirement to "make all such studies available to the public to the extent practicable" makes it difficult to understand the implications of the requirement. Criteria are needed to define the requirement. A question to be answered is whether making the scientific papers reporting these studies available without charge makes the studies "available," or whether all data from every measurement taken as part of the study need to be available to anyone to analyze. At one end of this range of interpretation the requirement is easily implementable. On the other end of the spectrum, meeting the requirement would be enormously expensive and time consuming at best and could be expected to result in the exclusion of much of the scientific literature from consideration (the machine data may no longer be available and/or the researchers may no longer be alive or in a position to assemble the data). The net effect could be minimal or complex.

The SAB recommends that the Final Rule describe in greater detail and clarity how the requirement can be met. To effectively implement the requirement, at the minimum, pivotal regulatory science and regulatory science (as defined in the Proposed Rule), as well as other types of data necessary to make regulatory determinations, should be publicly available with a scientific justification explaining why they were used for regulatory decisions. As stated in the Proposed Rule, EPA should indicate whether an independent peer review of all pivotal regulatory science has been conducted. As discussed below, the proposed rule provides for exceptions to be granted if it is not feasible to implement these requirements.

Scientific and technical challenges of making all studies supporting regulatory actions available to the public

The SAB finds that requiring the identification of all studies and regulatory science supporting regulatory actions and making them available for reanalysis will be a complex process.[10] As previously discussed, identifying and making "pivotal science data or studies" available could present challenges if some studies were considered in the regulatory decisions but were not used to determine the point of departure (POD) or reference dose (RfD) or other regulatory/technical level. It is not clear how much information and which studies should be included in the requirement to identify and make studies available. This should be clarified in the Proposed Rule. As further discussed in Sections 3.2 and 3.3 of this report, if the intent of the rule is to make available for reanalysis "all underlying pivotal science supporting influential scientific information and/or pivotal regulatory science" that contributed to the ultimate regulatory decision, then practical procedures must be established for an independent validation.

EPA must also make certain that personally identifiable information (PII) and confidential business information (CBI) are not available to persons/groups not properly vetted, approved, and trusted by those owning the CBI information. Without adequate protection, industry data generated by one company can be used by other companies to fulfill regulatory requirements in other geographies. It would be beneficial for the EPA to develop some specific policies related to: the protection of PII and CBI, exceptions that would be appropriate where CBI cannot be released, and whether data compensation should be considered. When PII and CBI data or methods are made available, "the public" receiving the information should be a small group of

[10] The level of complexity will depend upon the studies and information that must be identified and made available for reanalysis.

people who have provided assurance that they will keep such information confidential and protected. EPA could consider developing tiers of public access that may provide a base level of information (e.g., a robust summary or final study report which is devoid of confidential protected information) for the general user or member of the public and then restrict the availability of additional information to a smaller group when access to confidential or copyrighted information is warranted. In matters of public health, it is common to require disclosure to a trusted third party. In the case of environmental epidemiological studies, microaggregation of data can be used to protect personal identity. The SAB notes that the supplemental proposal indicates the Agency will only use studies containing CBI and PII if there is tiered access to the data or if the data can be sufficiently de-identified.

It may not be practical for EPA to make all studies or other regulatory science used in a final regulatory action available to the public. Many studies are included as part of a regulatory evaluation, but some studies or data may not drive the final regulatory decision (e.g., because of insufficient sample size or doses that are too high). The EPA could consider producing a list of study data considered in an evaluation and then strive to provide data for critical studies driving regulatory limits (e.g., "pivotal studies"). The SAB notes that the Proposed Rule could benefit from use of the term "analysis dataset" to define data that should be made publicly available.[11] This term refers to data that have been collected and processed (e.g., cleaned and transformed) for analysis. In this way, the public could understand which studies/data were considered and used without expending the resources to gather and disseminate all available data. The level of detail required to allow the public to transparently reach the same conclusion(s) as the Agency will differ among individuals who seek the data. Some people may wish to verify that the EPA has selected the right "pivotal studies." However, a more expansive scope would increase the reporting burden in a way that may make the Proposed Rule untenable.

The scientific and technical challenges of making pivotal studies supporting regulatory actions available to the public consist of: (1) having EPA be explicit about which studies are pivotal to the recommended regulatory action (e.g., using a decision and risk analysis framework that explicitly derives recommended actions from study results, and thus enables the roles of studies in the regulatory actions to be precisely identified); and (2) making the data and models for the underlying pivotal studies publicly available. The first step may be relatively technically straightforward to implement if standard decision and risk analysis frameworks are used to derive policy recommendations from study results. However, there are a number of emerging technologies (e.g., use of transcriptomics data) that raise new challenges, including dealing with "big data" (i.e., extremely large datasets to be analyzed computationally). It may also be very challenging to identify pivotal studies if holistic judgments and weight-of-evidence frameworks are used. Likewise, the second step, making available the data and models underlying pivotal scientific studies, is also technically straightforward when the pivotal studies already provide the analysis dataset and document the models used to analyze it and to reach their conclusions.[12]

[11] The SAB notes that protecting privacy and confidentiality must be taken into consideration when data and models underlying pivotal regulatory science are made available to the public. Requirements for protection of privacy have been established under the Health Insurance Portability and Accountability Act (HIPAA). Although the proposed Rule suggests that privacy and confidentiality can be addressed through anonymization or de-identification, even de-identified datasets present risks of re-identification (Rocher et al. 2019).

[12] Some SAB members are concerned that the Proposed Rule contains contradictory statements about whether its effect would be to prohibit EPA from considering studies for which underlying data cannot be made publicly available. Major public health protections, like the National Ambient Air Quality Standards (NAAQS) for ozone and

Typically, an analysis dataset lists values of exposure, response, and covariate variables and uses multiple imputation and related methods for missing data and confidentiality (Reiter and Raghunathan, 2007). Such datasets have already been made publicly available in studies such as NHANES III (National Center for Health Data Statistics 2019). Publishing such analysis datasets and the models used to analyze them to produce conclusions that are pivotal to regulatory action seems to be an appropriate scope of coverage for the stated goal, i.e., "to increase transparency in the preparation, identification, and use of science in policymaking." [13]

It may not be feasible to identify and make available data in epidemiological studies that arise from small datasets or targeted geographic areas, especially if the Informed Consent Form indicated that only the particular researchers who conducted the study would have access to the information and data. If the participants agreed to grant only a select group of researchers access to their personal information, then that consideration should be respected, and such information should not be supplied to additional people for validation. It would probably be impractical, to go back to the participants and request their approval to provide additional people access to personal information. Therefore, the identification of "analysis dataset" when referring to data (i.e., analysis dataset is data that have been collected and processed, cleaned and transformed for analysis) could be useful in this regard as it does not contain personal identifiers.

Some of the major cohort studies, such as the Women's Health Initiative (WHI) (National Heart, Lung, and Blood Institute, 2020) or the Atherosclerosis Risk in Communities (ARIC) Study (ARIC Investigators, 1989), require researchers to write a manuscript proposal, study design or protocol – in effect, a document describing the study they intend to conduct, including data that are required and the rationale behind the study (Prentice et al., 1998; ARIC Investigators, 1989). The proposal is reviewed by a committee and, in most cases, is either approved without modification or returned to the investigators to address specific issues. It is possible that the EPA could work with holders of private datasets to develop such study designs or protocols or a similar system of broader applicability. That approach would provide a mechanism for interested researchers to access datasets for reanalysis under appropriate controls where relevant for EPA regulations. The SAB encourages the development of data center protocols to address procedures regarding data availability for independent validation.

3.3. Requirement to Ensure that Data and Models Underlying Pivotal Regulatory Science are Publicly Available in a Manner Sufficient for Independent Validation

The Proposed Rule states that when promulgating significant regulatory decisions, the Agency shall ensure that data and models underlying pivotal regulatory science are publicly available in a manner sufficient for independent validation. The supplemental proposal (Revised §30.3) has indicated that this requirement would apply to "data and models, underlying pivotal science supporting influential scientific information and/or underlying pivotal regulatory science," not

oxides of nitrogen, drinking water standards for arsenic, and Toxic Substances Control Act (TSCA) standards for formaldehyde emissions in composite wood rely on studies that include confidential data.

[13] One SAB member suggests that, to accomplish this, the EPA should not fund new research unless researchers file a research protocol, provide analysis datasets to EPA and the public, and provide analysis code. The SAB member also suggests that previous research should not be considered to support regulatory actions unless researchers provide sharable datasets, analyze them, provide analysis code, and file these materials with the EPA. The SAB notes, however, that institutional review boards at host universities may not agree with the requirement that data be deposited with EPA as a condition of funding.

just dose response data. The SAB recommends that the EPA clarify the definitions of "data and models" and "pivotal regulatory science." It would be useful to develop a guidance document that includes examples of the types of data and models of interest and requirements for reporting this information. It would be particularly useful to clarify specific requirements for reporting information from animal toxicity and/or environmental epidemiology studies. The Proposed Rule indicates that pivotal regulatory science refers to studies or analyses that drive the requirements and/or quantitative analysis of EPA final significant regulatory decisions. Although this definition is adequate for some practical purposes, the EPA should clarify whether this includes all the hazard characterization and dose response models that the Agency evaluated and captured in its analysis or only the final model ultimately selected.[14] As previously discussed, data might be defined as the "analysis dataset" (i.e., listing values of causally relevant exposure, response, and covariate variables, and uses of multiple imputation and related methods for representing missing data and confidential data). Dose-response model might be defined as "the model or algorithm used to calculate conditional probability of a stated health response caused by stated exposures together with the values of any other direct causes of the response (e.g., stated levels of causally relevant covariates such as co-exposures and co-morbidities)." The SAB notes that the Final Rule should define and clarify when the requirements are applicable to animal toxicity studies or environmental epidemiology studies.

When defining these terms, the following questions should be considered with regard to dose-response data. Are there any requirements for the number of dose levels or dose spacing? Does this mean that single dose studies will be excluded? If not, under what circumstances would these studies be included? What types of models are "in scope?" What type of information is needed for each model type – animal toxicity or epidemiological data? Is the goal to provide equations, allowing the public to replicate the math or to provide models with assumptions that the public can evaluate in detail? Are there standard approaches for modeling dose-response relationships (benchmark dose? other?) If so, under what conditions are different dose-response approaches implemented? Can a framework be prepared to outline the EPA's approach? To make the Proposed Rule more practical, pivotal regulatory science could be defined as the study (or studies if necessary) on which the regulatory limit is based. Ideally, the EPA should provide a brief explanation of why the selected study is the pivotal study and why other studies were not selected.

Protecting privacy and confidentiality

As previously discussed, protecting privacy and confidentiality must be taken into consideration when data and models underlying pivotal regulatory science are made available to the public. Although the Proposed Rule suggests that privacy and confidentiality issues can be addressed through anonymization or de-identification, this is not always the case. The U.S. Office of Management and Budget (OMB, 2013), Health Effects Institute[15], National Academies of

[14] A technical issue to be considered when defining models and data is how to separate datasets and models that were the actual basis of the calculations used to drive the quantitative assessment, e.g., point of departure (POD) or reference dose (RfD), from additional datasets and models that may have been ancillary but part of the weight of evidence used in the regulatory action. Reasons for the choice of the primary datasets and models for the quantitative assessment and reasons for placing the other options in the weight of evidence (WOE) category should be made available.

[15] Letter from Daniel S. Greenbaum, Health Effects Institute, to Lek Kadeli, Environmental Protection Agency (Aug. 27, 2013) ("noting increasingly granular data used in health studies and stating that "these characteristics -

Science (NRC, 2005), and independent experts (Rothstein, 2010; Commission on Evidence-Based Policy Making, 2017; Rocher et al., 2019) have all found that even de-identified datasets present significant risks of re-identification given modern techniques for combining these datasets with other sources of individual information (also known as a "mosaic effect"). Where risk of personal identification is high, the SAB recommends using approaches such as independent analysis by a third party, e.g., Health Effects Institute (HEI).

Scientific and technical challenges and feasibility of making data and models publicly available in a manner sufficient for independent validation

The SAB notes that a number of specific scientific and technical challenges must be addressed, and questions answered, in order to make data and models available in a manner sufficient for independent validation. There are experimental considerations that must be known in order to determine whether experimental data are valid (e.g., appropriate controls, protocols employed, where data fall within the standard curve of the target analyte, limit of quantification, dynamic range of the instrument, calibration of instruments, condition of experimental animals, blinding of reading of slides or behavioral observations, stability of samples, qualifications and approach of researchers obtaining epidemiological data). Also, there is a need to know what information was used in models (e.g., whether all data points were used, how confounders were handled in epidemiological studies, what statistical models were used, what predictive models were used, the fit of the model to the data). Independent validation requires sufficient information about how the original data were collected and analyzed in order to know whether the validation procedures are likely to yield the same result as the original calculations.

Therefore, it would be useful for the EPA to specifically define "independent validation." EPA's supplemental proposal indicates that independent validation means the reanalysis of study data by subject matter experts who have not contributed to the development of the study to demonstrate that the same analytic results reported in the study are capable of being substantially reproduced. However, the specific definition of independent validation drives the feasibility of whether EPA can make data and models available for independent validation. For example, the EPA should consider the following questions: How much information is sufficient for independent validation? Will this information consist of equations where reviewers can verify the math, more detailed models where assumptions and limitations are described, or code to allow the public to evaluate and run the models if desired? Is this information simply the dose-response data for the endpoint of concern driving a regulatory limit, or is it availability of all data from a pivotal study to allow reviewers to examine the potential contributions of other variables on the primary endpoint of concern? Endpoint data are seldom evaluated in isolation so providing sufficient study information to allow an independent assessment seems important to meet the goals of the Proposed Rule. For example, an effect on pup body weights in a toxicology study should be examined with knowledge of maternal gestational body weight gains, litter size, food consumption, maternal/litter clinical signs, etc. Sample size and variability also play a key role in data interpretation. The SAB notes that EPA Data Quality Guidelines include definitions of replication and validation; this information must be included in the "analysis dataset" with guidance on how to read/interpret it.

---

which have in general enhanced the quality and the sensitivity of the studies - increase the difficulty of providing a fully "de-identified" dataset while also enabling a different investigator to conduct a full replication and sensitivity analysis of the original study results.")

It may be beneficial for EPA to develop a guidance document with case studies based on past risk assessments to clarify some scenarios and how the requirement to make data and models publicly available in a manner sufficient for independent validation would be managed (i.e., what is 'in scope' vs. 'out of scope' and rationale). A guidance document may offer clarity on the implementation of this rule. The SAB finds that the following specific questions should be answered or addressed in a post-rule guidance document in order to assess the technical feasibility of making data and models available in a manner sufficient for independent validation.

1. Good laboratory practice (GLP) studies include full disclosure of study data, statistics, instrument calibration, positive control data, etc. Would the EPA make all of these data available to the public for any study judged to be "pivotal?" Would this include GLP audit findings? Will any information in the study be judged by EPA to be CBI?

2. With respect to non-GLP/investigational studies, laboratories may not have data formatted in a manner that makes these data easily shared with the public. How will EPA manage this, particularly when several years may elapse between conduct of a study and its designation as a "pivotal study?"

3. How will the EPA manage and release laboratory data for public review? Will these data be used in risk assessment? Approaches need to be developed to avoid potential bias in risk assessments or regulatory decision-making.

4. How will the EPA manage international studies where there is no requirement for laboratories to provide data to EPA? EPA may have little to no leverage in these situations. Again, the inability to use "pivotal studies" from other geographies could bias risk assessment decisions.

5. How will the EPA manage conclusions drawn from a meta-analysis? Do all studies included in the analysis become "pivotal" studies? This could markedly increase the number of datasets that must be publicly available.

6. How will the EPA justify identification of a "pivotal study" and dose-response without clarifying why other studies were not pivotal? This will require transparent evaluation and reporting of data quality/reliability for available studies that allows reviewers to understand EPA's selection of the "pivotal" study.

7. Some data and models may rely on proprietary software that may not be readily accessible or available and scientists may need to develop their own proprietary code, while other software may be accessible but require considerable data storage or download that may limit utility and availability. How will these models be validated?

Assessing the validity of epidemiological studies

The SAB finds that assessing the validity of epidemiological studies for the purposes of the Proposed Rule poses particular scientific and technical challenges. In general, for the purposes of the Proposed Rule, validation should be defined to include both internal validity and external validity, in the senses defined by Campbell and Stanley (1963). Issues to be addressed include

understanding bias, confounding factors, measurement errors and exposure characterization. All of these factors will play a role in defining what would be appropriate for access and validation purposes. Specifically, one would need the following information: how measurements were taken; how confounders were assessed and dealt with; the institutional review board (IRB) application and subsequent approvals or concerns; the Informed Consent Form (or assent process for children capable of providing assent) or the consent of parents or guardians for information collection on children; the qualifications of the researchers obtaining personal or health information and the consistency of multiple researchers for collection of PII; how truncated datasets for longitudinal studies were handled when participants dropped out of the study or missed sampling times; how environmental samples or human blood samples were taken, handled, stored and analyzed; criteria for how any data points were deemed outliers and eliminated from the dataset; how participants were selected and what the selection/exclusion criteria were. As previously discussed, to address some of these issues, the SAB recommends using approaches such as conducting independent analysis by a third party, e.g., Health Effects Institute (HEI), where risk of personal identification is high, and/or adopting the term analysis datasets as described below.

Challenges and costs of processing and documenting of data prior to public release, maintenance and administration of datasets so they are publicly available, and handling historical datasets

There will be technical, scientific and resource challenges associated with assessing and disseminating data as required by the Proposed Rule. Costs of processing and documenting data will be difficult to assess in advance until EPA has developed a system for dealing with the requirements of the rule. However, the SAB notes that the Agency should consider seeking input from experts in library science, data curation management and data retention to identify best practices and tools to ensure efficiency and utility of data that are made available.

Obtaining data in a useable format with adequate documentation may be difficult. Funding agencies may have different time limits for retaining data. It is likely that not all data will be in a format suitable for public data sharing. Data problems may include: non-traditional data formats, PII, CBI, and inadequate documentation of data or methods. The SAB notes that there may be solutions for some of these data issues, but not for others. Historical datasets might not be available at the level of detail needed for recalculation. Some of the data or computational methods may have been discarded if they were deemed not necessary to maintain. IRB applications usually indicate when individual records can be discarded.

The people or groups processing and handling the data (EPA staff or independent non-EPA consultants) would need to be identified, their credentials and any conflicts of interest with the particular case identified, and documentation secured that they will not reveal confidential information without appropriate permission from the owners of the CBI or PII. The processing might include ways to strip some of the PII of potential identifying information or aggregating the information if these methods would still allow for the validation to be performed. Similarly, the people or groups who would maintain the datasets and provide them to independent public people or groups for validation processes would need to be identified and provide assurance that the data would be maintained as confidential PII or CBI and only released to authorized people or groups. A mechanism would need to be developed for public requests for data access and for approval or disapproval of the requests.

The SAB suggests that EPA consider establishing an office (or virtual office) on data sharing and a peer review panel or workgroup to assist EPA in this process. For example, the American Society for Testing and Materials (ASTM) and the International Organization for Standardization (ISO) have workgroups and approaches to establish valuable consensus standards. This group could provide advice on: standard data formats (data templates), how to report methods/procedures used, uncertainty, when and how to implement greater data protections for PII/CBI, etc. It would be beneficial to build experience and expertise in a group charged to meet this goal. This office could work directly with laboratories and researchers to provide the necessary information in a "user friendly" format. This office also could build and manage data archives and pursue critical historical datasets if deemed important. There will be costs associated with the establishment of such an office as well as researchers' time to collate data and work with EPA to make these data publicly available. It is unclear how the EPA will manage these additional costs. In the future, it might be possible for EPA to develop a reporting framework for laboratories so that study data are collected in a format that requires less rework if a study is subsequently judged to be a "pivotal study." Some laboratories/researchers may not want to organize historical data for public release as they may see this activity as a diversion from their research priorities. There would likely also be additional costs that occur at an institutional level (i.e., Institutional Review Boards) that would be substantial. The EPA may need to find creative ways to offset the expense associated with data submission for pivotal studies.

Processing and documenting data and models developed prior to the effective date of the rule

The SAB notes that processing and documenting data and models developed prior to the effective date of the rule will pose challenges. It is likely that some flexibility is going to be required as the standards on transparency are evolving, and data collection expectations do not apply to historical studies or investigations completed 10 or 20 years ago. It is reasonable to apply modern standards of transparency and public availability to current and future studies, but it will not always be possible to apply these same standards retrospectively.

It is not clear: (1) how many of the studies EPA currently relies upon to take important regulatory actions would meet the public disclosure standards stated in the Proposed Rule, and (2) whether EPA has assessed the feasibility of making underlying information from the studies publicly available, or what the impact of precluding those studies would be on EPA's decision making and its ability to protect public health/environment. The SAB could advise EPA on how to use data from historical studies in regulatory decisions. This may require case-by-case approaches.

Prospective or retrospective application of provisions for ensuring the public availability of data and models underlying pivotal regulatory science

As previously noted, the retrospective application of modern transparency standards is a challenge. A large amount of work would be required to locate, curate and retrospectively make datasets available for public access. This requirement could adversely affect the ability to move this program forward in a meaningful capacity. EPA could decide not to apply the Proposed Rule and its specific requirements retrospectively given the potential difficulty accessing, reviewing and making data available that were not originally intended to be disseminated in such a manner

as defined in a future rule. The EPA could consider designating a "start date" and begin collecting and releasing pivotal study data at that time. When the EPA updates an existing risk assessment after the start date, the EPA could collect and release pivotal data.

Feasibility of developing criteria to specify exceptions to the requirement to make information available to the public

The Proposed Rule indicates that the EPA Administrator may grant exceptions to the requirements of the Proposed Rule on a case-by-case basis if it is determined that compliance is impracticable because: (1) it is not feasible to ensure that all pivotal regulatory science is publicly available in a manner sufficient for independent validation, consistent with law, protects privacy, confidentiality, confidential business information, and is sensitive to national and homeland security; or (2) It is not feasible to conduct independent peer review on all pivotal regulatory science used to justify regulatory decisions. The SAB understands the need for exceptions and recommends that EPA develop specific criteria for such exceptions as part of the Final Rule. Although it may be difficult to develop criteria for exceptions, outlining such criteria would benefit EPA and help ensure that the principles of transparency outlined in the Proposed Rule are accomplished.[16]

The SAB notes that the proposal to use a case-by-case "waiver" may not be an effective mechanism for ensuring that the EPA can appropriately consider important studies, including those that rely on confidential data. Without pre-defined criteria for such waivers, a case-by-case waiver may create concerns about inappropriate exclusion of scientifically important studies. Reference to a vague "feasibility" standard suggests that such waiver decisions are to be made solely by the Administrator. In the absence of clear guidance, such waivers might appear to be inconsistent or lacking objectivity. A framework and/or guidance document could also help EPA to clarify how current scientific review procedures will be affected by this rule. It might be useful for the EPA to consider recommendations from a scientific advisory committee when making waiver decisions.

The SAB finds that exclusion of segments of the scientific literature, with the possibility of inclusion of other selected information without pre-defined criteria, could allow systematic bias to be introduced with no easy remedy. The proposed exception process applies no constraints on how this mechanism could be used or that it be restricted to the issue of confidential data. Such a proposal is inconsistent with the scientific method that requires all credible data be used to understand an issue and to allow systematic review to evaluate past research. It may be useful for the SAB to peer review documentation containing the mechanisms for exclusions based on criteria defined by EPA and provide constructive considerations.

Although it would be difficult to develop, the EPA could benefit from preparing a framework and/or guidance that outline criteria to specify exceptions. While the EPA cannot address all circumstances and scenarios that could limit data sharing, the SAB recommends that EPA explore and document some reasonably anticipated scenarios that could be described, perhaps

[16] Some members of the SAB find that conclusions from studies that cannot be independently verified and reproduced due to unavailability of data or models should not be treated as pivotal for regulatory action if the regulatory action is supposed to be based on independently verifiable and reproducible science – which might be construed as part of the definition of "sound" science.

with case studies from previous risk assessments (e.g., What information would fall in scope based on past risk assessments? What data would have to be released to support a given risk assessment? Is it feasible to release this information? Why or why not?).

3.4. Requirement to Make the Data Underlying any Rule Publicly Available

The SAB notes that there are legitimate legal, ethical, professional and financial reasons why researchers may be unable or unwilling to fully share "data" - including statutes protecting participant privacy, experimental protocols assuring confidentiality of data for human subjects, and (for past studies) issues related to degradation and custody of data. The EPA, the U.S. Office of Management and Budget, and scientific institutions have recognized these legitimate concerns, and recognized that such constraints on availability of data do not prevent studies from being verified in other ways - or preclude those studies from being considered in regulatory decisions.[17,18,19,20]

Under the Proposed Rule, EPA would require that data underlying any proposed rule be made publicly available. The SAB finds that the requirement in the Proposed Rule that "data" be made publicly available is vague and, as a result, can be interpreted in different ways. The term "publicly available" is also vague and should be better defined. If "data" includes all machine output associated with analysis it would create demands on researchers that would be very onerous and could significantly slow down science-based decision-making. Even if the "data" were accessible, making it publicly available in a useable form would be costly and could be of limited utility based on past experience of the scientific community relative to the interpretation of the derivative data. If the data required to meet the "data" requirement is no more than the current standard of most journals (and in most cases provided in supplementary information) then the implications prospectively are minimal. Either way retrospective application of the requirement would be difficult to implement, could be expensive with no clear responsibility regarding who would cover the added costs, and could arbitrarily impact the conclusions drawn.

---

[17] U.S. EPA. 2016. *Plan to Increase Access to Results of EPA-Funded Scientific Research* states "some research data cannot be made fully available to the public but instead may need to be made available in more limited ways," but says the lack of full public availability "does not affect the validity of the scientific conclusions from peer-reviewed research publications." https://www.epa.gov/open/plan-increase-access-results-epa-funded-scientific-research

[18] U.S. EPA. 2002. *Information Quality Guidelines*: recognizes that sometimes "access to data and methods cannot occur due to compelling interests such as privacy, trade secrets, intellectual property, and other confidentiality protections." But where that is the case "EPA should, to the extent practicable, apply especially rigorous robustness checks to analytic results and carefully document all checks that were undertaken." https://www.epa.gov/sites/production/files/2019-08/documents/epa-info-quality-guidelines_1.pdf

[19] U.S. Office of Management and Budget. 2002. *IQA Guidelines*: "[making the data and methods publicly available will assist in determining whether analytic results are reproducible...the objectivity standard does not override other compelling interests such as privacy, trade secrets, intellectual property, and other confidentiality protections... where public access to data and methods will not occur due to other compelling interests, agencies shall apply especially rigorous robustness checks to analytic results and document what checks were undertaken."

[20] National Academy of Sciences. 2018. Letter to Acting Administrator Wheeler from Marcia McNutt, President of the National Academy of Sciences, C.D. Mote, Jr., President of the National Academy of Engineering, and Victor J. Dzau, President of the National Academy of Medicine 2 (July 16, 2018) stated that "Individual study quality should be evaluated on the basis of information that is available in standard journal articles, such as the study design elements, analytical techniques, and statistical methods. Researchers may be contacted to answer questions about the conduct of the study or be asked to provide additional data. If the study data are not available, their absence may affect how the study is rated and used in the analysis, but the study should not necessarily be eliminated from the assessment."

The Proposed Rule should be explicitly prospective and follow evolving norms developed by the scientific community as well as federal agencies (e.g., National Science Foundation, National Institutes of Health, National Oceanic and Atmospheric Administration, Department of Energy). The Proposed Rule should state and/or compare its objectives with existing and evolving federal procedures to address the underlying purpose of the Proposed Rule, which is to increase transparency of scientific studies utilized in regulatory actions. This additional explanation will enhance the characterization of the proposed regulation and will help EPA meet its objectives effectively and efficiently.

There appears to be consistency among analyses of how to address transparency that are orthogonal to the Proposed Rule. There is minimal justification provided in the Proposed Rule for why existing procedures and norms utilized across the U.S. scientific community, including the federal government, are inadequate, and how the Proposed Rule will improve transparency and the scientific integrity of the regulatory outcomes in an effective and efficient manner. It is plausible that in some situations, the Proposed Rule will decrease efficiency and reduce scientific integrity, determining if in fact that will be the case requires a thorough and thoughtful examination that is currently absent in the Proposed Rule. Moving forward with altered transparency requirements beyond those already in use, in the absence of such a robust analysis, risks serious and perverse outcomes.

Definition of data

The term "data" was not defined in the Proposed Rule and the nature of the data that must be made publicly available to fulfill the requirements of the Proposed Rule is not clear. It is difficult to develop a singular definition of "data" that would meet EPA's objectives in the Proposed Rule. The SAB notes that EPA has further defined "data"[21] in the supplemental proposal and, as previously discussed, the SAB finds that EPA could benefit from using the term "analysis dataset" (i.e., data that have been collected and processed for analysis) to define data. However, the definition of data would likely differ based on available datasets. For example, the original data for an in vivo study could include all the individual animal body weight or individual pathology data while a dataset for an in vitro study may include multiple samples and assays assessed, and for epidemiology data it may include individual exposure monitoring data or biological samples.

"Raw" data (also known as original or primary data) could include individual sample values collected on individual study subjects or various instruments and include each measurement on sampled endpoints (and each time it is sampled) in a given study. The original or "raw" data would not be manipulated in any fashion (e.g., removing outliers - data would appear in the raw data with a scientific rationale on why a data point was removed from subsequent analysis).

Notably, in 40 CFR part 792 -TSCA which describes Good Laboratory Practices (GLP), raw data is defined as "any laboratory worksheets, records, memoranda, notes, or exact copies thereof,

[21] Supplemental proposal definition of data: "Data means the set of recorded factual material commonly accepted in the scientific community as necessary to validate research findings in which obvious errors, such as keystroke or coding errors, have been removed and that is capable of being analyzed by either the original researcher or an independent party."

that are the result of original observations and activities of a study and are necessary for the reconstruction and evaluation of the report of that study." All or none of this original information may or may not be needed to adequately understand the results of the study or EPA's use of the information. Thus, making all this information publicly available may be resource intensive and provide limited utility. The SAB recommends that EPA use the term "analysis dataset."

The SAB notes that the requirement to make data available for public inspection will be more easily implemented for some datasets than others (e.g., GLP studies include individual sample data as part of standard reporting). The expectation that data and methods will be available for all endpoints may be unrealistic. EPA could consider designating scores for information availability with the highest tier assigned to studies that define methods (e.g., protocol) as well as raw (individual sample level) data. Additional scores could be driven by sharing data on other related endpoints in the pivotal study to facilitate data interpretation. Scores could subsequently be used to evaluate data utility, uncertainty, etc. Ultimately, there are some datasets where CBI and/or PII data will require more onerous steps to protect data confidentiality. Based on EPA responses to SAB questions, it appears that the Agency is seeking approaches to manage these data issues.

In defining data, it might be reasonable to consider the initial compilation of data (original data) in a spreadsheet as "raw" data. While this might still require some interpretation, e.g., abbreviations used in spreadsheet column headings, and footnotes about missing data points (if they were discarded because of legitimate reasons, such as known mistakes or statistically-determined outliers), the spreadsheets are still very close to the initial raw data coming out of an instrument or written by a researcher when working with a subject in an epidemiological study.

As previously stated, it is difficult to evaluate the impacts of a definition of data a priori. There is extensive work required to understand the implications of different definitions across a diversity of fields, data types and data of different ages. Such an effort is beyond the scope of what the SAB can undertake with the resources and time available. However, the SAB finds that such an analysis is foundational to the development of any transparency rule that goes beyond well-established norms and procedures.

Another aspect to consider is the practical aspect of actually conducting a reanalysis of a major epidemiological study. Such an enterprise requires an enormous amount of work even for a well-qualified researcher. The Health Effects Institute (HEI) established a model for conducting such a reanalysis in its 2000 reanalysis of the Six Cities and American Cancer Society datasets (HEI, 2000). However, HEI has not repeated this kind of exercise. EPA could consider using the HEI model for funding its own reanalysis for datasets that are deemed critically important for regulation.

3.5. Requirement to Describe and Document any Assumptions and Methods Pertaining to the use of Data and Models Underlying Pivotal Regulatory Science

The Proposed Rule requires the EPA to describe and document any assumptions and methods that pertain to the use of data and models underlying pivotal regulatory science and to describe variability and uncertainty. The SAB notes that high quality scientific studies identify the assumptions used in models and methods, the variability of the replications, and any other confounders that add to the uncertainty of the final dataset, so these are not unusual or inappropriate factors to be addressed. Specifically, it is good practice to identify: (1) testable

assumptions (e.g., that residuals in a regression model are normally distributed with constant variance); (2) results of tests of assumptions; and (3) any untested assumptions made in deriving conclusions from data (e.g., that there are no unmeasured confounders, or that dose-response functions are linear below the lowest dose for which data are available). Results of tests should be presented where they are available, and assumptions that have not been tested, or that have been tested but not supported, should be identified. Assumptions are often made about (a) error distributions for exposure estimates (most commonly, that they can be ignored); (b) model specification errors and uncertainties; and (c) causal interpretations of modeling results. These assumptions should be explicitly stated, and results of tests presented. Epidemiologists (e.g., Sander Greenland), statisticians, and risk analysts have written at length over several decades about how to test, validate, and document assumptions and methods for dose-response modeling and uncertainty and variability characterization, and these modern methods should be applied to make the factual and assumption-based foundations of pivotal studies as clear as possible. However, the SAB finds that there are scientific and technical challenges to be overcome and provides suggestions to implement this requirement.

Currently, in various chemical assessment processes, EPA program offices attempt to document the methods, assumptions, variability and uncertainty associated with the use of various dose-response models and data inputs utilized. EPA has generally done this in a qualitative format and should continue to refine and document this information. One would anticipate variability in the reporting of assumptions, methods, variability, and uncertainty across laboratories. Therefore, EPA could offer guidance on how to report these parameters. When this information is received from submitting laboratories, EPA could review the information to determine if methods, uncertainty, etc. are adequately addressed, and the Agency could follow up as needed. There are numerous resources available from which the EPA could develop guidance, including some defined by the EPA (Maurissen, 2010; U.S. EPA, 2012; U.S. EPA, 2019). The SAB notes that it may be difficult to make detailed information available to the public for some parameters like methods. This will vary depending on factors such as how new/novel the method is, how many variables impact outcome, etc. Many laboratories may not want to share standard operating procedures for public release by the EPA.

Dose response models

The following language in the preamble of the Proposed Rule has been interpreted differently by some members of the SAB:

> "…there is growing empirical evidence of non-linearity in the concentration-response function for specific pollutants and health effects. The use of default models, without consideration of alternatives or model uncertainty, can obscure the scientific justification for EPA actions. To be even more transparent about these complex relationships, EPA should give appropriate consideration to high quality studies that explore: A broad class of parametric concentration-response models with a robust set of potential confounding variables; nonparametric models that incorporate fewer assumptions; various threshold models across the exposure range; and spatial heterogeneity. EPA should also incorporate the concept of model uncertainty when needed as a default to optimize low dose risk estimation based on major competing models, including linear, threshold, and U-shaped, J-shaped, and bell-shaped models."

Some SAB members view this as a policy change indicating that evidence of nonlinearity should not be ignored when there is scientific evidence to support it. They note that EPA has explicitly identified low-dose linear models, along with others, in the list of models to be considered, and that support for alternative models is indicated in the EPA's 2005 Cancer Guidelines (U.S. EPA, 2005). On page 1-8 the EPA Cancer Guidelines state:

> "When there are alternative procedures having significant biological support, the Agency encourages assessments to be performed using these alternative procedures, if feasible, in order to shed light on the uncertainties in the assessment, recognizing that the Agency may decide to give greater weight to one set of procedures than another in a specific assessment or management decision. Encouraging risk assessors to be receptive to new scientific information, NRC discussed the need for departures from default options when a sufficient showing is made."

Additionally, in the National Academies 2009 Report *Science and Decisions* (page 207) (NRC, 2009) the Committee recommended that "EPA should continue and expand use of the best, most current science to support or revise its default assumptions." The Committee also identified several factors that EPA should take into consideration, including: "(1) the extent to which the current default is inconsistent with available science; (2) the extent to which a revised default would alter risk estimates; and (3) the public health (or ecologic) importance of risk estimates that would be influenced by a revision to the default."

Conversely, other members of the SAB view the language in the preamble of the Proposed Rule cited above as inappropriately codifying certain required scientific approaches into regulation. In support of this view, some members note that: (1) the preamble of the Proposed Rule asserts without providing any evidence that "there is growing empirical evidence of non-linearity in the concentration-response function for specific pollutants and health effects," and (2) the preamble does not acknowledge the value of default models (in this regard, these members also note that the National Academies' 2009 Report *Science and Decisions* (NRC, 2009) specifically recommended the use of non-threshold and linear models as a default for both cancer and non-cancer dose-response analysis). It is critical that the justification for alternative approaches is well documented and transparent, and some members of the SAB find that is not currently the case.

It is important that the EPA provide clarity as to the Agency's intent. Therefore, the SAB recommends that the preamble of the Proposed Rule include some specific examples or case studies where there is evidence in the published literature demonstrating threshold or other low-dose nonlinear responses which provide support for the preamble statements (e.g., Sweeney et al., 2009; Johnson et al., 2014; Cohen et al., 2016). The SAB notes that the EPA's 2005 Guidelines for Carcinogen Risk Assessment allow the Agency to consider nonlinear approaches after an analysis of available data under the guidance provided in the framework for mode of action analysis. Notably, where alternative approaches have significant biological support, and no scientific consensus favors a single approach, an assessment may present results using alternative approaches (i.e., a nonlinear approach). It may be useful for the Agency to consider

developing guidance or criteria delineating when data may be sufficient to apply alternative approaches. [22]

In general, it is difficult to define universal "rules of good behavior" for a many-faceted question such as choosing the right dose-response model. In some situations, direct biological arguments may support a particular model such as a linear or log-linear dose-response model. Where information on kinetics (e.g., saturable metabolism) as it relates to the dose-response is available, this might help guide selection of the most appropriate model. In other cases, the use of any model as a default choice may be inappropriate. The biostatistics literature includes many procedures for identifying and validating appropriate models, including techniques such as Bayesian Model Averaging which avoid the specification of a single model. There is no "one size fits all" approach that could be applied to all problems of this nature.

3.6. Protecting Sensitive Data and Copyrighted or Confidential Business Information

The Proposed Rule preamble states that nothing in the Proposed Rule compels the disclosure of any confidential or private information in a manner that violates applicable legal and ethical protections**.** The SAB provides suggestions and recommendations to protect sensitive data and copyrighted or confidential business information.

It seems reasonable that the standards applied by the EPA should be the same as the standards applied by editors of reputable scientific journals. A report of the National Research Council (NRC, 2003) discusses responsibilities of authors to share data, software, and materials related to their publications. Techniques and practices to protect sensitive data have been developed by researchers involved in studies with human subjects. Other federal agencies have utilized specific and explicit data transfer agreements with entities that sought access to protected information for the purposes of reviewing and running analysis on the dataset. The EPA could employ similar approaches. Any new rule needs to build on those efforts. Coding data to protect any personal identifiers is routinely done and may have already been done with the datasets under consideration. If aggregation of some of the data (e.g., by age bands) does not compromise the ability of the calculations to be done, then this would protect personal information. Existing methodologies and technologies already in widespread use such as those used in NHANES III (National Center for Health Data Statistics, 2019), including multiple imputation, can be used to provide protected access to data. Additional technologies worth considering include differential privacy techniques (Dankar and El Emam, 2012) and perhaps Bayesian deep learning and other approaches that model joint distributions for variables in a dataset and that can use them to generate anonymous data exchangeable with the original data.

As previously mentioned, the Agency could develop tiers of public access that may include a base level of information (e.g., a robust summary or final study report which is devoid of

[22] One SAB member notes that the Proposed Rule emphasis on consideration of studies that use a diversity of models is arbitrary. The member points out that there is no scientific basis provided in the Proposed Rule for giving greater weight to studies that use a wider variety of models without regard for goodness-of-fit, confidence bounds, biological plausibility, attention to untested assumptions. The member notes that the lack of a scientific basis for this approach and the lack of transparency with respect to the logic underlying the proposed approach undercuts confidence that goals of the proposal can be accomplished without perverse effects. The member also notes that this indicates the need for a peer-reviewed transparent analysis in which all assumptions, data and conclusions are made available to the public before the Proposed Rule is enacted.

confidential protected information) for the general user or member of the public and then require additional information or when access to confidential or copyrighted information is warranted. Copyright would protect the particular program or model from being used by others without permission. Signing a non-disclosure agreement would protect other CBI information. A model for accessing personal information contained in federal datasets already exists in the Federal Statistical Research Data Centers (FSRDCs), where individual researchers can access individual-level data under conditions that guarantee the confidentiality of personal information. EPA could consider collaborating with other federal agencies to provide access through the FSRDCs for federal data that are used in epidemiological research (e.g., Medicare data).

The SAB notes that many publications supported by federal grants are freely available in the public literature and this may reduce the concern about publication of these data. For data that are published, but not freely available, the Agency could consider paying a fee to make these data publicly available (e.g., an open access fee). This might protect the rights of journals to copyrighted material.

For CBI and PII, it is possible that access may be possible, but limited in some cases. For example, members of the public could petition the EPA for access to some sensitive data and the Agency could take countermeasures to provide only the information permissible and control the settings around which these data are made available (e.g., onsite access only). This option would likely require the EPA to maintain a “data office” which would require substantial resources to establish and support.

REFERENCES

ARIC Investigators. 1989. The Atherosclerosis Risk in Community (ARIC) Study: Design and Objectives. *American Journal of Epidemiology* 129 (4):687–702. [Available at: https://doi.org/10.1093/oxfordjournals.aje.a115184]

Campbell, D. T., and J. Stanley. 1963. *Experimental and Quasi-experimental Designs for Research*. Chicago, IL: Rand McNally.

Cohen, S.M., A. Chowdhury, and L.L. Arnold. 2016. Inorganic arsenic: A non-genotoxic carcinogen. *Journal of Environmental Sciences*,49:28-37.

Commission on Evidence-Based Policymaking. 2017. *The Promise of Evidence-Based Policymaking*. [Available at: https://www.cep.gov/content/dam/cep/report/cep-final-report.pdf (Accessed, August 18, 2019)]

Dankar, F.D., and K. El Emam. 2012. The application of differential privacy to health data. In: *Proceedings of the 2012 Joint EDBT/ICDT Workshops*, pages 158–166. ACM [Available at: http://www.mathcs.emory.edu/pais12/pdf/PAIS12_Paper2.pdf]

HEI (Health Effects Institute). 2000. *Reanalysis of the Harvard Six Cities Study and the American Cancer Society Study of Particulate Air Pollution and Mortality*. Special Report, Health Effects Institute, Boston, MA.

Johnson, N.M., P.A. Egner, V.K. Baxter, M.B. Sporn, R.S. Wible, T.R. Sutter, J.D. Groopman, T.W. Kensler, and B.D. Roebuck. 2014. Complete protection against aflatoxin B1-induced liver cancer with a triterpenoid: DNA adduct dosimetry, molecular signature, and genotoxicity threshold. *Cancer Prevention Research*, 7(7):658-665.

Maurissen, J. 2010. Practical considerations on the design, execution and analysis of developmental neurotoxicity studies to be published in Neurotoxicology and Teratology. *Neurotoxicology and Teratology* 32:121–123.

National Center for Health Data Statistics. 2019. *NHANES III (1988-1994)*. [Available at: https://wwwn.cdc.gov/nchs/nhanes/nhanes3/Default.aspx (Accessed August 23, 2019)]

National Heart, Lung, and Blood Institute. 2020. *Women's Health Initiative (WHI)*. [Available at: https://www.nhlbi.nih.gov/science/womens-health-initiative-whi]

NRC (National Research Council). 2003. *Sharing Publication-Related Data and Materials: Responsibilities of Authorship in the Life Sciences*. National Academies Press, Washington DC [Available at: https://www.nap.edu/catalog/10613/sharing-publication-related-data-and-materials-responsibilities-of-authorship-in]

NRC (National Research Council). 2005. *Expanding Access to Research Data: Reconciling Risks and Opportunities*. The National Academies Press. [Accessed 3/27/2020]

NRC (National Research Council). 2009. *Science and Decisions: Advancing Risk Assessment (Chapter 5, pages 177-182)*. Washington, DC: The National Academies Press. [Available at: https://doi.org/10.17226/12209]

Prentice R,, J. Rossouw, C. Furberg, S. Johnson, M. Henderson, S. Cummings, J. Manson, L. Freedman, A. Oberman, L. Kuller, and G. Anderson. 1998. Design of the WHI Clinical Trial and Observational Study. *Controlled Clinical Trials* 19:61-109.

Reiter, J.P. and T.E. Raghunathan. 2007. The multiple adaptations of multiple imputation. *Journal of the American Statistical Association* 102:1462-1471 [Available at: https://pdfs.semanticscholar.org/6383/01ae5456419b1e49effec158bf48f32fdf3b.pdf]

Rocher, L, J.M. Hendrickx, and Y.-A. de Montjoye. 2019. Estimating the success of re-identifications in incomplete datasets using generative models. *Nature Communications* 10:3069 [Available at: https://doi.org/10.1038/s41467-019-10933-3]

Rothstein, M. 2010. Is deidentification sufficient to protect health privacy in research? 10.9 *The American Journal of Bioethics* 3-11, 6.

Sweeney, L.M., C.R. Kirman, R.J. Albertini, Y.M. Tan, H.L. Clewell III, J.G. Filser, G. Csanády, L.H. Pottenger, M.I. Banton, C.J. Graham, and L.S. Andrews. 2009. Derivation of inhalation toxicity reference values for propylene oxide using mode of action analysis: example of a threshold carcinogen. *Critical Reviews in Toxicology* 39(6):462-486.

U.S. EPA. 2005. *Guidelines for Carcinogen Risk Assessment*. EPA/630/P-03/001F. Risk Assessment Forum, U.S. Environmental Protection Agency Washington, D.C.

U.S. EPA 2012. *Guidance for Considering and Using Open Literature Toxicity Studies to Support Human Health Risk Assessment*. U.S. EPA Office of Pesticide Programs, Washington, D.C. [Available at: https://www.epa.gov/sites/production/files/2015-07/documents/lit-studies.pdf]

U.S. EPA. 2019. *Uncertainty and Variability*. [Available at: https://www.epa.gov/expobox/uncertainty-and-variability (Accessed August 18, 2019)]

U.S. Office of Management and Budget. 2013. Office of Management and Budget Memorandum M-13-13, *Memorandum for the Heads of Executive Departments and Agencies on Open Data Policy-Managing Information as an Asset 4-5.*

APPENDIX A: DISSENTING OPINIONS

1. Dissenting Opinion from Dr. John Graham

Dr. Graham's individual comments on EPA's proposed rule, Strengthening Transparency in Regulatory Science are provided in:

U.S. EPA Science Advisory Board. 2019. *Consultation on Mechanisms for Secure Access to Personally Identifying Information (PII) and Confidential Business Information (CBI) Under the Proposed Rule, Strengthening Transparency in Regulatory Science.* EPA-SAB-19-005. pages B-16 to B-20. [Available at: https://yosemite.epa.gov/sab/sabproduct.nsf/41042C652229CA398525848500595458/$File/EPA-SAB-19-005.pdf]

2. Dissenting Opinion from Dr. Donald van der Vaart

Addendum to the SAB's consideration of EPA's proposed rule titled *Strengthening Transparency in Regulatory Science*

By Donald R. van der Vaart

While the points raised by the SAB are valuable, the criticism of the transparency rule is overly one-sided. What must be weighed against the inherent challenges for such a rule is the serious problem that currently exists with EPA science and that has eluded the Data Quality Act. Instances of fraudulent science in academic institutions are a reality. As a relatively small illustration, Duke University Medical Center recently settled a whistle-blower suit for more than $100 million for allegations research was fabricated, some of which was done for EPA standard setting. Speaking to the press the whistleblower's lawyer noted that his client's suit was motivated, in part, by,

"…his concerns that the university wasn't being transparent enough, after "Duke's administration and researchers faced the reality that seven years of data were false or unreliable." https://www.npr.org/2019/03/25/706604033/duke-whistleblower-gets-more-than-33-million-in-research-fraud-settlement

Transparency is not something academic institutions crave, but when they accept hundreds of millions of taxpayer dollars those taxpayers deserve some level of meaningful transparency. EPA's proposed rule seeks to do that. Studies undertaken to protect the health and welfare of Americans are crucially important. While conventional wisdom appears to assume added scrutiny would necessarily weaken a standard, I am equally concerned standards might not be protective enough as a result of spurious or fraudulent science.

The proposal should be applauded in its efficiency. When an agency such as EPA relies more on some studies than others, it is natural to seek added scrutiny to those "pivotal" studies. As an example, when only a few studies formed the basis of costly standards for PM 2.5, such as the Harvard Six-City and CPS-II studies, the importance of allowing peer scientific groups to review

and potentially replicate the results is greater, and every effort should be made to provide for such review.

While the SAB's comments should be considered and answered in either the final rule or in subsequent guidance, the importance of the rule should not be lost. The EPA proposal is a thoughtful approach to the admittedly difficult challenge.