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 agency 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 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. The SAB’s advice and comments are provided in the enclosed report.

The SAB recognizes that enhanced public access to scientific data and analytic methods helps ensure scientific integrity and facilitates robust analysis. Strengthening transparency in regulatory science is a worthy goal. However, the SAB finds that key considerations that should inform the Proposed Rule have been omitted from 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 Proposed Rule. The SAB’s major comments and recommendations are as follows:

- The Proposed Rule requires the EPA to clearly identify and make available to the public all studies (or other regulatory science) relied upon when it takes any significant final agency action. This requirement could be cumbersome and impractical if some studies were used in a weight of evidence consideration but not used to determine specific regulatory endpoints. The lack of criteria for what
might satisfy the requirement makes it difficult to understand the implications. The proposed rule should describe in greater detail and clarity how the requirement can be met.

- The Proposed Rule indicates that there may be exceptions to the requirement to make information available to the public. Case-by-case exceptions may exacerbate concerns about inappropriate exclusion of scientifically important studies. Although it will be difficult to develop criteria for exceptions, the EPA would benefit from a framework and guidance that outline criteria to specify exceptions.

- The Proposed Rule states that when promulgating significant regulatory actions, the agency shall ensure that dose response data and models underlying pivotal regulatory science are publicly available in a manner sufficient for independent validation. Greater clarity is needed in definitions of “dose response data and models” and “pivotal regulatory science.” The definitions provided in the Proposed Rule are not adequate and may be better supported in the context of a guidance document.

- 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). Techniques and practices to protect sensitive data have been well-developed by researchers involved in studies with human subjects. The Proposed Rule should build on those efforts.

- A number of questions must be answered in order to assess the feasibility of making dose response data and models available in a manner sufficient for independent validation (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 manage international studies, and how to manage conclusions drawn from meta-analysis). Experimental considerations (such as the appropriateness of controls, protocols employed, limits of quantitation, and other considerations) must be made known to determine whether data are valid. It would be useful for the 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 dose response data and models available in a manner sufficient for validation.

- There will be costs associated with assessing and disseminating data as required in the Proposed Rule. 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.

- The requirement in the Proposed Rule that “raw data” be made available for public inspection 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 raw data for the purposes of the Proposed Rule. The SAB notes that historical data sets 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.
• Personal data used in epidemiological studies generally fall into two categories: data held by federal agencies such as the Centers for Disease Control and Prevention or the Department of Health and Human Services (Medicare data), and those created by universities or private societies. Where this is not already the case, SAB recommends that EPA negotiate with other agencies for individual-level data to be made 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 them from public dissemination. However, no comparable system currently exists for datasets that are owned by universities or societies, that are themselves protected by strong privacy and confidentiality requirements. If the Proposed Rule is implemented without addressing this issue, such datasets risk being excluded entirely from the regulatory process.

• If the EPA wants to see reanalyses of datasets that are critically important for regulation, the agency should consider funding a competition to conduct such reanalysis. A model for this was established by the Health Effects Institute (HEI) in its 2000 reanalyses of datasets from the Six Cities Study and the American Cancer Society. However, HEI has not repeated that exercise, and to SAB’s knowledge has no plans to do so.

• The Proposed Rule requires the EPA to describe and document any assumptions and methods that pertain to the use of dose response data and models underlying pivotal regulatory science and to describe variability and uncertainty. One would anticipate variability in the reporting of this information across laboratories; therefore, the SAB suggests that the EPA offer additional guidance on how to document assumptions, methods, variability, 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,

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 and Interim Chief, Environmental and Occupational Health Division, Department of Public Health Sciences, School of Medicine, University of California, Davis, CA
Dr. Frederick Bernthal, President Emeritus and Senior Advisor to the Board of Trustees, Universities Research Association, Washington, DC
Dr. Bob Blanz, Chief Technical Officer, Arkansas Department of Environmental Quality, North Little Rock, AR
Dr. Todd Brewer, Senior Manager, Partnership 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, 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 and John and Frances Larkin 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. 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. Anne Smith, Managing Director, NERA Economic Consulting, Washington, DC
Table of Contents

1. EXECUTIVE SUMMARY ................................................................................................................. 1

2. INTRODUCTION................................................................................................................................ 5

3. SAB ADVICE AND COMMENT ON THE PROPOSED RULE .................................................... 6

   3.1. General Comments .......................................................................................................................... 6

   3.2. Requirement to Identify All Studies and Regulatory Science Supporting Final Agency Actions...................... 6

   3.3. Requirement to Ensure that Dose Response Data and Models Underlying Pivotal Regulatory Science are Publicly Available in a Manner Sufficient for Independent Validation................................................. 10

   3.4. Requirement to Make the Raw Data Underlying any Proposed Rule Available for Public Inspection ............ 15

   3.5. Requirement to Describe and Document any Assumptions and Methods Pertaining to the use of Dose Response Data and Models Underlying Pivotal Regulatory Science ......................................................... 19

   3.6. Protecting Sensitive Data and Copyrighted or Confidential Business Information ................................................. 21

   3.7. Other SAB Comments on the Proposed Rule ................................................................................ 22

REFERENCES ........................................................................................................................................ 25
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 agency 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. This report provides the SAB findings and recommendations.

The SAB recognizes that enhanced public access to scientific data and analytic methods helps ensure scientific integrity and facilitates robust analysis. Strengthening transparency in regulatory science is a worthy goal. However, the SAB finds that key considerations that should inform the Proposed Rule have been omitted from 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 Proposed Rule. Given the relatively skeletal nature of the proposed rule, it is not possible at this time to define the implications of the rule with confidence. To ensure that the rule is evidence-based requires that EPA 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 Proposed Rule requires the EPA to clearly identify and make available to the public all studies (or other regulatory science) relied upon when it takes any final agency action. This requirement could be cumbersome and impractical if some studies were used in a weight of evidence consideration but not used to determine specific regulatory endpoints. The EPA should clarify how the requirement can be met. If the intent is to identify and make available pivotal studies that were relied upon, scientific and technical challenges of implementing the requirement consist of: (1) having EPA be explicit about which studies are pivotal to recommended regulatory action and (2) making the data and models for the underlying pivotal studies publicly available. Given the lack of clarity, the proposed rule could be viewed as a license to politicize the scientific evaluation required under the statute based on administratively determined criteria for what is practicable. The EPA must also make certain that personally identifying information (PII) and confidential business information (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 some specific policies related to the protection of CBI, exceptions that would be appropriate
where CBI cannot be released, and whether data compensation should be considered. To protect
PII, the identification of individual data in epidemiological studies that arise from small data sets
or targeted geographic areas might not be possible. The lack of criteria for what might satisfy the
requirement makes it difficult to understand the implications. The net effect could be minimal or
massive. The SAB recommends that the EPA develop definitions of terms and methods for
meeting the requirement.

The Proposed Rule indicates that there may be exceptions to the requirement to make
information available to the public. The SAB notes that it will be difficult to develop criteria for
exceptions. Case-by-case exceptions may exacerbate concerns about inappropriate exclusion of
scientifically important studies. However, the SAB finds that the EPA would benefit from a
framework and guidance that outline criteria to specify exceptions. There are some reasonably
anticipated scenarios that the EPA could describe, perhaps with case studies from previous risk
assessments.

The Proposed Rule states that when promulgating significant regulatory actions, the agency shall
ensure that dose response 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 “dose response data and models” and “pivotal regulatory science.” 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 dose response
data and models of interest and the requirements for reporting this information. A technical issue
to be considered is how to separate data sets and models that were the basis of calculations used
to drive the quantitative assessment from ancillary data and models that were part of the weight
of evidence.

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 issues can be addressed through anonymization or de-identification, even de-
identified data sets present risks of re-identification (Rocher et al. 2019). 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. 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 data set. The EPA could
employ similar approaches. Any new rule needs to build on those efforts.

A number of questions must be answered in order to assess the feasibility of making dose
response data and models available in a manner sufficient for independent validation (such as
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 manage international
studies, and how to manage conclusions drawn from meta-analysis). Experimental considerations
(such as the appropriateness of controls, protocols employed, limits of quantitation, and other considerations) must be made known to determine whether data are valid. The SAB finds that assessing the validity of epidemiological studies for the purposes of the Proposed Rule poses particular scientific and technical challenges. 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. It would be useful for the EPA to consider these questions and define “independent validation” because this definition drives the feasibility of whether the EPA can make dose response data and models available for validation. It may be useful for the EPA to develop a guidance document to clarify some scenarios and how the requirement would be managed.

There will be costs associated with assessing and disseminating data as required by the Proposed Rule. 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. The EPA could consider requesting proposals to undertake such work. Obtaining data in a useable format with adequate documentation may be difficult. It is likely that not all data will be in a format suitable for public data sharing. Historical data sets might not be available at the level of detail needed for recalculation. Some of the raw 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. 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 dose response 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. However, it is most likely that some flexibility is going to be required. Standards on transparency are evolving, and modern expectations do not apply to studies from 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.

The SAB finds that the requirement in the proposed rule that “raw data” be made available for public inspection is vague and, as a result, can be interpreted in different ways. If “raw 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 without any demonstrable benefit. There are legitimate legal, ethical, professional, and financial reasons why researchers may be unable or unwilling to fully share “raw data” - including statutes protecting patient privacy, experimental protocols assuring confidentiality of data for human subjects, and (for past studies) issues related to degradation and custody of data. Historical data sets 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.

It is difficult to develop a singular definition of “raw data” that would meet EPA’s objectives in proposing this rule. The definitions of raw data would likely differ based on the available data set. However, the SAB recommends the development of definitions to clarify the requirement to
make raw data available. “Raw 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 raw data for the purposes of the Proposed Rule. 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 development of any transparency rule that goes beyond established norms and procedures.

The Proposed Rule requires the EPA to describe and document any assumptions and methods that pertain to the use of dose response 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 data set, so these are not unusual or inappropriate factors that need to be addressed. However certain scientific and technical challenges must be surmounted. One would anticipate variability in the reporting across laboratories; therefore, EPA could offer guidance on how to document assumptions, methods, variability, 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 agency 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. This report provides the SAB findings and recommendations.
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 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 has questions and concerns about the feasibility of implementing some requirements of the Science and Transparency Rule. In general, the SAB finds that the EPA has not fully identified the problem to be addressed by the Proposed Rule. The EPA must comply with federal transparency and data integrity laws and, as discussed in this report, some additional requirements of the Proposed Rule may not add transparency, and even may make some kinds of research more difficult.

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 action. The Proposed Rule states that the EPA should make all such studies available to the public to the extent possible. The Proposed Rule also states that the data and models underlying scientific studies that are pivotal to the regulatory action must be made publicly available in a manner sufficient for independent validation. While the objective of making information available to the public is worthy, the SAB finds that there are scientific and technical challenges that must be addressed, and feasibility questions that must be considered, when implementing this requirement. The SAB notes that determining the number, scope and range of studies considered in a regulatory action must involve some judgment on the part of the agency (for rule determination) or the authors (for a peer reviewed paper that may be used by regulators). A rule to determine which studies are appropriate would not increase transparency by making such determinations.

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

The SAB finds that requiring identification of all studies and regulatory science supporting regulatory actions could be needlessly cumbersome. Identifying and making “all studies” available could be cumbersome if some studies were used in the weight of evidence consideration but were not used to determine the point of departure (POD) or reference dose (RfD) or other regulatory number. It is unclear how much information and which studies would be included in the requirement to identify and make available studies that are “relied upon.” This
needs to be clarified in the Proposed Rule. If the intent is truly to identify and make available “all studies” that contributed to the ultimate decision, then the amount of data to be made available might be too large to be practical for an independent validation. EPA must also make certain that personally identifying 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 CBI, exceptions that would be appropriate where CBI cannot be released, and whether data compensation should be considered. Where PII and CBI data or methods are concerned, “the public” receiving the information needs to be a small group of people who have provided assurance that they will keep such information confidential and protected. EPA could consider developing 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 confidential protected information) to the general user or member of the public and then require additional information or data sharing agreements 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 epidemiology, microaggregation of data can be used to protect personal identity.

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 data may not be impactful to the final decision (e.g., because of insufficient sample size, doses too high, etc.). The EPA could consider producing a list of study data considered in an evaluation, then strive to provide raw data for critical studies driving regulatory limits (e.g., “pivotal studies”). 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. However, this approach may not be agreeable to all parties. The level of detail required to allow the public to transparently reach the same conclusion as the agency will differ among individuals who seek the data (e.g., 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.

If the intent is to identify and make available pivotal studies that were relied upon, scientific and technical challenges of implementing the requirement consist of: (1) having EPA be explicit about which studies are pivotal to recommended regulatory action (e.g., using a decision and risk analysis framework that explicitly derives recommended actions from study results, and that 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.” It may also be very challenging if holistic judgments and weight-of-evidence frameworks are used instead. 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 data set and document the models used to analyze it and to reach their conclusions. Typically, an analysis data set 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 data sets are already made
publicly available in studies such as NHANES III (National Center for Health Data Statistics
2019). Publishing such analysis data sets 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.” One SAB member suggests that, to accomplish this, the EPA should
not fund new research unless researchers file a research protocol, provide analysis data sets 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
shareable data sets, 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 raw data be deposited with EPA as a condition of funding.

The identification of individual data in epidemiological studies that arise from small data sets or
targeted geographic areas where identification of individuals might not be possible, 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 should be respected
and such information should not be supplied to additional people for validation. It would
probably be impractical, if not impossible, to go back to the participants and request their
approval to provide additional people access to personal information.

Some of the major cohort studies, such as the Women’s Health Initiative (WHI) or the
Atherosclerosis Risk in Communities (ARIC) Study, require intending researchers to write a
“manuscript proposal” – in effect, a document describing the study they intend to conduct,
including what data are required and the rationale behind the study. 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 a similar system of broader applicability, that would provide a
mechanism for interested researchers to access datasets for reanalysis under appropriate controls,
where relevant for EPA regulations.

The lack of any criteria for what might satisfy the requirement to “make all such studies
available to the public to the extent practicable” makes it difficult to understand the implications
of this 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 does all data from every measurement taken as part of the study need to be
available to anyone to manipulate? At one end of this range of interpretation the requirement is
easily implementable, on the other end of the spectrum it would be enormously expensive and
time consuming at best and can be expected to result in the exclusion of much of the scientific
literature from consideration (the machine data are no longer available and/or the researchers are
no longer alive or in a position to assemble the data). The net effect could be minimal or
massive. This is hard to evaluate because either outcome is plausible. Given the lack of clarity,
the proposed rule could be viewed as a license to politicize the scientific evaluation required
under the statute based on administratively determined criteria for what is practicable.
The SAB recommends that the EPA develop definitions of terms and methods for making information available. In the Proposed Rule it is stated that “for the science pivotal to its significant regulatory actions, EPA will ensure that the data and models underlying the science is publicly available in a manner sufficient for validation and analysis.” As noted above, it is unclear exactly what approaches the agency will utilize to ensure that information is available to the public and to what extent this information should be accessible. EPA should clearly define what is meant by “publicly available,” identify the method(s) that it will utilize to make this information available, and discuss the level of effort required by the generator of the data to assist EPA in complying with this proposed mandate. Additionally, it is also unclear what it meant by “validation and analysis” so the Agency should clarify these terms.

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

The Proposed Rule indicates that there may be exceptions to the requirement to make information available to the public. The SAB understands that exceptions may be needed but notes that it would be difficult to develop criteria for exceptions because it seems almost inevitable that unanticipated situations could come up that would not be covered by the criteria. Alternatively, it is not clear that exceptions are appropriate if the goal is truly “to increase transparency in the preparation, identification, and use of science in policymaking.” 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.

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 studies that rely on confidential data. A case-by-case waiver may exacerbate concerns about inappropriate exclusion of scientifically important studies. Reference to a vague “feasibility” standard means that such waivers are decided upon solely by the Administrator, in the absence of clear guidance. This approach could easily have the effect of politicizing science because key decisions about what science is relevant to decision making could be made with no requirement to base decisions on scientific evaluation. It might be useful for the EPA to consider recommendations from an advisory committee when making waiver decisions.

The SAB finds that exclusion of segments of the scientific literature with the possibility of inclusion of selected elements based on non-scientific considerations represents a significant shift in science-based decision making. Such a change could easily undercut the integrity of environmental laws, as it will 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.
Although they 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, there are some reasonably anticipated scenarios that the EPA could describe, perhaps with case studies from previous risk assessments (e.g., What information would fall in scope based on past risk assessments? What raw data would have to be released to support a given risk assessment? Is it feasible to release this information? Why or why not?). EPA might also consider being more explicit regarding criteria for making studies publicly available. Instead of developing specific exceptions to this requirement, EPA might consider being more explicit regarding the criteria for inclusion and evaluate specific exceptions on a case by case basis.

3.3. Requirement to Ensure that Dose Response 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 actions, the agency shall ensure that dose response 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 “dose response data and models” and “pivotal regulatory science.” The definitions are not adequate in the current Federal Register notice and may be better supported in the context of a guidance document that includes realistic examples of the types of dose-response and models of interest and requirements for reporting this information. It would be particularly useful to clarify specific requirements for reporting information from 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. A technical issue to be considered when defining dose response models and data is how to separate data sets and models that were the actual basis of the calculations used to drive the quantitative assessment (e.g., POD or RfD), from additional data sets and models may have been ancillary but part of the weight of evidence used in the regulatory action, if appropriate to the case under consideration. Reasons for the choice of the primary data sets and models for the quantitative assessment and reasons for placing the other options in the weight of evidence (WOE) category should be made available. Dose-response data might be defined as the analysis data set (listing values of causally relevant exposure, response, and covariate variables and uses multiple imputation and related methods for missing data and confidentiality); and 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).”

However, when defining these terms, some questions should be addressed: Dose-response data are referenced. 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? 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

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. 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 Institute¹, National Academies of 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”). Without a specific plan for addressing
these substantive concerns it is difficult to understand the implications of the proposed rule with
respect to leading to exclusion of important data, delaying development of protective regulations
or risking peoples’ privacy. As previously noted, microaggregation of data can protect personal
identity in epidemiology studies.

Even where feasible, de-identification does not address all legitimate restrictions or constraints
on disclosure of data (such as proprietary information). The SAB finds that EPA must address
the critical issues of protecting privacy and confidentiality prior to taking action on the proposed
rule because the evaluation of potential solutions to the issues raised should involve a transparent
and open process. One SAB member notes that a contradiction in the proposed rule is that, in the
name of transparency, it puts forward opaque requirements that may allow decisions to be made
without transparency. The member notes that the proposal raises critical issues that potentially
challenge the integrity of the science used in making science-based decisions as required by law.

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

The SAB notes that a number of scientific and technical challenges must be addressed and
questions answered in order to make dose response data and models available in a manner
sufficient for independent validation. There are experimental considerations that must be made

¹Letter 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 - which
have in general enhanced the quality and the sensitivity of the studies - increase the difficulty of providing a fully
"de-identified" data set while also enabling a different investigator to conduct a full replication and sensitivity
analysis of the original study results.”).
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 quantitation, 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.

It would be useful for the EPA to define “independent validation.” This drives the feasibility of whether EPA can make dose response data and models available for independent validation. For example, how much information is sufficient for independent validation? Will this consist of equations where reviewers can verify the math or 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 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 finds that the following specific questions should be answered 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 on 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 move forward if laboratories refuse to collate and release their data for public review? Will these data be excluded from use in risk assessment? If so, there is potential to bias risk assessments based on the exclusion of data that investigators chose not to make publicly available. In contrast, perhaps these data, without public scrutiny, should not be used for 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 data sets 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 dose response data and models may rely on proprietary software that may not be readily accessible or available 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?

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 dose response 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). The lack of details on implementation of this rule allows for an endless number of hypothetical scenarios, so a guidance document may offer clarity at least for the most common scenarios.

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 to know: 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 data sets 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 data set; how participants were selected and what the selection/exclusion criteria were.
Challenges and costs of processing and documenting of data prior to public release, maintenance and administration of data sets so they are publicly available, and handling historical data sets

There will be technical, scientific and resource challenges associated with assessing and disseminating data as required by the proposed rule. Costs are 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. It is likely that not all data will be in a format suitable for public data sharing (e.g., non-traditional data formats, PII, CBI, 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 data sets might not be available at the level of detail needed for recalculation. Some of the raw data or computational methods may have been discarded if they were deemed not necessary to maintain. Certainly 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; PII should not be revealed. 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 data sets and provide them to independent public people or groups for validation processes would need to be identified and they must assure that the data will 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 access and approval or disapproval of the request.

The SAB suggests that EPA consider establishing an office (or virtual office) on data sharing. This group could identify 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/researchers to provide the necessary information in a “user friendly” format. This office also could build and manage data archives and pursue critical historical data sets if deemed important. There will be costs associated with the establishment of such an office as well as researchers’ time to collate raw 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. The EPA may find creative ways to offset the expense associated with data submission for pivotal studies.

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

The SAB notes that processing and documenting dose response 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. However, it is most likely that some flexibility is going to be required. 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. In those situations, EPA should seek, support, and encourage future studies to replace historical studies lacking transparency, but this cannot be done instantly. It will be a process over several years.

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

The SAB notes that the retrospective application of modern transparency standards is a challenge. A large amount of work would be required to locate, curate and retrospectively make data sets 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.

3.4. Requirement to Make the Raw Data Underlying any Proposed Rule Available for Public Inspection

Under the Proposed Rule, EPA would require that raw data underlying any proposed rule be made available for public inspection. The SAB finds that the requirement in the Proposed Rule that “raw data” be made available for public inspection is vague and, as a result, can be interpreted in different ways. If “raw 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 without any demonstrable benefit. Even if the “raw 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 “raw 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 impact the conclusions drawn arbitrarily.

The Proposed Rule should be explicitly prospective and follow evolving norms developed by the scientific community as well as federal agencies (National Science Foundation, National Institutes of Health, National Oceanic and Atmospheric Administration). The lack of reference to and comparison with existing and evolving federal procedures to address the underlying purpose of the Proposed Regulation (increased transparency) within scientific studies supported and utilized by the federal government undermines the confidence that the proposed regulation will meet its objectives effectively and efficiently.

The Proposed Rule is based on the premise that using scientific information that can be independently validated will lead to better outcomes and strengthen public confidence in the health and environmental protections underpinning EPA’s regulatory actions. The SAB notes that public disclosure of data is helpful in enabling reanalysis that can assure the quality of the work, but it is not a sine qua non for validating scientific studies. The scientific community uses numerous tools to validate studies without access to all underlying “raw data,” including rigorous peer review, replication of a study using the same methodology but with different data sources, or reproduction of a study’s conclusions using different methodologies and data. The EPA’s proposed policy of excluding from consideration any study for which underlying data are not made publicly available is not consistent with sound scientific practice.

The SAB notes that there are legitimate legal, ethical, professional and financial reasons why researchers may be unable or unwilling to fully share “raw data” - including statutes protecting patient 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 validated in other ways - much less require that the studies be ignored in regulatory decisions.\(^2\),\(^3\),\(^4\),\(^5\)

\(^2\) 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

\(^3\) 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

\(^4\) U.S. Office of Management and Budget. 2002. IQA Guidelines: "[m]aking 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."

\(^5\) 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
There appears to be consistency among analyses of how to address transparency that are orthogonal to the proposed rule. There is no justification in the Proposed Rule for why EPA finds that 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.

**Feasibility of making raw data available for public inspection**

In the absence of more detailed information, it is difficult to determine the feasibility of making raw data available for public inspection as required in the Proposed Rule. There are numerous reasons why it may be difficult to share raw data including PII (e.g., HIPAA requirements), CBI, non-standard data collection systems (e.g., not easily accessed outside). Time and cost associated with formatting, curating and sharing raw data also may pose difficulties for laboratories. The SAB recommends the development of definitions to clarify the requirement of making raw data available. Although EPA wants to make data available to allow others to reproduce results, privacy considerations (e.g., like those mandated in HIPAA) must be respected. “Raw data” should not be confused with personally-identifiable data.

Given the reasons why sharing raw data may be difficult, the method(s) that EPA identifies and utilizes for making “raw data” available for public inspection will impact significantly the feasibility of this requirement. In order to determine the feasibility, EPA must consider: (1) what level of detail it will plan to make data available (e.g. a robust summary of the available data, all of the in-life and necropsy data from a specific study, only data that are included in the scientific publication or publicly available report or publication); (2) the type of platform EPA will use to make this information available (e.g., free publicly available online database with searchable features, electronic cd rom or hard copies available by public request); and (3) requirements the public will have to adhere to/agree to in order to access the data to ensure protection of relevant information (e.g., formal data transfer agreements between the agency and entity requesting access to the data, general electronic agreement and check box noting that the data user agrees to meet certain pre-established requirements).

The SAB notes a number of additional concerns about the feasibility of making raw data available for public inspection. Raw data would likely include all data points collected, some of which might have been legitimately removed from the final analysis, e.g., there is knowledge of a mistake in the collection of a particular data point, or it is determined statistically to be an outlier that could be eliminated. Without documentation of such reasons in an extensive data set, 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.”

---
it would be difficult for the person/group verifying the data calculations to come to the same
conclusions as the original calculations. Additionally, if the data set were reliable but not
generated under GLP requirements (such as from an academic lab), the documentation in the raw
data might not be extensive enough regarding outliers or mistakes that were taken into
consideration when the final data set was used for modeling. If this were the case, then
verification would not produce the same results. Some raw data, again reliable data that comes
from a non-GLP laboratory, may have been discarded and be unavailable. Some raw data might
be print-outs and scans from a laboratory instrument that would require technical knowledge to
interpret, and might require handwritten notes in original laboratory notebooks that would be
difficult to obtain or interpret, especially if studies were performed by the laboratory without the
intention that they be used for regulatory purposes. Therefore this requirement might not be
feasible. Additionally, the labor involved in resorting and recalculating the information from
available raw data might be too time-consuming and cumbersome to be reasonable.

Regardless of how “raw data” is defined, the requirement to make data available for public
inspection will be more easily implemented for some data sets than others (e.g., GLP studies
include individual sample data as part of standard reporting). The SAB notes that the expectation
that raw 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 raw 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 data sets where CBI and/or PII data will require more
onerous steps to protect data confidentiality. Based on EPA responses to the SAB questions, it
appears that the agency is seeking approaches to manage these data issues.

**Definition of raw data**

It is difficult to develop a singular definition of “raw data” that would meet EPA’s objectives in
proposing this rule. Raw data could be defined as the *analysis data set*, as previously discussed.
Due to the depth of data included as “raw,” the definition of raw data would likely differ based
on the available data set. For example, raw data for an in vivo study could include all the
individual animal body weight or individual pathology data while a raw data set, 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.

Wikipedia defines raw data (also known as primary data) as “data collected from a source.” This
would 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. 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, raw data is
defined as “any laboratory worksheets, records, memoranda, notes, or exact copies thereof, 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 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 of this information publicly available may be resource intensive and provide limited utility.

It might be reasonable to consider the initial compilation of 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 down 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 raw 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 and to the best of our knowledge has no plans to do so in the future. EPA could consider using the HEI model for funding its own reanalyses, 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 Dose Response 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 dose response 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 used, the variability of the replications, and any other confounders that add to the uncertainty of the final data set, so these are not unusual or inappropriate factors that need 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 (e.g., that there are no unmeasured confounders, or that dose-response functions are linear below the lowest dose for which data are available) made in deriving conclusions from data. 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 with possible 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 identify the appropriate level of detail to define some parameters like methods. This will vary depending on factors like 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.

The Proposed Rule is viewed by some members of the SAB 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) recommended the use of linear non-threshold models as a default for both cancer and non-cancer responses). 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. 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 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.

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.

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) report 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 data set. 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 data sets 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 data set 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 confidential protected information) to the general user or member of the public and then require additional information or data sharing agreements 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 negotiating with other federal agencies to provide access through the FSRDCs for federal data that are used in epidemiological research, such as Medicare.
The SAB notes that many publications supported by federal grants are freely available in the public literature and this may lessen 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.

3.7. Other SAB Comments on the Proposed Rule

The SAB provides the following additional comments and concerns about the Proposed Rule.

1. As previously discussed, the following language in the preamble of the Proposed Rule has been interpreted by some members of the SAB as an abandonment of the linear no-threshold models. However, other SAB members view it 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 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). It is important that the EPA provide clarity as to the agency’s intent:

“…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.”

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.”

2. In general, it is difficult to define universal “rules of good behavior” for as many-faceted
a question 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. 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. The SAB is concerned that key considerations that should inform the rule appear to be
omitted or presented with no analysis - for example, it is not clear how many of the
studies EPA currently relies upon to take important regulatory actions would meet the
public disclosure standards in the proposal, whether EPA has assessed how many of those
studies would be feasible to provide underlying information, or what the impact of
precluding those studies would be on EPA's decision making and its ability to protect
public health / environment.

4. As previously discussed, the SAB is concerned that key terms and implementation issues
in the Proposed Rule are left undefined or too vague to enable informed comment (e.g.,
“sufficient for independent validation,” “feasible” (for purposes of Administrator's
waiver), “replication,” “validation,” “publicly available”). In particular the lack of a
functional definition of “data” and “raw data” in the Proposed Rule as it relates to peer-
reviewed studies is such an important omission that it seriously undercuts the ability to
understand the implications of the proposed rule. In particular the SAB notes that the
following questions should be answered: Would it suffice that researchers make available
the data on which they performed the bulk of their calculations (the analysis dataset),
which typically follows preprocessing or aggregation, or does EPA expect full raw data
down to the level of individual measurements, including data directly used to carry out
the reported statistical analysis and model development? What level of detail would be
provided (e.g., if subjects are removed from the study, would these subjects be identified
and reasons given)? How does the role of QA/QC methodology affect the choice of
“data” in the stages of aggregation that would be released?

5. The SAB is concerned that the net effect of the proposed rule could be to inappropriately
limit the science used by EPA and rigor of decision-making - impeding EPA's ability to
protect the public. Evidence to support this concern has been presented by the editors of scientific journals and the Congressional Budget Office.\textsuperscript{6,7,8}

6. The SAB is 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 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 rest on confidential data.

7. The Administrative Procedure Act requires that a proposed rule provide enough information for interested parties to “comment meaningfully.” As stated above the SAB finds that additional information about the proposed rule is needed.

\textsuperscript{6} Editors of scientific journals, the American Association for the Advancement of Science, the National Academies for Science, Engineering, and Medicine all raised concerns that proposed rule would not promote scientific integrity and would have harmful consequences.

\textsuperscript{7} Congressional Budget Office. 2015. Congressional Budget Office Cost Estimate, HR 1030, Secret Science Reform Act of 2015. Congressional Budget Office analysis of legislation that closely resembles the proposed rule concluded that it would significantly reduce the number of studies relied upon by EPA, perhaps by as much as 50%.

\textsuperscript{8} EPA officials have explained to Congressional Budget Office that the agency would implement H.R. 1430 with minimal funding. That approach to implementing legislation resembling the proposed rule would significantly reduce the number of studies relied on by EPA.
REFERENCES


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
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.

inhalation toxicity reference values for propylene oxide using mode of action analysis: example

Assessment Forum, U.S. Environmental Protection Agency Washington, D.C.

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]


M-13-13, *Memorandum for the Heads of Executive Departments and Agencies on Open Data
Policy-Managing Information as an Asset 4-5* (May 9, 2013)