



UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
WASHINGTON D.C. 20460

OFFICE OF THE ADMINISTRATOR
SCIENCE ADVISORY BOARD

EPA-SAB-20-xxx

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

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

Dear Administrator Wheeler:

As part of its statutory duties, the EPA Science Advisory Board (SAB) may provide advice and comment on the scientific and technical basis of certain planned EPA actions. The Environmental Research, Development, and Demonstration Authorization Act of 1978 (ERDDAA) requires the 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

1 might satisfy the requirement makes it difficult to understand the implications. The proposed rule
2 should describe in greater detail and clarity how the requirement can be met.
3

- 4 • The Proposed Rule indicates that there may be exceptions to the requirement to make information
5 available to the public. Case-by-case exceptions may exacerbate concerns about inappropriate
6 exclusion of scientifically important studies. Although it will be difficult to develop criteria for
7 exceptions, the EPA would benefit from a framework and guidance that outline criteria to specify
8 exceptions.
9
- 10 • The Proposed Rule states that when promulgating significant regulatory actions, the agency shall
11 ensure that dose response data and models underlying pivotal regulatory science are publicly
12 available in a manner sufficient for independent validation. Greater clarity is needed in definitions of
13 “dose response data and models” and “pivotal regulatory science.” The definitions provided in the
14 Proposed Rule are not adequate and may be better supported in the context of a guidance document.
15
- 16 • Protecting privacy and confidentiality must be taken into consideration when data and models
17 underlying pivotal regulatory science are made available to the public. Requirements for protection
18 of privacy have been established under the Health Insurance Portability and Accountability Act
19 (HIPAA). Techniques and practices to protect sensitive data have been well-developed by
20 researchers involved in studies with human subjects. The Proposed Rule should build on those
21 efforts.
22
- 23 • A number of questions must be answered in order to assess the feasibility of making dose response
24 data and models available in a manner sufficient for independent validation (such as how to treat
25 studies that are formatted in a manner that makes the data difficult to share, how to move forward if
26 laboratories refuse to collate and release data, how to manage international studies, and how to
27 manage conclusions drawn from meta-analysis). Experimental considerations (such as the
28 appropriateness of controls, protocols employed, limits of quantitation, and other considerations)
29 must be made known to determine whether data are valid. It would be useful for the EPA to consider
30 these questions and define independent validation in the Proposed Rule because this definition drives
31 the feasibility of whether the EPA can make dose response data and models available in a manner
32 sufficient for validation.
33
- 34 • There will be costs associated with assessing and disseminating data as required in the Proposed
35 Rule. The agency should consider seeking input from experts in library science, data curation
36 management, and data retention to identify best practices and tools to ensure efficiency and utility of
37 data that are made available.
38
- 39 • The requirement in the Proposed Rule that “raw data” be made available for public inspection is
40 vague and, as a result, can be interpreted in different ways. Extensive work is required, across a
41 diversity of fields, data types and data of different ages, to understand the implications of adopting
42 different definitions of raw data for the purposes of the Proposed Rule. The SAB notes that historical
43 data sets may be unavailable or may have been discarded if deemed not necessary to maintain. A
44 possible way to address this problem is to apply rule requirements only to information developed
45 after the effective date of a final rule.
46

Science Advisory Board (SAB) Draft Report (10/16/19) – Do Not Cite or Quote.

This draft has not been reviewed or approved by the chartered SAB and does not represent EPA policy.

- 1 • Personal data used in epidemiological studies generally fall into two categories: data held by federal
2 agencies such as the Centers for Disease Control and Prevention or the Department of Health and
3 Human Services (Medicare data), and those created by universities or private societies. Where this is
4 not already the case, SAB recommends that EPA negotiate with other agencies for individual-level
5 data to be made available through the system of Federal Statistical Research Data Centers, which are
6 already widely used by the Census Bureau to allow researchers to gain access to individual data
7 while protecting them from public dissemination. However, no comparable system currently exists
8 for datasets that are owned by universities or societies, that are themselves protected by strong
9 privacy and confidentiality requirements. If the Proposed Rule is implemented without addressing
10 this issue, such datasets risk being excluded entirely from the regulatory process.
11
- 12 • If the EPA wants to see reanalyses of datasets that are critically important for regulation, the agency
13 should consider funding a competition to conduct such reanalysis. A model for this was established
14 by the Health Effects Institute (HEI) in its 2000 reanalyses of datasets from the Six Cities Study and
15 the American Cancer Society. However, HEI has not repeated that exercise, and to SAB's
16 knowledge has no plans to do so.
17
- 18 • The Proposed Rule requires the EPA to describe and document any assumptions and methods that
19 pertain to the use of dose response data and models underlying pivotal regulatory science and to
20 describe variability and uncertainty. One would anticipate variability in the reporting of this
21 information across laboratories; therefore, the SAB suggests that the EPA offer additional guidance
22 on how to document assumptions, methods, variability, and uncertainty.
23
24

25 The SAB appreciates the opportunity to provide the EPA with advice and comment on the Proposed
26 Rule. We look forward to receiving the agency's response.
27

28 Sincerely,
29

30
31 Dr. Michael Honeycutt, Chair
32 Science Advisory Board
33
34

35 Enclosure

NOTICE

1
2
3
4
5
6
7
8
9
10
11

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>.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47

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, 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

1
2 **Dr. Otto C. Doering III**, Professor, Department of Agricultural Economics, Purdue University, W.
3 Lafayette, IN

4
5 **Dr. Susan P. Felter**, Research Fellow, Global Product Stewardship, Procter & Gamble, Mason, OH

6
7 **Dr. Joseph A. Gardella**, SUNY Distinguished Professor and John and Frances Larkin Professor of
8 Chemistry, Department of Chemistry, College of Arts and Sciences, University at Buffalo, Buffalo, NY

9
10 **Dr. John D. Graham**, Dean, O’Neill School of Public and Environmental Affairs, Indiana University,
11 Bloomington, IN

12
13 **Dr. John Guckenheimer**, Professor Emeritus and Interim Director, Center for Applied Mathematics,
14 Cornell University, Ithaca, NY

15
16 **Dr. Robert E. Mace**, The Meadows Center for Water and the Environment, Texas State University, San
17 Marcos, TX

18
19 **Dr. Clyde F. Martin**, Horn Professor of Mathematics, Emeritus, Department of Mathematics and
20 Statistics, Texas Tech University, Crofton, MD

21
22 **Dr. Sue Marty**, Senior Toxicology Leader, Toxicology & Environmental Research, The Dow Chemical
23 Company, Midland, MI

24
25 **Mr. Robert W. Merritt**, Independent Consultant, Houston, TX

26
27 **Dr. Larry Monroe**, Independent Consultant, Braselton, GA

28
29 **Dr. Thomas F. Parkerton**, Senior Environmental Scientist, Toxicology & Environmental Science
30 Division, ExxonMobil Biomedical Science, Spring, TX

31
32 **Dr. Robert Phalen**, Professor, Air Pollution Health Effects Laboratory, Department of Medicine,
33 University of California-Irvine, Irvine, CA

34
35 **Dr. Kenneth M. Portier**, Independent Consultant, Athens, GA

36
37 **Dr. Robert Puls**, Owner/Principal, Robert Puls Environmental Consulting, Bluffton, SC

38
39 **Dr. Kenneth Ramos**, Executive Director, Institute of Biosciences and Technology, Texas A&M
40 University, Houston, TX

41
42 **Dr. Tara L. Sabo-Attwood**, Associate Professor and Chair, Department of Environmental and Global
43 Health, College of Public Health and Health Professionals, University of Florida, Gainesville, FL

44
45 **Dr. Anne Smith**, Managing Director, NERA Economic Consulting, Washington, DC

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

3
4 **Dr. Jay Turner**, Professor and Vice Dean for Education, Department of Energy, Environmental and
5 Chemical Engineering, McKelvey School of Engineering, Washington University, St. Louis, MO

6
7 **Dr. Brant Ulsh**, Principal Health Physicist, M.H. Chew & Associates, Cincinnati, OH

8
9 **Dr. Donald van der Vaart**, Senior Fellow, John Locke Foundation, Raleigh, NC

10
11 **Dr. Kimberly White**, Senior Director, Chemical Products and Technology Division, American
12 Chemistry Council, Washington, DC

13
14 **Dr. Mark Wiesner**, Professor, Department of Civil and Environmental Engineering, Director, Center
15 for the Environmental Implications of NanoTechnology (CEINT), Pratt School of Engineering, Nicholas
16 School of the Environment, Duke University, Durham, NC

17
18 **Dr. Peter J. Wilcoxon**, Laura J. and L. Douglas Meredith Professor for Teaching Excellence, Director,
19 Center for Environmental Policy and Administration, The Maxwell School, Syracuse University,
20 Syracuse, NY

21
22 **Dr. Richard A. Williams**, Independent Consultant, McLean, VA

23
24 **Dr. S. Stanley Young**, Chief Executive Officer, CGStat, Raleigh, NC

25
26 **Dr. Matthew Zwiernik**, Professor, Department of Animal Science, Institute for Integrative Toxicology,
27 Michigan State University, East Lansing, MI

28
29
30 **SCIENCE ADVISORY BOARD STAFF**

31
32 **Dr. Thomas Armitage**, Designated Federal Officer, U.S. Environmental Protection Agency,
33 Washington, DC

34
35
36
37

Table of Contents

1
2
3 **1. EXECUTIVE SUMMARY 1**
4 **2. INTRODUCTION..... 5**
5 **3. SAB ADVICE AND COMMENT ON THE PROPOSED RULE 6**
6 **3.1. General Comments6**
7 **3.2. Requirement to Identify All Studies and Regulatory Science Supporting Final Agency Actions.....6**
8 **3.3. Requirement to Ensure that Dose Response Data and Models Underlying Pivotal Regulatory**
9 **Science are Publicly Available in a Manner Sufficient for Independent Validation.....10**
10 **3.4. Requirement to Make the Raw Data Underlying any Proposed Rule Available for Public Inspection15**
11 **3.5. Requirement to Describe and Document any Assumptions and Methods Pertaining to the use**
12 **of Dose Response Data and Models Underlying Pivotal Regulatory Science19**
13 **3.6. Protecting Sensitive Data and Copyrighted or Confidential Business Information21**
14 **3.7. Other SAB Comments on the Proposed Rule.....22**
15 **REFERENCES..... 25**
16
17
18
19

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

1 where CBI cannot be released, and whether data compensation should be considered. To protect
2 PII, the identification of individual data in epidemiological studies that arise from small data sets
3 or targeted geographic areas might not be possible. The lack of criteria for what might satisfy the
4 requirement makes it difficult to understand the implications. The net effect could be minimal or
5 massive. The SAB recommends that the EPA develop definitions of terms and methods for
6 meeting the requirement.

7
8 The Proposed Rule indicates that there may be exceptions to the requirement to make
9 information available to the public. The SAB notes that it will be difficult to develop criteria for
10 exceptions. Case-by-case exceptions may exacerbate concerns about inappropriate exclusion of
11 scientifically important studies. However, the SAB finds that the EPA would benefit from a
12 framework and guidance that outline criteria to specify exceptions, There are some reasonably
13 anticipated scenarios that the EPA could describe, perhaps with case studies from previous risk
14 assessments.

15
16 The Proposed Rule states that when promulgating significant regulatory actions, the agency shall
17 ensure that dose response data and models underlying pivotal regulatory science are publicly
18 available in a manner sufficient for independent validation. The SAB finds that greater clarity is
19 needed in definitions of “dose response data and models” and “pivotal regulatory science.” The
20 definitions provided in the Proposed Rule are not adequate and may be better supported in the
21 context of a guidance document that includes realistic examples of the types of dose response
22 data and models of interest and the requirements for reporting this information. A technical issue
23 to be considered is how to separate data sets and models that were the basis of calculations used
24 to drive the quantitative assessment from ancillary data and models that were part of the weight
25 of evidence.

26
27 The SAB notes that protecting privacy and confidentiality must be taken into consideration when
28 data and models underlying pivotal regulatory science are made available to the public.
29 Requirements for protection of privacy have been established under the Health Insurance
30 Portability and Accountability Act (HIPAA). Although the Proposed Rule suggests that privacy
31 and confidentiality issues can be addressed through anonymization or de-identification, even de-
32 identified data sets present risks of re-identification (Rocher et al. 2019). It seems reasonable that
33 the standards applied by the EPA to protect sensitive data and copyrighted or confidential
34 business information should be the same as the standards applied by editors of reputable
35 scientific journals (e.g., guidance from the International Committee of Medical Journal Editors).
36 Techniques and practices such as microaggregation to protect sensitive data have been developed
37 by researchers involved in studies with human subjects. Other federal agencies have utilized
38 specific and explicit data transfer agreements with entities that sought access to protected
39 information for the purposes of reviewing and running analysis on the data set. The EPA could
40 employ similar approaches. Any new rule needs to build on those efforts.

41
42 A number of questions must be answered in order to assess the feasibility of making dose
43 response data and models available in a manner sufficient for independent validation (such as
44 how to treat studies that are formatted in a manner that make the data difficult to share, how to
45 move forward if laboratories refuse to collate and release data, how to manage international
46 studies, and how to manage conclusions drawn from meta-analysis). Experimental considerations

1 (such as the appropriateness of controls, protocols employed, limits of quantitation, and other
2 considerations) must be made known to determine whether data are valid. The SAB finds that
3 assessing the validity of epidemiological studies for the purposes of the Proposed Rule poses
4 particular scientific and technical challenges. Issues to be addressed include understanding bias,
5 confounding factors, measurement errors and exposure characterization. All of these factors will
6 play a role in defining what would be appropriate for access and validation purposes.
7 It would be useful for the EPA to consider these questions and define “independent validation”
8 because this definition drives the feasibility of whether the EPA can make dose response data
9 and models available for validation. It may be useful for the EPA to develop a guidance
10 document to clarify some scenarios and how the requirement would be managed.

11
12 There will be costs associated with assessing and disseminating data as required by the Proposed
13 Rule. The agency should consider seeking input from experts in library science, data curation
14 management and data retention to identify best practices and tools to ensure efficiency and utility
15 of data that are made available. The EPA could consider requesting proposals to undertake such
16 work. Obtaining data in a useable format with adequate documentation may be difficult. It is
17 likely that not all data will be in a format suitable for public data sharing. Historical data sets
18 might not be available at the level of detail needed for recalculation. Some of the raw data or
19 computational methods may have been discarded if they were deemed not necessary to maintain.
20 The SAB suggests that the EPA consider establishing an office (or virtual office) on data sharing.
21 This group could identify standard data formats (data templates), how to report
22 methods/procedures used, uncertainty, and when and how to implement greater data protections
23 for PII/CBI. The SAB notes that processing and documenting dose response data and models
24 developed prior to the effective date of the rule will pose challenges. A possible way to address
25 this is to apply rule requirements only to information developed after the effective date of a final
26 rule. However, it is most likely that some flexibility is going to be required. Standards on
27 transparency are evolving, and modern expectations do not apply to studies from 10 or 20 years
28 ago. It is reasonable to apply modern standards of transparency and public availability to current
29 and future studies, but it will not always be possible to apply these same standards
30 retrospectively.

31
32 The SAB finds that the requirement in the proposed rule that “raw data” be made available for
33 public inspection is vague and, as a result, can be interpreted in different ways. If “raw data”
34 includes all machine output associated with analysis it would create demands on researchers that
35 would be very onerous and could significantly slow down science-based decision making
36 without any demonstrable benefit. There are legitimate legal, ethical, professional, and financial
37 reasons why researchers may be unable or unwilling to fully share “raw data” - including statutes
38 protecting patient privacy, experimental protocols assuring confidentiality of data for human
39 subjects, and (for past studies) issues related to degradation and custody of data. Historical data
40 sets may be unavailable or may have been discarded if deemed not necessary to maintain. A
41 possible way to address this problem is to apply rule requirements only to information developed
42 after the effective date of a final rule.

43
44 It is difficult to develop a singular definition of “raw data” that would meet EPA’s objectives in
45 proposing this rule. The definitions of raw data would likely differ based on the available data
46 set. However, the SAB recommends the development of definitions to clarify the requirement to

Science Advisory Board (SAB) Draft Report (10/16/19) – Do Not Cite or Quote.

This draft has not been reviewed or approved by the chartered SAB and does not represent EPA policy.

1 make raw data available. “Raw data” should not be confused with personally-identifiable data.
2 Extensive work would be required, across a diversity of fields, data types and data of different
3 ages, to understand the implications of adopting different definitions of raw data for the purposes
4 of the Proposed Rule. Such an effort is beyond the scope of what the SAB can undertake with the
5 resources and time available. However, the SAB finds that such an analysis is foundational to
6 development of any transparency rule that goes beyond established norms and procedures.
7
8 The Proposed Rule requires the EPA to describe and document any assumptions and methods
9 that pertain to the use of dose response data and models underlying pivotal regulatory science
10 and to describe variability and uncertainty. High quality scientific studies identify the
11 assumptions used in models, methods used, the variability of the replications, and any other
12 confounders that add to the uncertainty of the final data set, so these are not unusual or
13 inappropriate factors that need to be addressed. However certain scientific and technical
14 challenges must be surmounted. One would anticipate variability in the reporting across
15 laboratories; therefore, EPA could offer guidance on how to document assumptions, methods,
16 variability, and uncertainty.
17
18
19
20

2. INTRODUCTION

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17

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

1 needs to be clarified in the Proposed Rule. If the intent is truly to identify and make available “all
2 studies” that contributed to the ultimate decision, then the amount of data to be made available
3 might be too large to be practical for an independent validation. EPA must also make certain that
4 personally identifying information (PII) and confidential business information (CBI) are not
5 available to persons/groups not properly vetted, approved, and trusted by those owning the CBI
6 information. Without adequate protection, industry data generated by one company can be used
7 by other companies to fulfill regulatory requirements in other geographies. It would be beneficial
8 for the EPA to develop some specific policies related to the protection of CBI, exceptions that
9 would be appropriate where CBI cannot be released, and whether data compensation should be
10 considered. Where PII and CBI data or methods are concerned, “the public” receiving the
11 information needs to be a small group of people who have provided assurance that they will keep
12 such information confidential and protected. EPA could consider developing tiers of public
13 access that may include a base level of information (e.g., a robust summary or final study report
14 which is devoid of confidential protected information) to the general user or member of the
15 public and then require additional information or data sharing agreements when access to
16 confidential or copyrighted information is warranted. In matters of public health, it is common to
17 require disclosure to a trusted third party. In the case of environmental epidemiology,
18 microaggregation of data can be used to protect personal identity.
19

20 It may not be practical for EPA to make all studies or other regulatory science used in a final
21 regulatory action available to the public. Many studies are included as part of a regulatory
22 evaluation but data may not be impactful to the final decision (e.g., because of insufficient
23 sample size, doses too high, etc.). The EPA could consider producing a list of study data
24 considered in an evaluation, then strive to provide raw data for critical studies driving regulatory
25 limits (e.g., “pivotal studies”). In this way, the public could understand which studies/data were
26 considered and used without expending the resources to gather and disseminate all available
27 data. However, this approach may not be agreeable to all parties. The level of detail required to
28 allow the public to transparently reach the same conclusion as the agency will differ among
29 individuals who seek the data (e.g., some people may wish to verify that the EPA has selected
30 the right “pivotal studies”); however, a more expansive scope would increase the reporting
31 burden in a way that may make the Proposed Rule untenable.
32

33 If the intent is to identify and make available pivotal studies that were relied upon, scientific and
34 technical challenges of implementing the requirement consist of: (1) having EPA be explicit
35 about which studies are pivotal to recommended regulatory action (e.g., using a decision and risk
36 analysis framework that explicitly derives recommended actions from study results, and that thus
37 enables the roles of studies in the regulatory actions to be precisely identified); and (2) making
38 the data and models for the underlying pivotal studies publicly available. The first step may be
39 relatively technically straightforward to implement if standard decision and risk analysis
40 frameworks are used to derive policy recommendations from study results. However, there are a
41 number of emerging technologies (e.g., use of transcriptomics data) that raise new challenges,
42 including dealing with “big data.” It may also be very challenging if holistic judgments and
43 weight-of-evidence frameworks are used instead. Likewise, the second step, making available
44 the data and models underlying pivotal scientific studies, is also technically straightforward
45 when the pivotal studies already provide the analysis data set and document the models used to
46 analyze it and to reach their conclusions. Typically, an analysis data set lists values of exposure,

1 response, and covariate variables and uses multiple imputation and related methods for missing
2 data and confidentiality (Reiter and Raghunathan 2007), Such data sets are already made
3 publicly available in studies such as NHANES III (National Center for Health Data Statistics
4 2019). Publishing such analysis data sets and the models used to analyze them to produce
5 conclusions that are pivotal to regulatory action seems to be an appropriate scope of coverage for
6 the stated goal, i.e., “to increase transparency in the preparation, identification, and use of
7 science in policymaking.” One SAB member suggests that, to accomplish this, the EPA should
8 not fund new research unless researchers file a research protocol, provide analysis data sets to
9 EPA and the public, and provide analysis code. The SAB member also suggests that previous
10 research should not be considered to support regulatory actions unless researchers provide
11 sharable data sets, analyze them, provide analysis code, and file these materials with the EPA.
12 The SAB notes, however, that institutional review boards at host universities may not agree with
13 the requirement that raw data be deposited with EPA as a condition of funding.
14

15 The identification of individual data in epidemiological studies that arise from small data sets or
16 targeted geographic areas where identification of individuals might not be possible, especially if
17 the Informed Consent Form indicated that only the particular researchers who conducted the
18 study would have access to the information and data. If the participants agreed to grant only a
19 select group of researchers access to their personal information, then that should be respected
20 and such information should not be supplied to additional people for validation. It would
21 probably be impractical, if not impossible, to go back to the participants and request their
22 approval to provide additional people access to personal information.
23

24 Some of the major cohort studies, such as the Women’s Health Initiative (WHI) or the
25 Atherosclerosis Risk in Communities (ARIC) Study, require intending researchers to write a
26 “manuscript proposal” – in effect, a document describing the study they intend to conduct,
27 including what data are required and the rationale behind the study. The proposal is reviewed by
28 a committee and, in most cases, is either approved without modification or returned to the
29 investigators to address specific issues. It is possible that the EPA could work with holders of
30 private datasets to develop a similar system of broader applicability, that would provide a
31 mechanism for interested researchers to access datasets for reanalysis under appropriate controls,
32 where relevant for EPA regulations.
33

34 The lack of any criteria for what might satisfy the requirement to “make all such studies
35 available to the public to the extent practicable” makes it difficult to understand the implications
36 of this requirement. Criteria are needed to define the requirement. A question to be answered is
37 whether making the scientific papers reporting these studies available without charge makes the
38 studies “available,” or does all data from every measurement taken as part of the study need to be
39 available to anyone to manipulate? At one end of this range of interpretation the requirement is
40 easily implementable, on the other end of the spectrum it would be enormously expensive and
41 time consuming at best and can be expected to result in the exclusion of much of the scientific
42 literature from consideration (the machine data are no longer available and/or the researchers are
43 no longer alive or in a position to assemble the data). The net effect could be minimal or
44 massive. This is hard to evaluate because either outcome is plausible. Given the lack of clarity,
45 the proposed rule could be viewed as a license to politicize the scientific evaluation required
46 under the statute based on administratively determined criteria for what is practicable.

1
2 The SAB recommends that the EPA develop definitions of terms and methods for making
3 information available. In the Proposed Rule it is stated that “for the science pivotal to its
4 significant regulatory actions, EPA will ensure that the data and models underlying the science is
5 publicly available in a manner sufficient for validation and analysis.” As noted above, it is
6 unclear exactly what approaches the agency will utilize to ensure that information is available to
7 the public and to what extent this information should be accessible. EPA should clearly define
8 what is meant by “publicly available,” identify the method(s) that it will utilize to make this
9 information available, and discuss the level of effort required by the generator of the data to
10 assist EPA in complying with this proposed mandate. Additionally, it is also unclear what it
11 meant by “validation and analysis” so the Agency should clarify these terms.

12
13 *Feasibility of developing criteria to specify exceptions to the requirement to make information*
14 *available to the public*

15
16 The Proposed Rule indicates that there may be exceptions to the requirement to make
17 information available to the public. The SAB understands that exceptions may be needed but
18 notes that it would be difficult to develop criteria for exceptions because it seems almost
19 inevitable that unanticipated situations could come up that would not be covered by the criteria.
20 Alternatively, it is not clear that exceptions are appropriate if the goal is truly “to increase
21 transparency in the preparation, identification, and use of science in policymaking.” Some
22 members of the SAB find that conclusions from studies that cannot be independently verified
23 and reproduced due to unavailability of data or models should not be treated as pivotal for
24 regulatory action if the regulatory action is supposed to be based on independently verifiable and
25 reproducible science – which might be construed as part of the definition of “sound” science.

26
27 The SAB notes that the proposal to use a case-by-case “waiver” may not be an effective
28 mechanism for ensuring that the EPA can appropriately consider studies that rely on confidential
29 data. A case-by-case waiver may exacerbate concerns about inappropriate exclusion of
30 scientifically important studies. Reference to a vague “feasibility” standard means that such
31 waivers are decided upon solely by the Administrator, in the absence of clear guidance. This
32 approach could easily have the effect of politicizing science because key decisions about what
33 science is relevant to decision making could be made with no requirement to base decisions on
34 scientific evaluation. It might be useful for the EPA to consider recommendations from an
35 advisory committee when making waiver decisions.

36
37 The SAB finds that exclusion of segments of the scientific literature with the possibility of
38 inclusion of selected elements based on non-scientific considerations represents a significant
39 shift in science-based decision making. Such a change could easily undercut the integrity of
40 environmental laws, as it will allow systematic bias to be introduced with no easy remedy. The
41 proposed exception process applies no constraints on how this mechanism could be used or that
42 it be restricted to the issue of confidential data. Such a proposal is inconsistent with the scientific
43 method that requires all credible data be used to understand an issue and to allow systematic
44 review to evaluate past research.

45

1 Although they would be difficult to develop, the EPA could benefit from preparing a framework
2 and/or guidance that outline criteria to specify exceptions. While the EPA cannot address all
3 circumstances and scenarios that could limit data sharing, there are some reasonably anticipated
4 scenarios that the EPA could describe, perhaps with case studies from previous risk assessments
5 (e.g., What information would fall in scope based on past risk assessments? What raw data
6 would have to be released to support a given risk assessment? Is it feasible to release this
7 information? Why or why not?). EPA might also consider being more explicit regarding
8 criteria for making studies publicly available. Instead of developing specific exceptions to this
9 requirement, EPA might consider being more explicit regarding the criteria for inclusion and
10 evaluate specific exceptions on a case by case basis.

11 **3.3. Requirement to Ensure that Dose Response Data and Models Underlying Pivotal**
12 **Regulatory Science are Publicly Available in a Manner Sufficient for Independent**
13 **Validation.**

14 The Proposed Rule states that when promulgating significant regulatory actions, the agency shall
15 ensure that dose response data and models underlying pivotal regulatory science are publicly
16 available in a manner sufficient for independent validation. The SAB finds that greater clarity is
17 needed in definitions of “dose response data and models” and “pivotal regulatory science.” The
18 definitions are not adequate in the current Federal Register notice and may be better supported in
19 the context of a guidance document that includes realistic examples of the types of dose-response
20 and models of interest and requirements for reporting this information. It would be particularly
21 useful to clarify specific requirements for reporting information from environmental
22 epidemiology studies. The Proposed Rule indicates that pivotal regulatory science refers to
23 studies or analyses that drive the requirements and/or quantitative analysis of EPA final
24 significant regulatory decisions. Although this definition is adequate for some practical purposes,
25 the EPA should clarify whether this includes all the hazard characterization and dose response
26 models that the agency evaluated and captured in its analysis or only the final model ultimately
27 selected. A technical issue to be considered when defining dose response models and data is how
28 to separate data sets and models that were the actual basis of the calculations used to drive the
29 quantitative assessment (e.g., POD or RfD), from additional data sets and models may have been
30 ancillary but part of the weight of evidence used in the regulatory action, if appropriate to the
31 case under consideration. Reasons for the choice of the primary data sets and models for the
32 quantitative assessment and reasons for placing the other options in the weight of evidence
33 (WOE) category should be made available. Dose-response data might be defined as the analysis
34 data set (listing values of causally relevant exposure, response, and covariate variables and uses
35 multiple imputation and related methods for missing data and confidentiality); and dose-response
36 model might be defined as “the model or algorithm used to calculate conditional probability of a
37 stated health response caused by stated exposures together with the values of any other direct
38 causes of the response (e.g., stated levels of causally relevant covariates such as co-exposures
39 and co-morbidities).”

40 However, when defining these terms, some questions should be addressed: Dose-response data
41 are referenced. Are there any requirements for the number of dose levels or dose spacing? Does
42 this mean that single dose studies will be excluded? If not, under what circumstances would
43 these studies be included? What types of models are “in scope?” What type of information is
44 needed for each model type? Is the goal to provide equations, allowing the public to replicate the

1 math or to provide models with assumptions that the public can evaluate in detail? Are there
2 standard approaches for modeling dose response relationships (benchmark dose? other?) If so,
3 under what conditions are different dose response approaches implemented? Can a framework
4 be prepared to outline the EPA’s approach.

5 To make the Proposed Rule more practical, pivotal regulatory science could be defined as the
6 study (or studies if necessary) on which the regulatory limit is based. Ideally, the EPA should
7 provide a brief explanation of why the selected study is the pivotal study and why other studies
8 were not selected.

9 *Protecting Privacy and confidentiality*

10
11 The SAB notes that protecting privacy and confidentiality must be taken into consideration when
12 data and models underlying pivotal regulatory science are made available to the public. Although
13 the proposed rule suggests that privacy and confidentiality issues can be addressed through
14 anonymization or de-identification, this is not always the case. The U.S. Office of Management
15 and Budget (OMB 2013), Health Effects Institute¹, National Academies of Science (NRC 2005),
16 and independent experts (Rothstein 2010; Commission on Evidence-Based Policy Making 2017;
17 Rocher et al. 2019) have all found that even de-identified datasets present significant risks of
18 re-identification given modern techniques for combining these datasets with other sources of
19 individual information (also known as a “mosaic effect”). Without a specific plan for addressing
20 these substantive concerns it is difficult to understand the implications of the proposed rule with
21 respect to leading to exclusion of important data, delaying development of protective regulations
22 or risking peoples’ privacy. As previously noted, microaggregation of data can protect personal
23 identity in epidemiology studies.

24
25 Even where feasible, de-identification does not address all legitimate restrictions or constraints
26 on disclosure of data (such as proprietary information). The SAB finds that EPA must address
27 the critical issues of protecting privacy and confidentiality prior to taking action on the proposed
28 rule because the evaluation of potential solutions to the issues raised should involve a transparent
29 and open process. One SAB member notes that a contradiction in the proposed rule is that, in the
30 name of transparency, it puts forward opaque requirements that may allow decisions to be made
31 without transparency. The member notes that the proposal raises critical issues that potentially
32 challenge the integrity of the science used in making science-based decisions as required by law.

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

35
36 The SAB notes that a number of scientific and technical challenges must be addressed and
37 questions answered in order to make dose response data and models available in a manner
38 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.").

1 known in order to determine whether experimental data are valid (e.g., appropriate controls,
2 protocols employed, where data fall within the standard curve of the target analyte, limit of
3 quantitation, dynamic range of the instrument, calibration of instruments, condition of
4 experimental animals, blinding of reading of slides or behavioral observations, stability of
5 samples, qualifications and approach of researchers obtaining epidemiological data). Also there
6 is a need to know what information was used in models (e.g., whether all data points were used,
7 how confounders were handled in epidemiological studies, what statistical models were used,
8 what predictive models were used, the fit of the model to the data). Independent validation
9 requires sufficient information about how the original data were collected and analyzed in order
10 to know whether the validation procedures are likely to yield the same result as the original
11 calculations.

12
13 It would be useful for the EPA to define “independent validation.” This drives the feasibility of
14 whether EPA can make dose response data and models available for independent validation. For
15 example, how much information is sufficient for independent validation? Will this consist of
16 equations where reviewers can verify the math or more detailed models where assumptions and
17 limitations are described, or code to allow the public to evaluate and run the models if desired? Is
18 this simply the dose-response data for the endpoint of concern driving a regulatory limit, or is it
19 availability of all data from a pivotal study to allow reviewers to examine the potential
20 contributions of other variables on the primary endpoint of concern? Endpoint data are seldom
21 evaluated in isolation so providing sufficient study information to allow an independent
22 assessment seems important to meet the goals of the Proposed Rule. For example, an effect on
23 pup body weights in a toxicology study should be examined with knowledge of maternal
24 gestational body weight gains, litter size, food consumption, maternal/litter clinical signs, etc.
25 Sample size and variability also play a key role in data interpretation.

26
27 The SAB finds that the following specific questions should be answered in order to assess the
28 technical feasibility of making data and models available in a manner sufficient for independent
29 validation

- 30
- 31 1. Good laboratory practice (GLP) studies include full disclosure of study data, statistics,
32 instrument calibration, positive control data, etc. Would the EPA make all of these data
33 available to the public on any study judged to be “pivotal”? Would this include GLP
34 audit findings? Will any information in the study be judged by EPA to be CBI?
35
 - 36 2. With respect to non-GLP/investigational studies, laboratories may not have data
37 formatted in a manner that makes these data easily shared with the public. How will EPA
38 manage this, particularly when several years may elapse between conduct of a study and
39 its designation as a “pivotal study”?
40
 - 41 3. How will the EPA move forward if laboratories refuse to collate and release their data for
42 public review? Will these data be excluded from use in risk assessment? If so, there is
43 potential to bias risk assessments based on the exclusion of data that investigators chose
44 not to make publicly available. In contrast, perhaps these data, without public scrutiny,
45 should not be used for regulatory decision-making.
46

- 1 4. How will the EPA manage international studies, where there is no requirement for
2 laboratories to provide data to EPA? EPA may have little to no leverage in these
3 situations. Again, the inability to use “pivotal studies” from other geographies could bias
4 risk assessment decisions.
5
- 6 5. How will the EPA manage conclusions drawn from a meta-analysis? Do all studies
7 included in the analysis become “pivotal” studies? This could markedly increase the
8 number of data sets that must be publicly available.
9
- 10 6. How will the EPA justify identification of a “pivotal study” and dose-response without
11 clarifying why other studies were not pivotal? This will require transparent evaluation
12 and reporting of data quality/reliability for available studies that allows reviewers to
13 understand EPA’s selection of the “pivotal” study.
14
- 15 7. Some dose response data and models may rely on proprietary software that may not be
16 readily accessible or available while other software may be accessible but require
17 considerable data storage or download that may limit utility and availability. How will
18 these models be validated?
19

20 It may be beneficial for EPA to develop a guidance document with case studies based on past
21 risk assessments to clarify some scenarios and how the requirement to make dose response data
22 and models publicly available in a manner sufficient for independent validation would be
23 managed (i.e., what is ‘in scope’ vs. ‘out of scope’ and rationale). The lack of details on
24 implementation of this rule allows for an endless number of hypothetical scenarios, so a
25 guidance document may offer clarity at least for the most common scenarios.
26

27 *Assessing the validity of epidemiological studies*

28

29 The SAB finds that assessing the validity of epidemiological studies for the purposes of the
30 proposed rule poses particular scientific and technical challenges. In general, for the purposes of
31 the Proposed Rule, validation should be defined to include both internal validity and external
32 validity, in the senses defined by Campbell and Stanley (1963). Issues to be addressed include
33 understanding bias, confounding factors, measurement errors and exposure characterization. All
34 of these factors will play a role in defining what would be appropriate for access and validation
35 purposes. Specifically, one would need to know: how measurements were taken; how
36 confounders were assessed and dealt with; the institutional review board (IRB) application and
37 subsequent approvals or concerns; the Informed Consent Form (or assent process for children
38 capable of providing assent) or the consent of parents or guardians for information collection on
39 children; the qualifications of the researchers obtaining personal or health information and the
40 consistency of multiple researchers for collection of PII, how truncated data sets for longitudinal
41 studies were handled when participants dropped out of the study or missed sampling times; how
42 environmental samples or human blood samples were taken, handled, stored and analyzed;
43 criteria for how any data points were deemed outliers and eliminated from the data set; how
44 participants were selected and what the selection/exclusion criteria were.
45

1 *Challenges and costs of processing and documenting of data prior to public release,*
2 *maintenance and administration of data sets so they are publicly available, and handling*
3 *historical data sets*
4

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

11
12 Obtaining data in a useable format with adequate documentation may be difficult. It is likely that
13 not all data will be in a format suitable for public data sharing (e.g., non-traditional data formats,
14 PII, CBI, inadequate documentation of data or methods). The SAB notes that there may be
15 solutions for some of these data issues, but not for others.

16
17 Historical data sets might not be available at the level of detail needed for recalculation. Some of
18 the raw data or computational methods may have been discarded if they were deemed not
19 necessary to maintain. Certainly IRB applications usually indicate when individual records can
20 be discarded.

21
22 The people or groups processing and handling the data (EPA staff or independent non-EPA
23 consultants) would need to be identified, their credentials and any conflicts of interest with the
24 particular case identified, and documentation secured that they will not reveal confidential
25 information without appropriate permission from the owners of the CBI; PII should not be
26 revealed. The processing might include ways to strip some of the PII of potential identifying
27 information or aggregating the information if these methods would still allow for the validation
28 to be performed.

29
30 Similarly the people or groups who would maintain the data sets and provide them to
31 independent public people or groups for validation processes would need to be identified and
32 they must assure that the data will be maintained as confidential PII or CBI and only released to
33 authorized people or groups. A mechanism would need to be developed for public requests for
34 access and approval or disapproval of the request.

35
36 The SAB suggests that EPA consider establishing an office (or virtual office) on data sharing.
37 This group could identify standard data formats (data templates?), how to report
38 methods/procedures used, uncertainty, when and how to implement greater data protections for
39 PII/CBI, etc. It would be beneficial to build experience and expertise in a group charged to meet
40 this goal. This office could work directly with laboratories/researchers to provide the necessary
41 information in a “user friendly” format. This office also could build and manage data archives
42 and pursue critical historical data sets if deemed important. There will be costs associated with
43 the establishment of such an office as well as researchers’ time to collate raw data and work with
44 EPA to make these data publicly available. It is unclear how the EPA will manage these
45 additional costs. In the future, it might be possible for EPA to develop a reporting framework for
46 laboratories so that study data are collected in a format that requires less rework if a study is

1 subsequently judged to be a “pivotal study.” Some laboratories/researchers may not want to
2 organize historical data for public release as they may see this activity as a diversion from their
3 research priorities. The EPA may find creative ways to offset the expense associated with data
4 submission for pivotal studies.

5
6 *Processing and documenting dose response data and models developed prior to the effective*
7 *date of the rule*
8

9 The SAB notes that processing and documenting dose response data and models developed prior
10 to the effective date of the rule will pose challenges. A possible way to address this is to apply
11 rule requirements only to information developed after the effective date of a final rule. However,
12 it is most likely that some flexibility is going to be required. Standards on transparency are
13 evolving, and modern expectations do not apply to studies completed 10 or 20 years ago. It is
14 reasonable to apply modern standards of transparency and public availability to current and
15 future studies, but it will not always be possible to apply these same standards retrospectively. In
16 those situations, EPA should seek, support, and encourage future studies to replace historical
17 studies lacking transparency, but this cannot be done instantly. It will be a process over several
18 years.

19
20 *Prospective or retrospective application of provisions for ensuring the public availability of dose*
21 *response data and models underlying pivotal regulatory science*
22

23 The SAB notes that the retrospective application of modern transparency standards is a
24 challenge. A large amount of work would be required to locate, curate and retrospectively make
25 data sets available for public access. This requirement could adversely affect the ability to move
26 this program forward in a meaningful capacity. EPA could decide not to apply the Proposed Rule
27 and its specific requirements retrospectively given the potential difficulty accessing, reviewing
28 and making data available that were not originally intended to be disseminated in such a manner
29 as defined in a future rule. The EPA could consider designating a “start date” and begin
30 collecting and releasing pivotal study data at that time. When the EPA updates an existing risk
31 assessment after the start date, the EPA could collect and release pivotal data.

32
33 **3.4. Requirement to Make the Raw Data Underlying any Proposed Rule Available for**
34 **Public Inspection**
35

36 Under the Proposed Rule, EPA would require that raw data underlying any proposed rule be
37 made available for public inspection. The SAB finds that the requirement in the Proposed Rule
38 that “raw data” be made available for public inspection is vague and, as a result, can be
39 interpreted in different ways. If “raw data” includes all machine output associated with analysis
40 it would create demands on researchers that would be very onerous and could significantly slow
41 down science-based decision making without any demonstrable benefit. Even if the “raw data”
42 were accessible, making it publicly available in a useable form would be costly and could be of
43 limited utility based on past experience of the scientific community relative to the interpretation
44 of the derivative data. If the data required to meet the “raw data” requirement is no more than the
45 current standard of most journals (and in most cases provided in supplementary information)
46 then the implications prospectively are minimal. Either way retrospective application of the

1 requirement would be difficult to implement, could be expensive with no clear responsibility
2 regarding who would cover the added costs and could impact the conclusions drawn arbitrarily.

3
4 The Proposed Rule should be explicitly prospective and follow evolving norms developed by the
5 scientific community as well as federal agencies (National Science Foundation, National
6 Institutes of Health, National Oceanic and Atmospheric Administration). The lack of reference to
7 and comparison with existing and evolving federal procedures to address the underlying purpose
8 of the Proposed Regulation (increased transparency) within scientific studies supported and
9 utilized by the federal government undermines the confidence that the proposed regulation will
10 meet its objectives effectively and efficiently.

11
12 The Proposed Rule is based on the premise that using scientific information that can be
13 independently validated will lead to better outcomes and strengthen public confidence in the
14 health and environmental protections underpinning EPA’s regulatory actions. The SAB notes
15 that public disclosure of data is helpful in enabling reanalysis that can assure the quality of the
16 work, but it is not a *sine qua non* for validating scientific studies. The scientific community uses
17 numerous tools to validate studies without access to all underlying “raw data,” including rigorous
18 peer review, replication of a study using the same methodology but with different data sources,
19 or reproduction of a study's conclusions using different methodologies and data. The EPA's
20 proposed policy of excluding from consideration any study for which underlying data are not
21 made publicly available is not consistent with sound scientific practice.

22
23 The SAB notes that there are legitimate legal, ethical, professional and financial reasons why
24 researchers may be unable or unwilling to fully share “raw data” - including statutes protecting
25 patient privacy, experimental protocols assuring confidentiality of data for human subjects, and
26 (for past studies) issues related to degradation and custody of data. The EPA, the U.S. Office of
27 Management and Budget, and scientific institutions have recognized these legitimate concerns,
28 and recognized that such constraints on availability of data do not prevent studies from being
29 validated in other ways - much less require that the studies be ignored in regulatory
30 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

1
2 There appears to be consistency among analyses of how to address transparency that are
3 orthogonal to the proposed rule. There is no justification in the Proposed Rule for why EPA finds
4 that existing procedures and norms utilized across the U.S. scientific community, including the
5 federal government, are inadequate, and how the Proposed Rule will improve transparency and
6 the scientific integrity of the regulatory outcomes in an effective and efficient manner. It is
7 plausible that in some situations, the Proposed Rule will decrease efficiency and reduce scientific
8 integrity, determining if in fact that will be the case requires a thorough and thoughtful
9 examination that is currently absent in the Proposed Rule. Moving forward with altered
10 transparency requirements beyond those already in use, in the absence of such a robust analysis,
11 risks serious and perverse outcomes.

12
13 *Feasibility of making raw data available for public inspection*

14
15 In the absence of more detailed information, it is difficult to determine the feasibility of making
16 raw data available for public inspection as required in the Proposed Rule. There are numerous
17 reasons why it may be difficult to share raw data including PII (e.g., HIPAA requirements), CBI,
18 non-standard data collection systems (e.g., not easily accessed outside). Time and cost associated
19 with formatting, curating and sharing raw data also may pose difficulties for laboratories. The
20 SAB recommends the development of definitions to clarify the requirement of making raw data
21 available. Although EPA wants to make data available to allow others to reproduce results,
22 privacy considerations (e.g., like those mandated in HIPAA) must be respected. “Raw data”
23 should not be confused with personally-identifiable data.

24
25 Given the reasons why sharing raw data may be difficult, the method(s) that EPA identifies and
26 utilizes for making “raw data” available for public inspection will impact significantly the
27 feasibility of this requirement. In order to determine the feasibility, EPA must consider: (1) what
28 level of detail it will plan to make data available (e.g. a robust summary of the available data, all
29 of the in-life and necropsy data from a specific study, only data that are included in the scientific
30 publication or publicly available report or publication); (2) the type of platform EPA will use to
31 make this information available (e.g., free publicly available online database with searchable
32 features, electronic cd rom or hard copies available by public request); and (3) requirements the
33 public will have to adhere to/agree to in order to access the data to ensure protection of relevant
34 information (e.g., formal data transfer agreements between the agency and entity requesting
35 access to the data, general electronic agreement and check box noting that the data user agrees to
36 meet certain pre-established requirements).

37
38 The SAB notes a number of additional concerns about the feasibility of making raw data
39 available for public inspection. Raw data would likely include all data points collected, some of
40 which might have been legitimately removed from the final analysis, e.g., there is knowledge of
41 a mistake in the collection of a particular data point, or it is determined statistically to be an
42 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.”)

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

14
15 Regardless of how “raw data” is defined, the requirement to make data available for public
16 inspection will be more easily implemented for some data sets than others (e.g., GLP studies
17 include individual sample data as part of standard reporting). The SAB notes that the expectation
18 that raw data and methods will be available for all endpoints may be unrealistic. EPA could
19 consider designating scores for information availability with the highest tier assigned to studies
20 that define methods (e.g., protocol) as well as raw (individual sample level) data. Additional
21 scores could be driven by sharing raw data on other related endpoints in the pivotal study to
22 facilitate data interpretation. Scores could subsequently be used to evaluate data utility,
23 uncertainty, etc. Ultimately, there are some data sets where CBI and/or PII data will require more
24 onerous steps to protect data confidentiality. Based on EPA responses to the SAB questions, it
25 appears that the agency is seeking approaches to manage these data issues.

26 27 *Definition of raw data*

28
29 It is difficult to develop a singular definition of “raw data” that would meet EPA’s objectives in
30 proposing this rule. Raw data could be defined as the *analysis data set*, as previously discussed.
31 Due to the depth of data included as “raw,” the definition of raw data would likely differ based
32 on the available data set. For example, raw data for an in vivo study could include all the
33 individual animal body weight or individual pathology data while a raw data set, for an in vitro
34 study may include multiple samples and assays assessed, and for epidemiology data it may
35 include individual exposure monitoring data or biological samples.

36
37 Wikipedia defines raw data (also known as primary data) as “data collected from a source.” This
38 would include individual sample values collected on individual study subjects or various
39 instruments and include each measurement on sampled endpoints (and each time it is sampled) in
40 a given study. Raw data would not be manipulated in any fashion (e.g., removing outliers - data
41 would appear in the raw data with a scientific rationale on why a data point was removed from
42 subsequent analysis).

43
44 Notably, in 40 CFR part 792 -TSCA which describes Good Laboratory Practices, raw data is
45 defined as “any laboratory worksheets, records, memoranda, notes, or exact copies thereof, that
46 are the result of original observations and activities of a study and are necessary for the

1 reconstruction and evaluation of the report of that study.” All or none of this information may or
2 may not be needed to adequately understand the results of the study or EPA's use of the
3 information. Thus making all of this information publicly available may be resource intensive
4 and provide limited utility.

5
6 It might be reasonable to consider the initial compilation of data in a spreadsheet as raw data.
7 While this might still require some interpretation, e.g., abbreviations used in spreadsheet column
8 headings, and footnotes about missing data points (if they were discarded because of legitimate
9 reasons, such as known mistakes or statistically- determined outliers), the spreadsheets are still
10 very close to the initial raw data coming out of an instrument or written down by a researcher
11 when working with a subject in an epidemiological study.

12
13 As previously stated, it is difficult to evaluate the impacts of a definition of raw data a priori.
14 There is extensive work required to understand the implications of different definitions across a
15 diversity of fields, data types and data of different ages. Such an effort is beyond the scope of
16 what the SAB can undertake with the resources and time available. However, the SAB finds that
17 such an analysis is foundational to the development of any transparency rule that goes beyond
18 well-established norms and procedures.

19
20 Another aspect to consider is the practical aspect of actually conducting a reanalysis of a major
21 epidemiological study. Such an enterprise requires an enormous amount of work even for a well-
22 qualified researcher. The Health Effects Institute (HEI) established a model for conducting such
23 a reanalysis in its 2000 reanalysis of the Six Cities and American Cancer Society datasets (HEI
24 2000). However, HEI has not repeated this kind of exercise and to the best of our knowledge has
25 no plans to do so in the future. EPA could consider using the HEI model for funding its own
26 reanalyses, for datasets that are deemed critically important for regulation.

27
28 **3.5. Requirement to Describe and Document any Assumptions and Methods Pertaining to**
29 **the use of Dose Response Data and Models Underlying Pivotal Regulatory Science**

30
31 The Proposed Rule requires the EPA to describe and document any assumptions and methods
32 that pertain to the use of dose response data and models underlying pivotal regulatory science
33 and to describe variability and uncertainty. The SAB notes that high quality scientific studies
34 identify the assumptions used in models and methods used, the variability of the replications, and
35 any other confounders that add to the uncertainty of the final data set, so these are not unusual or
36 inappropriate factors that need to be addressed. Specifically, it is good practice to identify: (1)
37 testable assumptions (e.g., that residuals in a regression model are normally distributed with
38 constant variance); (2) results of tests of assumptions; and (3) any untested assumptions (e.g.,
39 that there are no unmeasured confounders, or that dose-response functions are linear below the
40 lowest dose for which data are available) made in deriving conclusions from data. Results of
41 tests should be presented where they are available, and assumptions that have not been tested, or
42 that have been tested but not supported, should be identified. Assumptions are often made about
43 (a) error distributions for exposure estimates (most commonly, that they can be ignored); (b)
44 model specification errors and uncertainties; and (c) causal interpretations of modeling results.
45 These assumptions should be explicitly stated and results of tests presented. Epidemiologists
46 (e.g., Sander Greenland), statisticians, and risk analysts have written at length over several

1 decades about how to test, validate, and document assumptions and methods for dose-response
2 modeling and uncertainty and variability characterization, and these modern methods should be
3 applied to make the factual and assumption-based foundations of pivotal studies as clear as
4 possible. However the SAB finds that there are scientific and technical challenges to be
5 overcome and provides suggestions to implement this requirement.

6
7 Currently in various chemical assessment processes EPA program offices attempt to document
8 the methods, assumptions, variability and uncertainty associated with the use of various dose
9 response models and data inputs utilized. EPA has generally done this in a qualitative format and
10 should continue to refine and document this information.

11
12 One would anticipate variability in the reporting of assumptions, methods, variability, and
13 uncertainty across laboratories. Therefore, EPA could offer guidance on how to report these
14 parameters. When this information is received from submitting laboratories, EPA could review
15 the information to determine if methods, uncertainty, etc. are adequately addressed with possible
16 follow up as needed. There are numerous resources available from which the EPA could develop
17 guidance, including some defined by the EPA (Maurissen, 2010; U.S. EPA 2012; U.S. EPA.
18 2019). The SAB notes that it may be difficult to identify the appropriate level of detail to define
19 some parameters like methods. This will vary depending on factors like how new/novel the
20 method is, how many variables impact outcome, etc. Many laboratories may not want to share
21 standard operating procedures for public release by the EPA.

22
23 The Proposed Rule is viewed by some members of the SAB as inappropriately codifying certain
24 required scientific approaches into regulation. In support of this view, some members note that:
25 (1) the preamble of the Proposed Rule asserts without providing any evidence that “there is
26 growing empirical evidence of non-linearity in the concentration-response function for specific
27 pollutants and health effects,” and (2) the preamble does not acknowledge the value of default
28 models (in this regard, these members also note that the National Academies' 2009 report
29 *Science and Decisions* (NRC 2009) recommended the use of linear non-threshold models as a
30 default for both cancer and non-cancer responses). It is critical that the justification for
31 alternative approaches is well documented and transparent, and some members of the SAB find
32 that is not currently the case. Therefore, the SAB recommends that the preamble of the proposed
33 rule include some specific examples or case studies where there is evidence in the published
34 literature demonstrating threshold responses which provide support for the preamble statements
35 (e.g., Sweeney et al. 2009; Johnson et al. 2014; Cohen et al 2016). The SAB notes that the EPA’s
36 2005 Guidelines for Carcinogen Risk Assessment allow the agency to consider nonlinear
37 approaches after an analysis of available data under the guidance provided in the framework for
38 mode of action analysis. Notably, where alternative approaches have significant biological
39 support, and no scientific consensus favors a single approach, an assessment may present results
40 using alternative approaches (i.e. a nonlinear approach). It may be useful for the Agency to
41 consider developing guidance or criteria delineating when data may be sufficient to apply
42 alternative approaches.

43
44 One SAB member notes that the Proposed Rule emphasis on consideration of studies that use a
45 diversity of models is arbitrary. The member points out that there is no scientific basis provided
46 in the Proposed Rule for giving greater weight to studies that use a wider variety of models

1 without regard for goodness-of-fit, confidence bounds, biological plausibility, attention to
2 untested assumptions. The member notes that the lack of a scientific basis for this approach and
3 the lack of transparency with respect to the logic underlying the proposed approach undercuts
4 confidence that goals of the proposal can be accomplished without perverse effects. The member
5 also notes that this indicates the need for a peer-reviewed transparent analysis in which all
6 assumptions, data and conclusions are made available to the public before the Proposed Rule is
7 enacted.

8 9 **3.6. Protecting Sensitive Data and Copyrighted or Confidential Business Information**

10
11 The Proposed Rule preamble states that nothing in the Proposed Rule compels the disclosure of
12 any confidential or private information in a manner that violates applicable legal and ethical
13 protections. The SAB provides suggestions and recommendations to protect sensitive data and
14 copyrighted or confidential business information.

15
16 It seems reasonable that the standards applied by the EPA should be the same as the standards
17 applied by editors of reputable scientific journals. A report of the National Research Council
18 (NRC 2003) report discusses responsibilities of authors to share data, software, and materials
19 related to their publications. Techniques and practices to protect sensitive data have been
20 developed by researchers involved in studies with human subjects. Other federal agencies have
21 utilized specific and explicit data transfer agreements with entities that sought access to protected
22 information for the purposes of reviewing and running analysis on the data set. The EPA could
23 employ similar approaches. Any new rule needs to build on those efforts. Coding data to protect
24 any personal identifiers is routinely done, and may have already been done with the data sets
25 under consideration. If aggregation of some of the data (e.g., by age bands) does not compromise
26 the ability of the calculations to be done, then this would protect personal information. Existing
27 methodologies and technologies already in widespread use (such as those used in NHANES III
28 (National Center for Health Data Statistics 2019), including multiple imputation) can be used to
29 provide protected access to data. Additional technologies worth considering include differential
30 privacy techniques (Dankar and El Emam 2012) and perhaps Bayesian deep learning and other
31 approaches that model joint distributions for variables in a data set and that can use them to
32 generate anonymous data exchangeable with the original data.

33
34 As previously mentioned, the agency could develop tiers of public access that may include a base
35 level of information (e.g., a robust summary or final study report which is devoid of confidential
36 protected information) to the general user or member of the public and then require additional
37 information or data sharing agreements when access to confidential or copyrighted information is
38 warranted. Copyright would protect the particular program or model from being used by others
39 without permission. Signing a non-disclosure agreement would protect other CBI information. A
40 model for accessing personal information contained in federal datasets already exists in the
41 Federal Statistical Research Data Centers (FSRDCs), where individual researchers can access
42 individual-level data under conditions that guarantee the confidentiality of personal information.
43 EPA could consider negotiating with other federal agencies to provide access through the
44 FSRDCs for federal data that are used in epidemiological research, such as Medicare.

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

6
7 For CBI and PII, it is possible that access may be possible, but limited in some cases. For
8 example, members of the public could petition the EPA for access to some sensitive data and the
9 agency could take countermeasures to provide only the information permissible and control the
10 settings around which these data are made available (e.g., onsite access only). This option would
11 likely require the EPA to maintain a “data office” which would require substantial resources to
12 establish and support.

13 14 **3.7. Other SAB Comments on the Proposed Rule**

15
16 The SAB provides the following additional comments and concerns about the Proposed Rule.

- 17
18 1. As previously discussed, the following language in the preamble of the Proposed Rule
19 has been interpreted by some members of the SAB as an abandonment of the linear no-
20 threshold models. However, other SAB members view it as a policy change indicating
21 that evidence of nonlinearity should not be ignored when there is scientific evidence to
22 support it. They note that EPA has explicitly identified linear models, along with others,
23 in the list of models to be considered, and that support for alternative models is indicated
24 in the EPA’s 2005 Cancer Guidelines (U.S. EPA 2005). It is important that the EPA
25 provide clarity as to the agency’s intent:

26
27 “...there is growing empirical evidence of non-linearity in the concentration-response
28 function for specific pollutants and health effects. The use of default models, without
29 consideration of alternatives or model uncertainty, can obscure the scientific
30 justification for EPA actions. To be even more transparent about these complex
31 relationships, EPA should give appropriate consideration to high quality studies that
32 explore: A broad class of parametric concentration- response models with a robust set
33 of potential confounding variables; nonparametric models that incorporate fewer
34 assumptions; various threshold models across the exposure range; and spatial
35 heterogeneity. EPA should also incorporate the concept of model uncertainty when
36 needed as a default to optimize low dose risk estimation based on major competing
37 models, including linear, threshold, and U-shaped, J-shaped, and bell-shaped
38 models.”

39
40 On page 1-8 the EPA Cancer Guidelines state:

41
42 “When there are alternative procedures having significant biological support,
43 the Agency encourages assessments to be performed using these alternative
44 procedures, if feasible, in order to shed light on the uncertainties in the
45 assessment, recognizing that the Agency may decide to give greater weight to
46 one set of procedures than another in a specific assessment or management

1 decision. Encouraging risk assessors to be receptive to new scientific
2 information, NRC discussed the need for departures from default options when
3 a sufficient showing is made.”
4

- 5 2. In general, it is difficult to define universal “rules of good behavior” for as many-faceted
6 a question as choosing the right dose-response model. In some situations, direct
7 biological arguments may support a particular model such as a linear or log-linear dose-
8 response model. In other cases, the use of any model as a default choice may be
9 inappropriate. The biostatistics literature includes many procedures for identifying and
10 validating appropriate models, including techniques such as Bayesian Model Averaging
11 which avoid the specification of a single model. There is no “one size fits all” approach
12 that could be applied to all problems of this nature.
13
- 14 3. The SAB is concerned that key considerations that should inform the rule appear to be
15 omitted or presented with no analysis - for example, it is not clear how many of the
16 studies EPA currently relies upon to take important regulatory actions would meet the
17 public disclosure standards in the proposal, whether EPA has assessed how many of those
18 studies would be feasible to provide underlying information, or what the impact of
19 precluding those studies would be on EPA's decision making and its ability to protect
20 public health / environment.
21
- 22 4. As previously discussed, the SAB is concerned that key terms and implementation issues
23 in the Proposed Rule are left undefined or too vague to enable informed comment (e.g.,
24 “sufficient for independent validation,” “feasible” (for purposes of Administrator's
25 waiver), “replication,” “validation,” “publicly available”). In particular the lack of a
26 functional definition of “data” and “raw data” in the Proposed Rule as it relates to peer-
27 reviewed studies is such an important omission that it seriously undercuts the ability to
28 understand the implications of the proposed rule. In particular the SAB notes that the
29 following questions should be answered: Would it suffice that researchers make available
30 the data on which they performed the bulk of their calculations (the analysis dataset),
31 which typically follows preprocessing or aggregation, or does EPA expect full raw data
32 down to the level of individual measurements, including data directly used to carry out
33 the reported statistical analysis and model development? What level of detail would be
34 provided (e.g., if subjects are removed from the study, would these subjects be identified
35 and reasons given)? How does the role of QA/QC methodology affect the choice of
36 “data” in the stages of aggregation that would be released?
37
- 38 5. The SAB is concerned that the net effect of the proposed rule could be to inappropriately
39 limit the science used by EPA and rigor of decision-making - impeding EPA's ability to

Science Advisory Board (SAB) Draft Report (10/16/19) – Do Not Cite or Quote.

This draft has not been reviewed or approved by the chartered SAB and does not represent EPA policy.

1 protect the public. Evidence to support this concern has been presented by the editors of
2 scientific journals and the Congressional Budget Office.^{6,7,8}
3

4 6. The SAB is concerned that the Proposed Rule contains contradictory statements about
5 whether its effect would be to prohibit EPA from considering studies for which
6 underlying data cannot be made publicly available. Major public health protections, like
7 the National Ambient Air Quality Standards (NAAQS) for ozone and oxides of nitrogen,
8 drinking water standards for arsenic, and Toxic Substances Control Act (TSCA)
9 standards for formaldehyde emissions in composite wood rely on studies that rest on
10 confidential data.

11
12 7. The Administrative Procedure Act requires that a proposed rule provide enough
13 information for interested parties to “comment meaningfully.” As stated above the SAB
14 finds that additional information about the proposed rule is needed.

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.

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%.

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

- 1
2
3 Campbell, D. T., and J. Stanley. 1963. *Experimental and Quasi-experimental Designs for*
4 *Research*. Chicago, IL: Rand McNally.
5
6 Cohen, S.M., A. Chowdhury, and L.L. Arnold. 2016. Inorganic arsenic: A non-genotoxic
7 carcinogen. *Journal of Environmental Sciences*,49:28-37. Commission on Evidence-Based
8 Policymaking. 2017. *The Promise of Evidence-Based Policymaking*.
9 <https://www.cep.gov/content/dam/cep/report/cep-final-report.pdf>. [Accessed, August 18, 2019]
10
11 Dankar, F.D., and K. El Emam. 2012. The application of differential privacy to health data. In:
12 *Proceedings of the 2012 Joint EDBT/ICDT Workshops*, pages 158–166. ACM
13 [Available at: http://www.mathcs.emory.edu/pais12/pdf/PAIS12_Paper2.pdf]
14
15 HEI (Health Effects Institute). 2000. *Reanalysis of the Harvard Six Cities Study and the*
16 *American Cancer Society Study of Particulate Air Pollution and Mortality*. Special Report,
17 Health Effects Institute, Boston, MA.
18
19 Johnson, N.M., P.A. Egner, V.K. Baxter, M.B. Sporn, R.S. Wible, T.R. Sutter, J.D. Groopman,
20 T.W. Kensler, and B.D. Roebuck. 2014. Complete protection against aflatoxin B1-induced liver
21 cancer with a triterpenoid: DNA adduct dosimetry, molecular signature, and genotoxicity
22 threshold. *Cancer Prevention Research*, 7(7):658-665.
23
24 Maurissen, J. 2010. Practical considerations on the design, execution and analysis of
25 developmental neurotoxicity studies to be published in *Neurotoxicology and Teratology*.
26 *Neurotoxicology and Teratology* 32:121–123.
27
28 National Center for Health Data Statistics. 2019. *NHANES III (1988-1994)*.
29 <https://wwwn.cdc.gov/nchs/nhanes/nhanes3/Default.aspx> [Accessed August 23, 2019]
30
31 NRC (National Research Council). 2003. *Sharing Publication-Related Data and Materials :*
32 *Responsibilities of Authorship in the Life Sciences*. National Academies Press, Washington DC
33 [Available at: [https://www.nap.edu/catalog/10613/sharing-publication-related-data-and-](https://www.nap.edu/catalog/10613/sharing-publication-related-data-and-materials-responsibilities-of-authorship-in)
34 [materials-responsibilities-of-authorship-in](https://www.nap.edu/catalog/10613/sharing-publication-related-data-and-materials-responsibilities-of-authorship-in)]
35
36 NRC (National Research Council). 2005. *Expanding Access to Research Data: Reconciling*
37 *Risks and Opportunities*. The National Academies Press.
38
39 NRC (National Research Council). 2009. *Science and Decisions: Advancing Risk Assessment*.
40 Washington, DC: The National Academies Press.
41 [Available at: <https://doi.org/10.17226/12209>]
42
43 Reiter, J.P. and T.E. Raghunathan. 2007. The multiple adaptations of multiple imputation.
44 *Journal of the American Statistical Association* 102:1462-1471 [Available at:
45 <https://pdfs.semanticscholar.org/6383/01ae5456419b1e49effec158bf48f32fdf3b.pdf>]
46

- 1 Rocher, L, J.M. Hendrickx, and Y.-A. de Montjoye. 2019. Estimating the success of re-
2 identifications in incomplete datasets using generative models. *Nature Communications* 10:3069
3 <https://doi.org/10.1038/s41467-019-10933-3>
4
- 5 Rothstein, M. 2010. Is deidentification sufficient to protect health privacy in research? 10.9 *The*
6 *American Journal of Bioethics* 3-11, 6.
7
- 8 Sweeney, L.M., C.R. Kirman, R.J. Albertini, Y.M. Tan, H.L. Clewell III, J.G. Filser, G.
9 Csanády, L.H. Pottenger, M.I. Banton, C.J. Graham, and L.S. Andrews. 2009. Derivation of
10 inhalation toxicity reference values for propylene oxide using mode of action analysis: example
11 of a threshold carcinogen. *Critical reviews in toxicology* 39(6):462-486.
12
- 13 U.S. EPA. 2005. *Guidelines for Carcinogen Risk Assessment*. EPA/630/P-03/001F. Risk
14 Assessment Forum, U.S. Environmental Protection Agency Washington, D.C.
15
- 16 U.S. EPA 2012. *Guidance for Considering and Using Open Literature Toxicity Studies to*
17 *Support Human Health Risk Assessment*. U.S. EPA Office of Pesticide Programs, Washington,
18 D.C.
19 [Available at: <https://www.epa.gov/sites/production/files/2015-07/documents/lit-studies.pdf>]
20
- 21 U.S. EPA. 2019. *Uncertainty and Variability*.
22 <https://www.epa.gov/expobox/uncertainty-and-variability> [Accessed August 18, 2019]
23
- 24 U.S. Office of Management and Budget. 2013. Office of Management and Budget Memorandum
25 M-13-13, *Memorandum for the Heads of Executive Departments and Agencies on Open Data*
26 *Policy-Managing Information as an Asset 4-5 (May 9, 2013)*
27