July 25, 2019

EPA Responses to SAB Questions Concerning the Proposed Rule *Strengthening Transparency in Regulatory Science*

Science Advisory Board (SAB) Science and Transparency Work Group Questions for EPA Concerning the Proposed Rule *Strengthening Transparency in Regulatory Science*

SAB Questions and EPA Responses

Definition of Data

1. For the purposes of this rule, what is the definition of the "data" underlying a peer-reviewed study? In particular, would it suffice that researchers make available the data on which they performed the bulk of their calculations (the analysis dataset), which typically follows some initial preprocessing or aggregation, or does EPA expect full raw data down to the level of individual measurements, including data directly used to carry out the reported statistical analysis and model development? What level of detail would be provided (e.g., if subjects are removed from the study, would these subjects be identified and reasons given)?

   EPA Response

   The proposed rule includes a proposed definition of "research data," although does not separately propose a definition of "data." The proposed regulatory text generally uses the term "data." Several commenters raised questions that are similar to those that you ask and requested that EPA clarify the scope of "data" for purposes of the rulemaking. EPA is evaluating these public comments along with your questions and will clarify in its response to comments and the final rule the scope of the term "data". EPA's evaluation of this specific issue will also be informed by the advice the SAB will provide in the consultation on existing mechanisms for secure access to confidential business information and personally identifiable information.

2. How does the role of QA/QC methodology affect the choice of "data" in the stages of aggregation that would be released?

   EPA Response

   EPA received many public comments from a range of commenters on whether aggregated data would be sufficient to meet the proposed rule's goal of transparency. To a lesser extent, EPA received public comment on QA/QC. In developing the final rule EPA will consider public comments on data aggregation and on QA/QC related to making data available. It should be noted that the QA/QC EPA currently conducts does not drive a determination of what data would be released.

Validation of Studies

3. How does EPA define "replication," "validation," and "publicly available" for the purposes of this rule?
   - Does "replication" consist of anything other than verifying that applying the same calculations to the same data yields the same results that have been published?
   - Does "validation consist of more than verification of calculations?

EPA Response

EPA received extensive public comment on the terms replicate and validation from a wide range of commenters. Many commenters noted that replicate is defined in varying ways depending on scientific discipline. Commenters provided several different definitions for these terms. For the term replicate, several commenters pointed to the definition from a 2016 NAS workshop report[1]. EPA will consider the differing definitions for these terms provided in public comment. In responding to these public comments and as part of the final rule, EPA will clarify its interpretation of these terms.

4. Given that there are multiple ways to assess validity of epidemiological studies (some of which do not require public access to all data and methods), what does EPA consider to be adequate validation of a study?
   - Would the answer differ if it referenced toxicological studies or environmental characterization studies?
   - Does "validation" encompass validation of interpretations as well as validation of calculations?
   - Should "validation" include testing whether conclusions are robust to changes in methods of analysis?
   - Should methods to allow for third party validation be developed for studies that cannot share data?

EPA Response

EPA received many public comments relating to the validation of epidemiological studies. Some commenters identified the need for the type of independent validation described in the proposed rule. Other commenters pointed to alternative methods for validation of epidemiological and toxicological studies. EPA is currently evaluating these comments and the approaches they describe. The Agency intends to address these issues in developing the final rule.

Data Handling

5. Who or what body will bear the data handling costs associated with implementing this regulation? This includes:
   - Initial processing and documentation of data prior to public release (including stripping away or suppressing identifying information or confidential business information where necessary);
   - Maintenance and administration of data sets so they are "publicly available" (this includes preparing data, beyond what is currently required by a diversity of funding agencies for posting on a public server, oversight of "limited access" data if appropriate, and updates when they are required).

[1] National Academies of Sciences, Engineering, and Medicine, Principles and obstacles for sharing data from environmental health research: Workshop summary, The National Academies Press (2016), https://www.nap.edu/catalog/21703/principles-and-obstacles-for-sharing-data-from-environmental-health-research.

Response

The answers to these implementation questions are dependent upon the decisions EPA makes, after consideration of the extensive public comment received, on the final requirements at § 30.5 **What requirements apply to EPA's use of dose response data and models underlying pivotal regulatory science** and § 30.6 **What additional requirements pertain to the use of dose response data and models underlying pivotal regulatory science**. After, deciding the scope of these two regulatory provisions, EPA will address these important implementation issues in the final rule.

6. How does EPA propose to handle historical data sets that were created long before any of the new rules were put into place, and for which it is not possible to retrospectively apply the proposed procedures?

Response

EPA received extensive public comment on how the proposed rule would affect the use of historical data. EPA is currently evaluating these public comments and intends to address them in developing the final rule.

7. How long do the data sets need to be maintained and publicly available?

Response

The data sets would be linked to the significant regulatory action and thus to the public docket associated with the rulemaking. Information in the docket associated with a rulemaking would be publicly available as long as the docket for the rulemaking exists. Per EPA's records schedule, these records are permanent and are transferred to the National Archives 15 years after completion of the rulemaking (see https://www.epa.gov/sites/production/files/2019-05/documents/20190308_epa_records_schedules_in_final_status.pdf).

Criteria for Exceptions

8. What specific criteria would constitute grounds for an exception to the stipulation that data upon which regulation is based be made public?

Response

EPA received public comment on the authority EPA proposed for the Administrator to grant exemptions. Commenters also asked EPA to identify criteria for exemptions and provided suggestions for what should be included as criteria. EPA is currently evaluating these public comments and intends to address them in developing the final rule.

Collaboration with other Federal Agencies

9. The proposed rule contemplates that the "EPA should collaborate with other federal agencies to identify strategies to protect confidential and private information in any circumstance in which it is making information publicly available."
   - What kind of collaboration is contemplated and what is the expected process for collaboration with other agencies?
   - Which other agencies should be involved? Many researchers have worked with data from either the Centers for Disease Control or the Census Bureau, so it would seem logical to include those two agencies at least.

Response

EPA has been working primarily with the National Center for Health Statistics at the Centers for Disease Control and Prevention. EPA is developing an Interagency Agreement with the Research Data Center (RDC), CDC/NCHS to conduct a pilot study using their secure data enclave to host EPA datasets in a restricted use environment. EPA has identified candidate datasets for this pilot study.

EPA has also funded Oak Ridge Associated Universities (ORAU) to conduct an independent evaluation of other governmental processes and procedures and the infrastructure needed to optimize public access to research data and protecting PII. ORAU is evaluating processes at NIST, NIH/National Institute of Environmental Health Sciences, Department of Veterns Affairs, Department of Education, Department of Energy, Department of Defense and the United States Agency for International Development.

EPA is a partner with the Office of Science Technology Policy (OSTP) and the federal agencies represented on the Subcommittee on Open Science, a chartered subcommittee of the Committee on Science, National Science and Technology Council. The Subcommittee is examining Agency best practices for increasing public access to scientific publications and research data, including the appropriate protections of sensitive but unclassified information such as PII and CBI.

EPA is reviewing the Foundations for Evidence-Based Policymaking Act of 2018 which includes Section 302. Confidential information protection and statistical efficiency and Section 303 Increasing access to data for evidence. Among other items, this law calls upon the Director of OMB to develop clear and consistent standards, to the extent possible, for the protection and confidentiality of individually identifiable information.

EPA is working with the Office of the Secretary, Department of Health and Human Services. Section 2012 Privacy Protection for Human Research Studies of the 21th Century Cures Act requires the Secretary (HHS), in coordination with other Agencies, to issue a certificate of confidentiality (CoC) to protect the privacy of individuals who are subjects of research funded wholly or in part by the Federal government. A CoC prohibits [with exception] any person to whom a certificate is issued from providing to any other person

not connected with the research any information, biospecimen, or document about an individual.

Consultation with the Science Advisory Board

10. EPA Administrator Wheeler's April 19, 2019 letter to the Chair of the EPA Science Advisory Board states that "The EPA would benefit from a SAB consultation on existing mechanisms for secure access to confidential business information and personally identifiable information as discussed in the proposal."
    - What form of consultation does EPA envision?
    - What does "existing mechanisms" mean to EPA?

    Response

    EPA provides secure access to PII and, as described in the response to question 9, is currently in the process of developing a more formal framework. In the document "Strengthening Transparency in Regulatory Science Proposed Rule: Charge Questions for the SAB," EPA has included links to the laws and regulations that govern its existing mechanisms for secure access to CBI and PII.

Additional Questions on Confidential Financial Business Information and Personally Identifying Information Submitted by SAB Members after the SAB Meeting Held on June 5-6, 2019

Response

Thank you for your questions. We do not have responses to your questions at this time. We will consider them as we develop the final rule.