C R I T
collaboration for research integrity and transparency

Collaboration for Research Integrity and Transparency

A PROGRAM OF YALE LAW SCHOOL, YALE SCHOOL OF MEDICINE AND THE YALE SCHOOL OF PUBLIC HEALTH

August 13, 2018

Acting Administrator Andrew Wheeler
Environmental Protection Agency
1200 Pennsylvania Avenue, NW
Washington, D.C. 20460
submitted electronically

Re: Docket EPA-HQ-OA-2018-0259-0025, Strengthening Transparency in Regulatory Science

Thank you for the opportunity to comment on this proposed Environmental Protection Agency (EPA) regulation.

The Collaboration for Research Integrity and Transparency (CRIT) at Yale is dedicated to promoting health by improving the integrity and transparency of biomedical and clinical research. Although we are strong proponents of responsible data sharing, and seek to make data available to researchers, we request that the EPA withdraw the proposed regulation.

We believe that the proposed regulation will not advance transparency in regulatory science. Instead, we suspect that the proposed regulation may limit the EPA's ability to rely on many well-designed longitudinal and cross-sectional scientific studies. First, there are concerns that the preamble and proposed regulation are not aligned with existing data sharing policies and that the proposed rule directly contradicts established policy and law. Second, although the proposed regulation outlines the importance of "reproducibility", it does not provide a framework for how the reproducibility of an individual study can or should be assessed. Lastly, there are opportunities to promote data sharing without requiring that all data be made publicly available.

Consistent with the larger societal move toward open science, the U.S. federal government, including the EPA, has adopted policies that facilitate data sharing. For instance, the National Institutes of Health (NIH) requires that researchers conducting agency-funded research pre-register clinical trials, pre-specify study endpoints and statistical plans, and report research results. Moreover, all NIH applications for $500,000 or more in direct funding in any given year must contain a data sharing plan, and some

157 CHURCH STREET, 17TH FLOOR
NEW HAVEN, CONNECTICUT 06510-2100
203.436.8420
EMAIL CRIT.ADMIN@YALE.EDU

institutes of the NIH share data with researchers via a secure website.[1] Many federal agencies, including the EPA, require that any published articles based on federally-funded research be made available via the PubMed Central (PMC) website.[2] In addition to government entities, professional societies and journals have also adopted transparency policies. For instance, the International Committee of Medical Journal Editors (ICMJE) has adopted a policy requiring pre-registration of clinical trials, and publication of a data-sharing statement.[3] In 2015, the Institute of Medicine (IOM) published a report recommending that trials be registered prospectively (i.e., at or before the time of first patient enrollment), that summary-level results be shared with the public after trial completion, and that the metadata and patient-level data be responsibly shared with researchers 6 months after scientific publication, or 30 days after regulatory approval.[4] **None of these policies require that all individual-level underlying data and models be made available to the public.**

The EPA's mission necessitates the use of research findings from a variety of sources: research funded or conducted by the EPA or other U.S. government agencies; research funded or conducted by other governments; and research conducted without EPA or other government funding. With regard to studies conducted with human subjects, data may contain personal health information, and particularly with regard to longitudinal studies, may contain such detailed and specific information regarding each individual that the data cannot be properly de-identified for release to the public. In addition, the move toward data-sharing is relatively new, and historical studies with human subjects used written informed consent that did not contemplate sharing of data. Making data from such studies available to the public would violate federal Human Subjects regulations.

The EPA has already embarked on a well-thought-out plan to require and promote research transparency, as described in the EPA's 2016 *Plan to Increase Access to Results of EPA-Funded Scientific Research* ("*Plan*"). Although the preamble to the proposed regulation cites and "applies concepts and lessons learned" from the *Plan*, there are some concerning differences.

---

[1] U.S. Department of Health and Human Services, National Institutes of Health. *NIH Data Sharing Policy and Implementation Guidance* (Updated: March 5, 2003). https://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm. (Last accessed August 13, 2018).
[2] U.S. National Library of Medicine, *Funders and PMC*. https://www.ncbi.nlm.nih.gov/pmc/about/public-access. (Last accessed August 13, 2018).
[3] De Angelis C, Drazen JM, Frizelle FA, et al. Clinical Trial Registration: A Statement from the International Committee of Medical Journal Editors. *N Engl J Med.* 2004;351(12):1250-1251.
[4] Institute of Medicine (IOM). *Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk.* Washington, DC: The National Academies Press; 2015.

First, the *Plan* indicates that it does not apply to "scientific research data collected before its implementation."[5] In contrast, the preamble and the proposed regulation suggest that the regulation may apply retrospectively:

> …EPA seeks comments on how the prospective or retrospective application of the provisions for dose response data and models or pivotal regulatory science could inadvertently introduce bias regarding the timeliness and quality of the scientific information available.[6]

We do not recommend any retroactive application of this regulation. It is well-established that agency rules cannot be applied retroactively unless Congress expressly granted the agency that power.[7] There is no specific grant of power to apply this proposed regulation retroactively. Moreover, the two bills upon which this proposed regulation is modeled, the Honest and Open New EPA Science Treatment (HONEST) Act of 2017 (H.R. 1430 – 115th Congress, 2017-2018), and the Secret Science Reform Act of 2015 (H.R.1030 – 114th Congress 2015-2016), were not enacted.

Second, the preamble and proposed regulation appear to suggest that peer-reviewed studies are only valid if the underlying data and models are publicly available. In contrast to the *Plan*, the proposed regulation proceeds from the erroneous assumption that peer-reviewed research is only valid if it is able to be replicated by others. However, many key environmental studies cannot be replicated, for ethical or practical reasons. For example, the federal Office of Management and Budget has stated:

> OMB urges caution in the treatment of original and supporting data because it may often be impractical or even impermissible or unethical to apply the reproducibility standard to such data. For example, it may not be ethical to repeat a ''negative'' (ineffective) clinical (therapeutic) experiment and it may not be feasible to replicate the radiation exposures studied after the Chernobyl accident.[8]

---

[5] Environmental Protection Agency, *Plan to Increase Access to Results of EPA-Funded Scientific Research*, Version 1.1 (Nov. 29, 2016), p. 5. Available at https://www.epa.gov/sites/production/files/2016-12/documents/epascientificresearchtransperancyplan.pdf. (Last accessed July 31, 2018).
[6] *Strengthening Transparency in Regulatory Science*, 83 Fed. Reg. 18768, 18772 (Apr. 30, 2018).
[7] *Bowen v. Georgetown University Hospital*, 488 U.S. 204, 208 (1988); *see also* Richard J. Pierce, Jr., *Seven Ways to Deossify Agency Rulemaking*, 47 Admin. L. Rev. 59, 76 (1995).
[8] *Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies; Notice; Republication*, 67 FR 8452, 8456 (Feb. 22, 2002).

157 CHURCH STREET, 17TH FLOOR
NEW HAVEN, CONNECTICUT 06510-2100
203.436.8420
EMAIL CRIT.ADMIN@YALE.EDU

Furthermore, according to the *Plan*, not all research can be made fully, thus publicly, available:

> It is important to recognize that some research data cannot be made fully available to the public but instead may need to be made available in more limited ways, e.g., establishing data use agreements with researchers that respect necessary protections. Whether research data are fully available to the public or available to researchers through other means does not affect the validity of the scientific conclusions from peer-reviewed research publications.[9]

The preamble and proposal also acknowledge this concern, and highlight that strategies need to be identified for data that cannot be made publicly available:

> —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. These strategies should be cost-effective and may include: requiring applications for access; restricting access to data for the purposes of replication, validation, and sensitivity evaluation; establishing physical controls on data storage; online training for researchers; and nondisclosure agreements.[10]

Over the past few years, there have been growing efforts to promote the transparency and reproducibility of science.[11] However, the nomenclature surrounding "reproducibility" is complex.[12] Historically, there has been no clear criteria for what constitutes "successful replication and reproduction",[13] which makes certain components of the current proposal concerning. We believe that it is important to ensure that the "reproducible research" movement is not being co-opted for political purposes. Without a clear understanding of

---

[9] *Id.* at pp. 4-5.
[10] *Strengthening Transparency in Regulatory Science*, 83 Fed. Reg. 18768, 18771 (Apr. 30, 2018).
[11] Munafò MR, Nosek BA, Bishop DVM, et al. A manifesto for reproducible science. *Nature Human Behaviour.* 2017;1:0021.
[12] Goodman SN, Fanelli D, Ioannidis JPA. What does research reproducibility mean? *Science Translational Medicine.* 2016;8(341):341ps312.
[13] *Id.*

research reproducibility, different standards (i.e., "case-by-case" decisions) can be applied to discredit individual studies.

According to a recent effort to standardize the language of reproducibility,[14] there are three key areas of reproducibility:

1. Methods reproducibility
    - This is when a study provides enough information about the experimental and/or computational procedures so that future authors can repeat the study using the same data to obtain the same results.
2. Results reproducibility
    - This is when a new study produces corroborating results using experiment methods that are closely matched to a previous study.
3. Inferential reproducibility
    - This is when a new study draws qualitatively similar conclusions from either an independent replication or re-analyses.[15]

These three areas of reproducibility can have different meanings for various areas of scientific research. "Methods reproducibility" may be realistic for laboratory-based studies, where it is possible to implement the exact experimental and computation procedures with the same (or very similar) tools to obtain the same results. However, new clinical or environmental studies may use procedures that are closely matched, but not identical, to previous evaluations. When it comes to observational research, it is well known that some weaknesses are unavoidable. For instance, it is often difficult to eliminate all potential sources of confounding, or to adjust for the same confounders across studies from different time periods. Therefore, it would be unfair to argue that all aspects of a study, including the design and analyses characteristics, need to be reproduced exactly.

The goal of repeating a study should be to increase the total amount of evidence. Furthermore, if a hypothesis being tested has strong evidence from an existing study (i.e., rigorous methods, large sample sizes, transparent reporting), a considerable amount of high quality evidence is necessary to change prior reasoning. Moreover, one small study that does not validate the results of a previous study should not automatically imply that the original study lacks reproducibility.

The emphasis in the proposed regulation on accepting only reproducible studies with publicly available data could eliminate EPA reliance on well-conducted longitudinal studies such as the Six Cities Study, an example of rigorous and robust observational

[14] *Id.*
[15] *Id.*

research, which established the association between fine-particulate air pollution and mortality.[16] After the results were published, an independent group of investigators re-analyzed the results, assessed the robustness of the findings, and confirmed the original results.[17] We believe that similar rigorous re-analyses and replications are possible without mandating public availability of all data.

While data sources that do not pose significant human subject re-identification concerns should be shared publicly using existing data repositories, such as those employed by the National Center for Health Statistics, the Centers for Disease Control and Prevention, and the NIH's NHLBI Biologic Specimen and Data Repositories Information Coordinating Center (BioLINCC), the EPA should focus on creating a data sharing system for long term human subject studies where privacy concerns exist. Under the model that we proposed, only properly deidentified data should be shared, and redaction and de-identification needs to ensure anonymity of research participants according to the Health Insurance Portability and Accountability Act of 1996 (HIPAA). P.L. No. 104-191, 110 Stat. 1938 (1996).

In order to make data available to external researchers for scientific purposes, the EPA should follow previous efforts to facilitate the sharing of data.[18] In particular:

1. An independent intermediary group should be created and should partner with the EPA to oversee data requests. By allowing an independent group to have full jurisdiction to make decisions regarding data access, potential political motivations can be avoided.

2. The independent group should ensure that data are made available for scientific research that is aimed at advancing knowledge. Data requestors should provide basic information about the principal investigators, all study personnel and

---

[16] Dockery DW, Pope CA, Xu X, et al. An Association between Air Pollution and Mortality in Six U.S. Cities. *New England Journal of Medicine.* 1993;329(24):1753-1759. doi: 10.1056/NEJM199312093292401.

[17] Krewski D, Burnett RT, Goldberg MS, et al. Overview of the reanalysis of the Harvard Six Cities Study and American Cancer Society Study of Particulate Air Pollution and Mortality. *J Toxicol Environ Health A.* 2003;66(16-19):1507-51. doi: 10.1080/15287390306424. *See also* Health Effects Institute, *Reanalysis of the Harvard Six Cities Study and the American Cancer Society Study of Particulate Air Pollution and Mortality, Special Report, July 2000.* 2000. https://www.healtheffects.org/publication/reanalysis-harvard-six-cities-study-and-american-cancer-society-study-particulate-air. (Last accessed July 31, 2018).

[18] Krumholz HM, Ross JS. A Model for Dissemination and Independent Analysis of Industry Data. *JAMA.*2011;306(14):1593–1594. doi:10.1001/jama.2011.1459; Yale Open Data Access Project, *Welcome to the YODA Project*. 2018. http://yoda.yale.edu/welcome-yoda-project. (last accessed Aug. 10, 2018).

funders, and submit a detailed proposal outlining specific aims and specifying study methodology. All data requests should undergo external review, to facilitate feedback from independent experts in the field to verify scientific merit. To ensure transparency, the independent group should make all of the data requests publicly available, along with the reasons for granting or denying data requests.

3. All requestors should be required to sign a Data Use Agreement (DUA), which states that access to the data will be used to enhance knowledge and that all findings will be made publicly available through publications and meetings.

4. The independent group should ensure that the scope of the analyses is limited to the specific aims set out in the proposal and that additional objectives are outlined in new submissions.

Responsible data sharing is a complex endeavor and should be done in a manner that is safe and in the best interest of society. Although the IOM report *Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk* focuses on clinical trial transparency, some of the conclusions are applicable to the EPA proposed rule:

> The committee's position is that the benefits of data sharing belong primarily to the public in the form of valid scientific knowledge and improvement of clinical practice and public health. However, these benefits are not necessarily best attained by full open transparency…If full open transparency of clinical trial data carries on balance more risk than benefits, it does not serve the public good.[19]

Furthermore, the language in the EPA proposal is dangerous as it may allow for "case-by-case" determinations that would permit the Administrator to "exempt significant regulatory decisions on a case-by-case basis if he or she determines that compliance is impracticable."[20] Well-established legal precedent frowns upon case-by case determinations where, as here, a general rule would be appropriate.[21] In addition, this provision would allow the EPA to selectively choose studies to rely on and to disregard by granting exemptions to the public availability requirement. Moreover, CRIT believes

[19] Institute of Medicine (IOM). *Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk.* Washington, DC: The National Academies Press; 2015, p. 42.
[20] *Strengthening Transparency in Regulatory Science*, 83 Fed. Reg. 18768, 18772 (Apr. 30, 2018).
[21] *Securities Exchange Commission v. Chenery Corporation*, 332 U.S. 194, 202-03 (1995); *see also* Warren E. Baker, *Policy by Rule or Ad Hoc Approach- Which Should It Be?*, 22 L. & Contemp. Probs. 658, 659 (1957).

that any new regulation involving reproducibility and transparency proposed by the EPA should be applied to research from all sources, including studies conducted by industry.

Lastly, we are also concerned about the lack of scientific justification for the emphasis in the proposed regulation on privileging non-linear models for dose-response over the well-accepted linear models, and the inclusion of the previously non-existent concept of "pivotal regulatory science."[22] The preamble to the proposed regulation makes an unsupported claim that there is "growing empirical evidence of non-linearity in the concentration-response function for specific pollutants and health effects;"[23] and establishes new priorities for EPA funding of research that privilege non-linear models, in direct opposition to the accepted scientific methodology. The proposed regulation states:

> When available, EPA shall give explicit consideration to high quality studies that explore: A broad class of parametric dose-response or concentration-response models; a robust set of potential confounding variables; nonparametric models that incorporate fewer assumptions; various threshold models across the dose or exposure range; and models that investigate factors that might account for spatial heterogeneity.[24]

It is inappropriate to establish major changes in research priorities through insertion of additional language in a proposed regulation ostensibly on another topic, as was done here, rather than to use the normal channels of consultation with stakeholders, advisory boards, and reference to prior commissioned studies from the National Academies of Sciences.[25] These proposed changes are antithetical to the governing law, existing regulations, and well-established agency practice.

---

[22] *Strengthening Transparency in Regulatory Science*, 83 Fed. Reg. 18768, 18770 (Apr. 30, 2018).
[23] *Id*. at 18773.
[24] *Id*. at 18774.
[25] National Research Council. 2009. *Science and Decisions: Advancing Risk Assessment*. Washington, DC: The National Academies Press. https://doi.org/10.17226/12209.

Collaboration for Research Integrity and Transparency

A PROGRAM OF YALE LAW SCHOOL, YALE SCHOOL OF MEDICINE AND THE YALE SCHOOL OF PUBLIC HEALTH

Thank you for the opportunity to comment. We urge you to withdraw the proposed regulations.

Sincerely,

Margaret E. McCarthy
Executive Director, CRIT

Joseph S. Ross
Associate Professor of Medicine, Yale School of Medicine

Joshua D. Wallach
Research Fellow, CRIT
Assistant Professor of Epidemiology, Department of Environmental Health Sciences, Yale School of Public Health (appointment effective September 1, 2018)

157 CHURCH STREET, 17TH FLOOR
NEW HAVEN, CONNECTICUT 06510-2100
203.436.8420
EMAIL CRIT.ADMIN@YALE.EDU