State of Louisiana
DEPARTMENT OF JUSTICE
OFFICE OF THE ATTORNEY GENERAL
P.O. BOX 94005
BATON ROUGE
70804-9005

Jeff Landry
Attorney General

August 15, 2018

Via Electronic Submission and Federal Express

Andrew Wheeler, Acting Administrator
Environmental Protection Agency
EPA Docket Center
WJC West Building, Room 3334
1301 Constitution Avenue, NW
Washington, DC 20004

Re: **Comments of the Attorneys General of Louisiana, Alabama, Arkansas, Indiana, Kansas, Nebraska, Oklahoma, South Carolina, Texas, Utah, and Wisconsin on EPA's Proposal to Strengthen Transparency in Regulatory Science; 83 Fed. Reg. 18,768 (April 30, 2018); Docket ID No. EPA-HQ-OA-2018-0259**

Dear Administrator Wheeler:

The undersigned Attorneys General, as the chief legal officers of our States, write to express support for the Environmental Protection Agency's ("EPA") April 30, 2018, proposal to adopt a regulation to provide much-needed transparency regarding the scientific research and data on which the agency bases its regulatory actions ("the Proposed Rule").*

GENERAL COMMENT

The citizens of our States and the industry stakeholders so crucial to our States' economies bear the cost of regulatory compliance. Accordingly, the science on which those regulations are based should be high quality and available to the public. Public disclosures should be sufficient to permit independent validation and analysis of the data, methodology, computational code, and conclusions. Transparency of this sort not only comports most fully with the public participation in rulemaking contemplated by the Administrative Procedure Act, but is a bedrock principle for establishing public confidence in actions taken by the EPA.

The Proposed Rule addresses the troubling erosion of credibility of published scientific literature. EPA is obliged to rely upon the best available science in its regulatory actions. *See, e.g.*, Exec. Order No. 13,783, 82 Fed. Reg. 16,093 (Mar. 31, 2017); Exec. Order No. 13,563, 76 Fed. Reg. 3,821 (Jan. 21, 2011). However, absent open sharing of underlying data, methodology, and computational code, the credibility of the science is open to question. Moreover, reproducibility can be rendered impossible, further crippling the credibility of the research.

As reported in *Science*, reproducibility in scientific work is important to providing confidence in the conclusions reached, but a "troubling proportion" of peer-reviewed studies are not reproducible.[1] *Science* thus revisited the criteria for publication with the aim of increasing confidence in the studies it publishes. Similarly, a recent article in *Nature* identified lack of replication and lack of data sharing as "threats" to reproducible science.[2] Open and transparent sharing of the methodology applied, the data acquired, and the process of methodology implementation, data analysis and outcome interpretation was identified as the remedy. The type of open science called for in these publications—laying bare both content and process—is precisely what is achieved through the Proposed Rule.

<u>SPECIFIC COMMENT 1</u>: There is strong evidence that publication in a peer reviewed journal is not sufficient to ensure a study reflects valid, reproducible science.

Concern with the quality of published science goes back to at least 2005, when a study of 45 highly-cited articles in *New England Journal of Medicine*, *JAMA*, and *Lancet* concluded that at least 7 articles (16%) were contradicted by subsequent research and another 7 articles (16%) claimed stronger effects than were supported by subsequent research.[3] A widely-read essay published later that year suggested that most published research findings are false.[4] Consistent with that suggestion, Bayer Healthcare disclosed in 2011 that 43 (65%) of the company's attempts to reproduce 67 published studies resulted in inconsistent data.[5] In 2012, Amgen, Inc., similarly disclosed that it was only able to reproduce 6 (11%) of 53 studies that the company attempted to confirm.[6]

Consistent with the Bayer and Amgen disclosures, recent estimates for irreproducibility in preclinical and biomedical research range as high as 90% of that research, even for articles published in high-quality journals.[7, 8, 9] The problem does not appear to be limited to particular fields. For example, a recent survey of 804 ecologists and evolutionary biologists found that questionable research practices were widespread, with 64% of surveyed researchers reporting they had at least once failed to report results because they were not statistically significant ("cherry picking"); 42% reporting they had collected more data after inspecting whether results were statistically significant (a form of "p-hacking"); and 51% acknowledging they had reported an unexpected finding as though it was hypothesized from the start.[10] In short, a large fraction of the surveyed researchers admitted to manipulating their results. Not surprisingly, a 2016 survey by the journal *Nature* found that lack of reproducibility is a widespread concern among scientists.[11, 12]

<u>SPECIFIC COMMENT 2</u>: There is strong evidence of widespread, outcome-altering errors in the computational code underlying many scientific studies.

---

* Each publication cited herein is attached as an exhibit to the hardcopy of this comment.

1 M. McNutt, *Editorial: Reproducibility*, 343 SCIENCE 229 (2014).

2 M. R. Munafo et al., *A Manifesto for Reproducible Science*, 1 NATURE HUM. BEHAVIOR 21 (2017).

3 J. P. A. Ioannidis, *Contradicted and Initially Stronger Effects in Highly Cited Clinical Research*, 294(2) JAMA 218 (2005).

4 J. P. A. Ioannidis, *Why Most Published Research Findings Are False*, 2(8) PLOS MED. e124 (2005).

5 F. Prinz et al., *Correspondence: Believe It or Not: How Much Can We Rely on Published Data on Potential Drug Targets*, 10 NATURE REV. DRUG DISCOV. 712 (2011).

6 C. G. Begley & L. M. Ellis, *Comment: Drug Development: Raise Standards for Preclinical Cancer Research*, 483 NATURE 531 (2012).

7 C. G. Begley & J. P.A. Ioannidis, *Reproducibility in Science*, 116 CIRC. RES. 126 (2015).

8 L. P. Freedman et al., *The Economics of Reproducibility in Preclinical Research*, 13(6) PLOS BIOLOGY e1002165 (2015).

9 V. E. Johnson, *Revised Standards for Statistical Evidence*, 110(48) PNAS 19313 (2013).

10 H. Fraser et al., *Questionable Research Practices in Ecology and Evolution*, Open Science Framework (Preprint March 21, 2018).

11 M. Baker, *Is There a Reproducibility Crisis?*, 533 NATURE 452 (2016).

12 Editorial, *Reality Check on Reproducibility*, 533 NATURE 437 (2016).

ED_002389_00024316-00002

Outcome-altering errors in computational code have been suggested as a significant contributor to the replication crisis.[13] High profile retractions, technical comments, and corrections because of coding errors include papers in prominent journals such as *Science*, *PNAS*, the *Journal of Molecular Biology*, *Ecology Letters*, *Journal of Mammalogy*, *Journal of the American College of Cardiology*, *Hypertension*, and *American Economic Review*.[14] Such outcome-altering errors can arise from the simplest mistakes. For example, five retractions were based on a flipped minus sign.[15] Perhaps most famously, a 2010 study by Harvard University economists Carmen Reinhart and Kenneth Rogoff was undermined by errors in the authors' spreadsheet.[16, 17] The Reinhart-Rogoff study had significant public policy consequences, yet the errors went undetected until a graduate student sought and reviewed the authors' working spreadsheet.[18, 19]

<u>SPECIFIC COMMENT 3</u>: The Proposed Rule is consistent with the publication requirements of major scientific journals, which were adopted specifically to address the reproducibility crisis.

In its commentary accompanying the Proposed Rule, the EPA noted that its policies and the policies of open-science advocates "are informed by the policies recently adopted by some major scientific journals, spurred in some part by the 'replication crisis.'" 83 Fed. Reg. at 18770. The EPA cited the policies of *PNAS, PLOS One, Science,* and *Nature* as examples. *Id.* at 18770 & n.11. Those policies were specifically intended to address the inability of researchers to properly assess results in the absence of underlying data and methodology. The journal *Nature* explained:

> Over the past year, *Nature* has published a string of articles that highlight failures in the reliability and reproducibility of published research . . . . The problems arise in laboratories, but journals such as this one compound them when they fail to exert sufficient scrutiny over the results that they publish, and ***when they do not publish enough information for other researchers to assess results properly***.
>
> From next month, Nature and the Nature research journals will introduce editorial measures to address the problem by improving the consistency and quality of reporting in life-sciences articles. ***To ease the interpretation and improve the reliability of published results we will more systematically ensure that key methodological details are reported***, and we will give more space to methods sections. We will examine statistics more closely and ***encourage authors to be transparent, for example by including their raw data***.[20]

More recently, the editors of *Nature* recognized that—despite the adoption of transparency policies—resolution of the reproducibility crisis "is far from complete."[21] They accordingly encouraged "funders, researchers and journals to keep up the pressure towards the openness of complete data sets and any source code required to use them."[22] Although public disclosure of data, methodology, and computational code

---

13 D. A. W. Soergel, *Rampant Software Errors May Undermine Scientific Results*, 3 F1000RESEARCH 303 (2015).

14 G. Wilson et al., *Best Practices for Scientific Computing*, 12(1) PLOS BIOLOGY e1001745 (2014).

15 Z. Merali, *Why Scientific Programming Does Not Compute*, 467 NATURE 775 (2010).

16 C. M. Reinhart & K. Rogoff, *Growth in a Time of Debt*, American Economic Review: Papers and Proceedings 100 (2010) at 573.

17 T. Herndon et al., *Does High Public Debt Consistently Stifle Economic Growth? A Critique of Reinhart and Rogoff*, Political Economy Working Paper Series No. 322 (2013).

18 R. Alexander, *Reinhart, Rogoff … and Herndon: The Student Who Caught Out the Profs*, BBC News (Apr. 20, 2013).

19 P. Coy, *FAQ: Reinhart, Rogoff, and the Excel Error that Changed History*, Bloomberg News (Apr. 18, 2013).

20 Announcement, *Reducing Our Irreproducibility*, 496 NATURE 398 (2013).

21 Announcement, *Transparency Upgrade for Nature Journals*, 543 NATURE 288 (2017).

22 Editorial, *Not-So-Open Data*, 546 NATURE 327 (2017).

ED_002389_00024316-00003

may not be *sufficient* to ensure that scientific claims are valid, public disclosure is absolutely *necessary* for confidence in the validity those claims.[23, 24, 25, 26]

SPECIFIC COMMENT 4: Pre-publication peer review is not an adequate substitute for public availability of data, methodology, and computational code.

In response to the Proposed Rule, some commenters have suggested the Proposed Rule is an attempt to remove valid and relevant scientific evidence from the rule-making process. But that begs the question: in the light of overwhelming evidence that a significant portion of studies published in prominent peer reviewed journals are not reproducible, it is dubious to claim that any particular research is valid and relevant unless—at a minimum—other researchers and the public have access to the underlying data, methodology, and computational code.

More to the point, empirical evidence indicates that peer reviewers routinely fail to identify even major errors.[27] And peer review in less prominent journals may often occur in name only.[28] In short, the best available metascience—science about science—indicates that pre-publication peer review is not adequate to ensure the validity of published scientific claims.

SPECIFIC COMMENT 5: Promises to share data with qualified researchers are not an adequate substitute for public disclosure of data, methodology, and computational code.

Promises to provide other researchers with access to data, methodology, or computational code are not an adequate substitute for public availability. Such promises often go unfulfilled.[29] In an attempt to quantify the problem, researchers from Memorial Sloan-Kettering Cancer Center sought data from the corresponding authors of ten articles that were subject to a data sharing requirement.[30] Two of the authors could not be contacted. Of the remaining eight authors, four replied that sharing their data was not possible, three did not respond, and one asked for further details regarding the request. Despite a follow-up contact reminding the authors of their data sharing obligation, the Sloan-Kettering researchers received only a single data set of the ten originally sought.

Significantly, many uncredentialed members of the public are capable of reviewing and identifying errors in data, methodology, and computational code. The Reinhart-Rogoff errors, *supra*, were identified by a graduate student. Other serious, basic errors in high-profile studies have been identified by undergraduates.[31] If undoubtedly qualified Sloan-Kettering researchers were generally unable to obtain data, there is little reason to believe other researchers—formally qualified or otherwise—will be able to do so. When the EPA decides to regulate, the public is entitled to more than empty promises that data supports that decision.

---

23 G. Santori, *Correspondence: Journals Should Drive Data Reproducibility*, 535 NATURE 355 (2017).

24 S. M. Easterbrook, *Open Code for Open Science?*, 7 NATURE GEOSCI. 779 (2014).

25 G. J. Lithgow et al., *Comment: A Long Journey to Reproducible Results*, 548 NATURE 387 (2017).

26 J. S. Mogil & M. R. Macleod, *Comment: No Publication Without Replication*, 542 NATURE 409 (2017)

27 S. Schroter et al., *What Errors Do Peer Reviewers Detect, and Does Training Improve Their Ability to Detect Them?*, 101 J. R. SOC. MED. 507 (2008).

28 J. Bohannon, *Who's Afraid of Peer Review?*, 342 SCIENCE 60 (2013).

29 D. G. Roche, *Evaluating* Science's *Open Data Policy*, 357 SCIENCE 654 (2017).

30 C. J. Savage & A. J. Vickers, *Empirical Study of Data Sharing by Authors Publishing in PLoS Journals*, 4(9) PLOS ONE e7078 (2009).

31 R. Nuzzo, *Fooling Ourselves*, 526 NATURE 182 (2015).

ED_002389_00024316-00004

SPECIFIC COMMENT 6: The Proposed Rule is supported by and arguably required by the Administrative Procedure Act.

The Administrative Procedure Act ("APA"), 5 U.S.C. § 500 et seq., prohibits agency action, findings, and conclusions that are arbitrary, capricious, an abuse of discretion, or otherwise not in accordance with law. *See* 5 U.S.C. § 706(1). In view of the overwhelming evidence that (i) a very high percentage of published studies are not reproducible, (ii) pre-publication peer review is not sufficient to ensure that a study lacks major errors, and (iii) pre-publication peer review is not sufficient to ensure that the computational code underlying a study lacks major errors, we respectfully submit that the Proposed Rule is *necessary* and arguably *required* by the APA. Indeed, an agency action based on a study without fully disclosed data, methodology, and computational code is arguably so lacking in reasoned foundation as to be arbitrary and capricious.

SPECIFIC COMMENT 7: The Proposed Rule should apply to all agency actions that are based on science.

The replication crisis and public confidence rationales that underlie the Proposed Rule suggest that data, methodology, and computational code should be disclosed for all science relied upon to support agency actions. As a first step, the phrase "dose response data and models" should be replaced with "data and models, including dose response data and models" throughout the Proposed Rule.

SPECIFIC COMMENT 8: EPA should identify all studies (or other regulatory science) it will rely upon at the time it proposes any regulation.

In order to meaningfully comment on proposed agency action, the public should be informed of the detailed scientific basis for that action when the action is proposed. Accordingly, Proposed Section 30.4 should be amended to provide:

> EPA shall clearly identify all studies (or other regulatory science) relied upon to support ~~when it takes~~ any ~~final~~ agency action. When EPA proposes any agency action, EPA ~~should~~ shall make all such studies (or other regulatory science) available to the public to the extent practicable.

SPECIFIC COMMENT 9: The proposed rule should not include exceptions that are not required by statute or the Executive Branch's inherent powers.

In view of the widespread problems with replicating even peer-reviewed studies published in prominent journals, the EPA should minimize exceptions to the disclosure requirements set forth in the Proposed Rule. Courts have frequently rejected attempts to withhold information based on claims to privacy, etc., unless protections are specifically provided by law. *See, e.g.*, *Johnson v. Dovey*, 2011 U.S. Dist. LEXIS 128577, at *6-7 (E.D. Cal. Nov. 7, 2011) (rejecting vague assertion of privacy rights: "CDCR's objections are not specific in any way, and are thus insufficient to assert a privilege."); *see also Donovan v. Nat'l Bank of Alaska*, 696 F.2d 678 (9th Cir. 1983) ("The bank cannot refuse to comply with the subpoena as a whole on the basis of its vague allegations that it might be required . . . to produce records in violation of the Financial Privacy Act."). EPA should similarly reject such claims to "privacy, confidentiality, [and] confidential business information" vis-à-vis data and models, unless those claims are clearly supported by law or the Executive Branch's inherent powers. Accordingly:

> (a) EPA should amend the second sentence of Section 30.5 to provide that
>
> Where the Agency is making data or models publicly available, it shall do so in a fashion that is consistent with law, ~~protects privacy, confidentiality, confidential business information,~~ and is sensitive to national and homeland security.

ED_002389_00024316-00005

(b) EPA should amend the penultimate sentence of Section 30.5 to provide that

> The agency shall make all reasonable efforts to explore methodologies, technologies, and institutional arrangements for making such data available before it concludes that doing so in a manner consistent with law and ~~protection of privacy, confidentiality,~~ national and homeland security is not possible.

Further, in view of the evidence that even studies published in prominent journals are often false or overstate results, the EPA should carefully consider the weight given to studies, data, and models that are not disclosed for review by the public and other researchers, even where non-disclosure is required by law.

SPECIFIC COMMENT 10: In view of the replication crisis, the EPA should re-evaluate existing regulations using the standards set forth in the Proposed Rule.

In the commentary accompanying the Proposed Rule, the EPA states that the Proposed Rule "is intended to apply prospectively . . . ." 83 Fed. Reg. at 18771. In view of the overwhelming evidence that (i) a very high percentage of published studies are not reproducible, (ii) peer review is not sufficient to ensure that a study lacks major errors, and (iii) peer review is not sufficient to ensure that the computational code underlying many studies lacks major errors, we submit that the EPA should establish a plan to review the scientific basis for existing regulations under the standards of the Proposed Rule.

SPECIFIC COMMENT 11: The EPA should not delay implementation of this rule unless required to do so by statute.

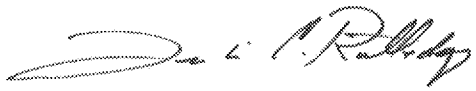
Certain commenters claim the EPA has not complied with specific statutory obligations. We take no position on the EPA's statutory obligations, but we encourage the EPA to carefully comply with applicable procedural requirements so as to minimize the risk of dilatory litigation.

* * * * *

The Proposed Rule makes great inroads towards the EPA's ensuring public confidence in the regulations it enacts. We support making critical scientific data available in a way that is both timely and transparent to stakeholders and the public at large. To that end, we fully support the Proposed Rule. We appreciate your consideration of this comment and look forward to the prompt enactment of this important regulation.

Sincerely,

Jeff Landry
Louisiana Attorney General



Leslie Rutledge
Arkansas Attorney General

Steve Marshall
Alabama Attorney General

Curtis T. Hill, Jr.
Indiana Attorney General

ED_002389_00024316-00006

Derek Schmidt
Kansas Attorney General

Alan Wilson
South Carolina Attorney General

Douglas J. Peterson
Nebraska Attorney General

Ken Paxton
Texas Attorney General

Mike Hunter
Oklahoma Attorney General

Sean Reyes
Utah Attorney General

Brad D. Schimel
Wisconsin Attorney General

cc: Brittany Bolen, Office of Policy (via email)
Tom Sinks, Office of the Science Advisor (via email)

ED_002389_00024316-00007