The American Statistical Association Urges the EPA Proposed Rule, *Strengthening Transparency in Regulatory Science*, Be Withdrawn or Rejected

August 15, 2018

The American Statistical Association strongly supports reliable research informing policymaking. Transparency of the scientific procedures and the underlying data is an important component of the policymaking process. Of course, transparency requires appropriate protections of privacy and intellectual property as well as scientifically sound peer review procedures. Just as importantly all relevant, reliable science should be considered in evidence-based policymaking. Regrettably, because less science will ultimately be considered in the proposed process and because of its introduction of potential bias, we view this proposed rule (Docket ID:EPA-HQ-OA-2018-0259) as a step backwards for evidence-based policymaking. We urge the proposed rule not be adopted.

1. **Proposed Rule Hampers the Use of Evidence.** The proposed rule will hamper the use of relevant scientific evidence in policymaking and regulation by restricting the evidence that could be considered. Important, useful, and valid studies that have not, or cannot, make available the data on which they are based still have scientific value.

   We recognize and appreciate the efforts in the proposed rule to potentially make data available while protecting confidentiality through such means as “controlled access in federal research data centers” and EPA “collaborat[ion] with other federal agencies to identify strategies to protect confidential and private information.” Such approaches to providing restricted access should be pursued across the federal government if cost-effective and not overly burdensome but such efforts in this proposal do not justify its approval for the many reasons made here.

Despite the efforts in the rule to make data available, some scientific studies will still be excluded—whether by cost of making the data available, the terms in data confidentiality policies used to collect the information, or other limitations—thereby undermining EPA work. The following points underscore further how this rule hampers the use of evidence for rulemaking.

a. **Private sector studies may not be considered**. Research studies designed to provide more knowledge and better understanding—versus studies for supporting regulations—may not be considered because they do not meet the proposed EPA criteria. Yet, these studies could be well designed, carefully conducted, and provide valuable information for the EPA to consider. One should also consider that researchers often publish their work on one topic and move on to other topics. If such research is ultimately relevant to EPA's rulemaking or regulations, what efforts would the EPA undertake to work with the researchers to make the data available? How would the researchers be incentivized to undertake the significant amount of work this rule would likely require to have their research used by the EPA? How would the EPA approach situations where HIPAA or FERPA informed consent requirements apply? Would the researchers' time be compensated?

b. **Concerns about protecting privacy may deter researchers from providing studies to the EPA.** Researchers have a responsibility to assure the confidentiality of data they collect from human subjects. Combining or merging datasets can increase risk to confidentiality by increasing the likelihood of identifying individuals. Researchers may therefore be reluctant to submit their research to the EPA because producing the required data may violate informed-consent provisions.

c. **Increasingly prevalent meta-analyses may be excluded**. Meta-analyses are studies that use scientific and statistical methods to combine results from multiple studies. Published meta-analyses have grown immensely over the years (to more than 250,000)[1] to become an important and standard research method in epidemiology, medicine, and public health. If any of the component studies 9fail meet the proposed EPA data requirements, the insights from the meta-analysis is likely to be lost.

[1] J Gurevitch, J Koricheva, S Nakagawa, G Stewart. Nature 2018; 555:175-182.

2. **Discriminating what science to use for drafting regulation risks introducing potential bias in the rulemaking process.** Bias is likely to be introduced through the process of identifying what science can and cannot be used because there is not a bright line for what studies this proposed rule would include or exclude. Specifically, the requirement to make a decision for what constitutes "high quality studies" is problematic because the tem is not defined in the proposed rule nor are operational criteria for its definition specified. Recent work in systematic reviewing of research in many areas of science has demonstrated considerable variation in judgments about "quality" by competent scientists unless the criteria for quality are specified precisely. This variation in judgement risks arbitrary and capricious decisions as to what evidence to include or exclude. Several good resources for avoiding bias and gaming of the regulatory process in specifying the standard for "high quality" are available.[2] The exclusion of otherwise valid and reliable research from consideration would also often result in regulatory analyses that suffer from biased selection and inadequate statistical power.

   The proposed rule's statement, "EPA shall conduct independent peer review on all pivotal regulatory science used to justify regulatory decisions," also risks the introduction of potential bias through the selection of the peer reviewers. The reviewers should be individuals who have the relevant expertise and are impartial with full disclosure of any conflicts of interest. Having had industry, government, or other funding for work on the topic at hand should not by itself be disqualifying to serving in peer review of scientific research.

3. **Providing access while protecting confidentiality is challenging and complicated.** Important advances have been made in the last decade in enabling protection of the confidentiality of data protections. However, advances in information technology, proliferation of linkable auxiliary data, and other factors make it increasingly difficult and sometimes impossible to ensure that confidential data are protected. A "one-size-fits-all" approach for access to regulatory data does not work. Rather a framework for providing varied levels of access to regulatory research data are essential. As has been learned from confidential data centers providing access to data in a secure environment for some statistical agencies, such access is administratively complex, takes considerable time to implement, and is costly to both agencies and the scientists who use it. We support the use of such access to data across the federal government but they should be implemented and developed independent of this proposed rule.

[2] http://us.cochrane.org/resources; https://campbellcollaboration.org; https://ies.ed.gov/ncee/wwc/WhoWeAre; https://www.russellsage.org/publications/handbook-research-synthesis-and-meta-analysis-second-edition.

4. **New studies may be discouraged for privacy and/or cost considerations:**
    a. The need to provide data will make it harder to obtain consent from subjects worried about personal identification. Recent methods have proven very powerful in overcoming even the most sophisticated techniques to protect privacy. It is possible that even very expensive measures such as use of the Federal Research Data Centers may not be adequate to maintain adequate privacy.

    b. With the stringent data requirements, and the increased concern about privacy, it is likely that the costs of compliant research would go up. This may discourage researchers from undertaking future studies.

5. **Costs and Benefits are not fully considered.** Cost is cited over a dozen times in the proposed rule, yet no explicit cost benefit analysis is provided to help understand how the benefits compare to the costs. A proposed rule of this scope should be informed by a cost benefit analysis that helps demonstrate that the benefits of its adoption outweigh the costs imposed.

6. **Rescinding regulations may have very detrimental health and environment effects**. If some rules and regulations are rescinded due to failure to comply with the data requirements, health and environmental damage may occur. These rules and regulations were subject to public comment and were promulgated to promote human health and environmental protection. For example, according to the EPA, the Clean Air Act will prevent as many as 230,000 early deaths in 2020.[3] Rescinding such rules would result in loss of such health and environment benefits, which should be fully considered in a cost-benefit analysis. The failure to adopt new rules because of the faults in this proposed rule poses a similar risk to health and the environment.

We recognize data sharing is important and the United States should continue to pursue broader data sharing but it should not be at the cost of restricting the science that EPA can use to fulfill its mission to protect the environment and the health of the U.S. population.

[3] https://www.epa.gov/clean-air-act-overview/benefits-and-costs-clean-air-act-1990-2020-second-prospective-study