Administrator Scott Pruitt  
Environmental Protection Agency  
1200 Pennsylvania Avenue, NW  
Washington, DC 20460


Dear Administrator Pruitt:

Regulatory actions taken by the Environmental Protection Agency (EPA) should be informed by the best available science. By restricting valid scientific information, however, the Agency’s proposed rulemaking, “Strengthening Transparency in Regulatory Science” undermines that notion and would adversely affect the Agency’s goal of protecting human health and the environment.

The proposed rule requires that scientific information used in regulatory decisions (specifically “dose response data” and models underlying “pivotal regulatory science”) must be publicly available for independent validation. This will limit the impact of many environmental peer-reviewed studies that have relied on personally identifiable data to assess health outcomes. Such examples include the Harvard Six Cities study and the American Cancer Society Cancer Prevention Study, both of which showed that exposure to fine particulate matter is associated with mortality. While results from both of these studies have been subjected to independent reanalysis and sensitivity analyses [1, 2], public access of these data will be difficult given that they were collected decades ago with assurances of confidentiality for research participants. A more recent example includes the Multi-Ethnic Study of Atherosclerosis and Air Pollution (MESA-Air), a 10-year prospective cohort study of more than 6,000 participants. Using these data, MESA-Air researchers have produced over 100 scientific publications examining the cardiovascular impacts of long-term exposure to air pollution [3], all of which could be excluded from consideration in the regulatory process. While the proposed rule does not state who would be responsible for the redaction of identifiable data, the costs associated with the redaction for 6,000 participants, each with repeated measurements over years of follow-up, would be a burden in both time and cost. Additionally, the proposed rule does not state the Agency’s plans for disseminating publicly available data that are under the jurisdiction of other federal agencies, such as Medicaid and Medicare. A recent study that might be excluded from consideration showed that long-term exposure to particulate matter and ozone at levels below the annual standards was associated with an increase in mortality in the Medicare population. [4]

Restricting studies that have relied on personally identifiable data would weaken the evidence used to draw causal inferences, which are necessary for creating regulatory actions. As a graduate student in epidemiology, I am taught that determination of a cause-effect relationship is based on a critical evaluation of all available studies. With each study, one must assess its validity (e.g., potential for selection bias, uncontrolled confounding, measurement error, and
sample size) and its consistency (or inconsistency) will all available science. In developing
“regulations for which the public is likely to bear the cost of compliance”, why should decisions
be based on anything but a complete assessment of all available science? Only from a complete
assessment can the “best studies” be determined.

The proposed rule states that “tools and methods to de-identify private information” are
available; however, it does not address issues of analysis due to information loss. After “simple
data masking, coding, and de-identification”, it is not guarantee that statistical and inferential
reproducibility will be achieved. Based on my experience with de-identified databases (e.g., the
Pediatric Health Information System, PEDSnet, and public data sources), identifiable
information are often aggregated, recoded, or set to missing in order to protect individuals’
identities. In environmental epidemiology, data on individuals’ residencies, medical histories,
demographics (e.g., race/ethnicity, age, gender, smoking status, and measurements of
socioeconomic status) are collected in order to obtain valid estimates of the relationship
between environmental exposures and health outcomes. In order to publicly release these
data, researchers will have to achieve de-identification in accordance with the Health Insurance
Portability and Accountability Act (HIPAA) Privacy rule by removing information on all
residencies, except for the initial three digits of the ZIP code. [5] Additionally, geographical units
containing 20,000 or fewer people will have to be aggregated or changed to an unidentifiable
code. As such, replication and validation of these de-identified data may produce less precise
estimates of individual exposures, and the impreciseness of exposure ascertainment will be
greater in rural communities. Additionally, information on confounders, effect modifiers, and
rare health outcomes may be missing due their identifiability. In conclusion, de-identification in
compliance with HIPAA leads to a substantial loss of information. Even if the original analysis
was unbiased, independent validation may not lead to similar conclusions.

While the proposed rule states that the EPA administrator may grant exemptions on a "case-by-
case basis if he or she determines that compliance is impracticable”, the proposed rule does not
provide information on how exemptions will be granted or if there will be a formal process for
exemptions. Leaving such decisions to the EPA administrator without more formal guidelines
will weaken the transparency of the Agency’s regulatory science as exemptions will be prone to
conflicts of interest and political influence.

Instead of enacting the proposed rulemaking, there are other ways the Agency could
strengthen the transparency of regulatory science without restricting scientific information. The
scientific community has been actively working to increase the transparency and reproducibility
of research, and the Agency could promote the resources that are currently being implemented
within the community. One example includes promotion of the STROBE statement, aimed at
STrengthening the Reporting of OBservational studies in Epidemiology by providing authors a
checklist of requirements necessary for complete and adequate reporting of research. [6]
Additionally, the Agency could encourage the registration of research protocols and create a
public repository where environmental scientists from industry and academia can declare
sources of funding. These tasks would help improve the quality of the science used in the
regulatory process and allow the Agency to make strong evidence-based regulations that would protect public health and the environment.

Thank you for your time and consideration.

Sincerely,

Vi Le
University of Washington

References