WRITTEN STATEMENT FROM ROY N. GAMSE TO THE US EPA SAB
REGARDING "STRENGTHENING TRANSPARENCY IN REGULATORY SCIENCE"
AUGUST 27, 2019

Thank you for the opportunity to provide comments for your consideration as you respond to the charge from EPA on the proposed regulation on Strengthening Transparency in Regulatory Science.

I am Roy Gamse. I worked for EPA for 10 years during the Nixon, Ford, Carter, and Reagan Administrations. I was Deputy Assistant Administrator, Deputy Associate Administrator, and Acting Assistant Administrator from 1977 to 1981, which included responsibility for the regulation development process and for economic and statistical analysis at EPA.

SUMMARY

1. **This proposal** has been justified on the false premise that identities of individuals in health research studies can be reliably masked to protect their privacy as required by law. They cannot be. Hence, this proposal **would result in many relevant studies being disallowed, including many that provide the basis for current EPA standards** that must be reviewed under applicable legislation.
2. **If this is such a good idea, why is it proposed just for EPA** rather than as legislation or regulation applying to all health-regulating agencies? Imagine that research used for FDA drug approvals had to pass the same requirements. FDA drug approvals would grind to a halt, just as would EPA health-protective regulations.
3. **The single most relevant document for the SAB to consider in responding to the EPA charge is the Comments of the International Society for Environmental Epidemiology on EPA's proposed rule on Strengthening Transparency in Regulatory Science.** Since the ISEE submitted it previously to EPA, but the SAB members have not seen it, I am attaching it to this submission. **SAB members <u>must</u> read this document. The ISEE recommends that EPA withdraw the proposal** in part because the "masking" of personal identities cannot be reliably done and still allow reanalysis of the research.

DISCUSSION

Your review of this proposed regulation is critical because, if adopted, it runs the risk of sidetracking EPA's development of health-protective regulations. If this self-regulation were in place throughout EPA's history in combination with today's advanced techniques for "big data" analysis, many of EPA's vital current health protective regulations could not have been adopted. If it were required for all EPA regulations in the future, many regulations needed to protect health could not be promulgated, which would be a tragedy, negatively affecting the health and longevity of many in the US. Most critical, health-based regulations which are required by legislation to undergo periodic review (e.g., Clean Air Act ambient air quality regulations) could be decimated as the studies used as a basis for setting them in the past would be rejected under this rule.

As the record shows and will show, privacy considerations would disallow the use of much data that provide the basis for EPA's health-protective regulations. That is especially the case for epidemiological studies which track or measure the health of groups of individuals with different levels of exposure to pollutants, such as the studies used as the basis for EPA's ambient air quality standards and hazardous pollution standards.

Some people will tell you that masking personal information can be done to hide the identity of individuals. That is nice in theory, and it may have been more practical in the 1970s and 80s when EPA regulation of health impacting pollution was in its early stages. But you should learn, if you have not already, that recent techniques of data manipulation in the era of "big data" analysis will now allow specific identification of many individuals who are participants in studies that provide the basis of EPA regulation.

I direct your attention first to Dr. Joel Schwartz's article in the New England Journal of Medicine entitled " 'Transparency' as Mask? The EPA's Proposed Rule on Scientific Data," published August 29, 2018 [https://www.nejm.org/doi/full/10.1056/NEJMp1807751].

Dr. Schwartz, Professor at Environmental Epidemiology at the Harvard School of Public Health and a MacArthur Foundation "genius award" winner, spells out with examples how individuals can be identified even after deleting their names and other identifying information, unless so much information is deleted that the research results cannot be verified.

Besides his technical explanations of how analysis can reveal identities of people who have a right to anonymity, Dr. Schwartz cites a few examples of how current analytic techniques can defeat efforts to hide individual identities:

- How addresses of victims of Hurricane Katrina were identified without published information on names or addresses. [Curtis, Mills, Leitner, Spatial confidentiality and GIS: re-engineering mortality locations from published maps about Hurricane Katrina, International Journal of Health Geographics, October 10, 2006; [https://ij-healthgeographics.biomedcentral.com/articles/10.1186/1476-072X-5-44]
- A National Academy of Sciences experiment that showed confidentiality could not be preserved even after all participant information not required to allow scientists to replicate a study's findings were deleted. [National Research Council, Access to research data in the 21st century: an ongoing dialogue among interested parties, National Academy Press, 2002; https://www.nap.edu/catalog/10302/access-to-research-data-in-the-21st-century-an-ongoing]
- An environmental health study in California using data considered under HIPAA to be sufficiently de-identified to be made public, yet more than 25% of participants were correctly identified. [Sweeney, Perovich, Boronow et al, Re-identification risks in HIPAA safe harbor data: a study of data from one environmental health study, Technology Science, August 28, 2017; https://techscience.org/a/2017082801/]
- A Canadian Community Health Survey in which most people could be identified from relevant facts excluding names and other key identifiers [Pinault, Tjepkema, Crouse, et al, Risk estimates of mortality attributed to

low concentrations of ambient fine particulate matter in the Canadian community health survey cohort, Environmental Health, 2016; https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4750218/]

Since drafting these comments, I have discovered an even more vital document: *"Comments of the International Society for Environmental Epidemiology on EOAs proposed rule on Strengthening Transparency in Regulatory Science." [http://www.youreventinfo.org/ISEE/Documents/ISEE_Comments_on_EPA-HQ-OA-2018-0259-0001FINAL_ISEE_submitted.pdf].* These comments were submitted to EPA but not to the Science Advisory Board, so I am forwarding them here for the convenience of the SAB members, including the newer ones. The ISEE members are the true experts in the field of environmental epidemiology including whether and how identities of research subjects can be protected. They conclude (see especially Section 3 [pages 7 – 11] that modern data science can now overcome efforts to redact personal data and still allow the research to be re-analyzed. They also have very critical comments on EPA's dose-response analysis proposal in Section 4. **The ISEE conclusion: "We urge EPA to withdraw the proposal."**

I urge the SAB members to read the ISEE submission. If you are not persuaded by the ISEE document itself, I urge you to hold a follow-on review directly with expert members of the ISEE (who cannot participate in this session due to a directly conflicting international meeting overseas).

The examples described by Dr. Schwartz, and others included in the ISEE comments, show that masking doesn't adequately hide identities of research participants. Even if in some instances it could, the fact that in many instances it cannot will have a chilling effect on participation in the studies that are needed as a basis for EPA regulations. Ask yourself, would you allow your 8-year-old to be a participant in a study of the effects of lead exposure on children's intelligence if there was a not insignificant probability that her identity could be unmasked despite the efforts to protect it? Of course not. So fielding usable studies would be increasingly difficult.

The result is that this seemingly well-intentioned proposal will either stifle the research needed by EPA or prevent its use by EPA. One suspects that

presumably well-meaning EPA leaders have been misled or fooled by proponents of this disastrous proposal.

Dr. Schwartz states " 'the gold standard' of science is not reanalysis, but replication," a view echoed by the ISEE. The SAB should redirect EPA's focus in that direction.

****

There is one more issue to be confronted in this consideration. **Why is this self-regulation being proposed just for EPA and not for all agencies involved in health-protective regulation?** There is no reason whatsoever that the same logic would not apply to regulations promulgated by the Occupational Safety and Health Administration, the Consumer Product Safety Commission, the National Highway Traffic Safety Administration, the Food and Drug Administration, and others. Why would EPA be regulated in this manner and not any other regulatory agency? Shouldn't this be a matter of legislation or regulation applying to every regulatory agency?

Simple answer: the effects of this regulatory proposal would cut both ways. Imagine the Food and Drug Administration being forbidden to approve new drugs unless the research justifying their use were subject to these same regulatory prohibitions. The privacy protections imagined in this proposal would not protect patient identities any more than they would subjects of research considered by EPA. Drug approvals would likely grind to a halt. As would EPA regulations for air quality standards, hazardous materials exposures, pesticides exposures, drinking water concentrations, etc.

This proposed self-regulation would make no sense for the FDA and it makes no sense for EPA. **Your answer to EPA should be that there is no way to provide certainty that privacy protections would not be violated, so approving this approach would eliminate the best use of research to underlie EPA regulation. It should be rejected.**

Signed: Roy N. Gamse