Beate Ritz., M.D., Ph.D., *President*
Manolis Kogevinas, M.D., Ph.D., *Past President*
Sara Adar, Sc.D, *Secretary-Treasurer*

Secretariat
Infinity Conference Group
(Contact: Doreen Albertson)
1035 Sterling Road
Herndon, VA 20170

703.925.0178 (Voice) 703.925.9453 (Fax)
secretariat@iseepi.org (E-Mail)
www.iseepi.org (website)

Dr. Michael Honeycutt, Chair
Science Advisory Board
U.S. Environmental Protection Agency Washington, DC

Re: Comments for the August 27, 2019 SAB Teleconference

Dear Dr. Honeycutt and members of the Science Advisory Board,

On behalf of the North American Chapter of the International Society for Environmental Epidemiology (ISEE), I wish to make two major points today with regard to the EPA's proposed approach to develop mechanisms for the secure access to Personally Identifying Information (PII) and Confidential Business Information (CBI):

1) In today's world of big data, it is no longer possible to publicly release CBI or PII datasets in a way that the privacy of those data can be guaranteed, and;
2) If the data are made available to vested interests as part of this process, past experience has shown that those data will likely be misused by their consultants in a way that serves the interests of those vested interests, and seeks to undermine the original study conclusions if the vested interest doesn't like those study results.

First, a fatal flaw in the EPA's data release proposal is that is no longer possible to effectively maintain the privacy of data made public in so called "de-identified" or "depersonalized" databases. Indeed, a study recently published in the journal Nature Communications (Rocher et al, 2019) shows this clearly. The study concludes; "Our results suggest that even heavily sampled anonymized datasets are unlikely to satisfy the modern standards for anonymization set forth by GDPR and seriously challenge the technical and legal adequacy of the de-identification release-and- forget model." A New York Times article on this study, entitled "*Your Data Were 'Anonymized'? These Scientists Can Still Identify You*" (July 23, 2019, also attached), states: "The investigators developed a method to re-identify individuals from just bits of what were supposed to be anonymous data". Thus, there is no such thing

anymore as truly de-identified personal health data, and the EPA proposal to make research data public is not achievable in a secure way.

Second, *abuse of this process by vested interests for their own purpose is likely the most dangerous aspect of this proposal.* This is not a case of concern this might happen, it has already happened in the past in real world situations. Just one past example of abuse by consultants to a vested interest (the tobacco industry) resulted when the State of Georgia set up an Open Records Law, as documented in a paper I previously published in the Tulane Environmental Law Journal when this idea was previously floated within the US Congress (Thurston, 1998). In that case, consultants for the RJ Reynolds Company used a Georgia state law that similarly allowed anyone to obtain research data in order to attack study findings that the use of cartoon characters (such as "Joe Camel") in tobacco advertising influences children's product recognition. That research was later validated in other studies, but the damage was done, and the physician involved left research for private practice. Thus, this data release approach has already been tried in the past, and has been shown to be too easily abused by vested interests.

Overall, we already know this EPA proposal would needlessly create new perils for the privacy of research participants' personal health data, as well as enable vested interests to misuse the data to serve their financial interests. Moreover, because the present EPA system of research study "weight of evidence" evaluation approach is already extremely rigorous, the proposed EPA data release requirements are not even needed. We strongly encourage EPA to withdraw this proposal.

Sincerely,

Professor George D. Thurston
NYU School of Medicine
On behalf of the ISEE North American Chapter

References:

Kolata G. (2019) Your Data Were 'Anonymized'? These Scientists Can Still Identify You. New York Times, July 23,2019.
https://www.nytimes.com/2019/07/23/health/data-privacy-protection.html

Roche L, Hendrickx JM, de Montjoye, YA. (2019) Estimating the success of re-identifications in incomplete datasets using generative models. Nature Communication. Jul 23;10(1):3069.
https://www.nature.com/articles/s41467-019-10933-3.pdf

Thurston, G.D. (1998) Mandating the release of health research data: issues and implications. *Tulane Environmental Law Journal.* 11(2):331-354.
https://yosemite.epa.gov/sab/sabproduct.nsf/8556A1302E44EFD18525829700752F08/$File/TulaneELJ_1998.pdf