Our OMB teleconference meeting on **30, September at 3:00 pm ET** for thirty minutes. Conference ID: 7943783

Good afternoon, I am George Thurston, Chair of the Policy Committee of the North American Chapter of the International Society for Environmental Epidemiology. We have a half dozen points our committee wishes to make regarding the severe problems with this EPA Regulation. I will introduce each member of this committee and their institution, though my views and theirs do not necessarily reflect those of the institutions where they do their research work.

1. **Dr. Carrie Breton, USC School of Medicine, will discuss the regulation's Privacy Conflicts** with HIPAA and with research Signed Consent forms:

- HIPAA was created to protect the privacy of patient health data and in many cases it does not allow for sharing of private health data without express permission from the patient. To participate in human subjects studies, people sign informed consents. Informed consents exist to ensure individuals understand what science they are participating in and how their data will be used and protected and shared.
- Large amounts of data from scientific studies in which informed consents were signed cannot be shared with the EPA even if we scientists want to because we must honor and comply with human subjects rules. This means EPA is knowingly excluding a large body of very good science.
- The proposed rule would onerously apply retroactively, not just to future research and publications. For example, when EPA reassesses its standards every five years, it will downgrade consideration of studies from the past that do not conform to the new rules. Since many past studies cannot practically comply with this rule, as they cannot ethically release the personal health data collected under informed consents which did not expressly state their data would be shared with EPA, this will again sideline much of the best available science from consideration by the EPA during pollutant standard revisions. This will weaken the scientific foundations of the prevailing environmental standards, and almost certainly lead to diminished public health protections in the United States.
- Another point of concern has to do with use of electronic medical records in secondary studies. If EMR data is used for a study and then shared with the EPA, this will undermine trust in physicians, and could potentially harm individuals if their sensitive data is accessed and misused. Moreover it may be in violation of HIPAA law.
- The EPA has no plan for how they will protect the very sensitive information being requested by this rule. There is no system in place so how can they guarantee privacy.
- This is simply a barrier knowingly being put in place to exclude some of the best science that exists in support of environmental regulations.

2. **Dr. Joan Casey, Columbia School of Public Health, will discuss the impracticability of the regulation**.

I use large electronic health record datasets to study the relationship between emerging exposures such as wildfires, increasingly strong and frequent hurricanes, power outages, and drought and temperature

spikes and health. I am concerned about re-identification of individuals from supposedly de-identified, limited, or full datasets.

We know that people are concerned. A 2017 study (Pereira et al. Plos One) found that about 70% of surveyed Americans already reported being concerned about the privacy and security of their health information.

Earlier work on the topic of balancing data privacy with potential social benefits of governmental databases. One study (Duncan 2004 Privacy and Social Benefits) found that 67% of respondents were concerned about medical record privacy and that over a quarter of patient would withhold information from their provider or delay care seeking based on privacy/data security concerns. This is not good news for research or health.

Particularly related to environmental health data, we have big concerns with being able to identify specific individuals in a study, even if we have tried to de-identify their information. To study environmental health, researchers generally have access to spatial information or where a study participant lives. When paired with other information about an individual, re-identification becomes fairly simple. A study in 2010 found that about 95% of study participants could be uniquely identified with just a geographic coordinate, age, disability status, income, martial status, and schooling information. Newly available data, such a genotype complicate our ability as researchers to de-identify datasets further. A 2013 study in Science showed how short tandem repeats on the Y chromosome can be used to determine participant surname even when data were fully de-identified.

As a researcher who relies on high-quality secondary electronic health record data to study health including that of pregnant women, people with dementia, those who are food insecure, having access to spatial data is critical. I find the risk of placing such key datasets in a central repository where they have a high probability of being re-identified despite our efforts very problematic. Study participants deserve better and researchers need assurance to be able to conduct the highest quality science with the best available data.

3. **Dr. Kelvin Fong, Yale School of Environment, will discuss the onerous cost of this regulation, which represents an unfunded mandate**.

In 2015, The Congressional Business Office (CBO) evaluated the cost to the government of H. R. 1030, referred to as the Secret Science Reform Act, that was similar to this proposed regulation in that it that would have prohibited the EPA from "proposing, finalizing, or disseminating regulations or assessments based upon science that is not transparent or reproducible. That estimate therefore provides insight into the costs of EPA's now proposed regulation. HR 1030 The CBO estimated that the EPA would need to spend $250 million dollars a year to cover the costs of data collection, database construction, and other activities to meet the legislation's requirements.

Not only is this cost large and unfunded, but $250 million would represent more than what the EPA requested in its 2020 fiscal year budget for the combined costs of its research programs in air quality (31.7 million), safe and sustainable water resources (70), sustainable and health communities (65.5), and chemical safety (63.9). Redirecting any money towards meeting the this or a similar legislation's "transparency" requirements would take away from funding these research programs and almost certainly undermine the goal of achieving "the best available science" needed to protect the lives of Americans.

The Congressional Budget Office notes that EPA reviews around 50,000 scientific studies per year for determining the National Ambient Air Quality Standards. It's estimated it would cost the EPA between $10,000 and $30,000 for each scientific study. To arrive at the 250 million cost estimate, we would need to assume the lower $10,000 estimate for cost per study AND cut the number of studies reviewed for reviewing the air quality standards by half. Please keep in mind that this is for reviewing air quality estimates alone, and it discounts EPA's other activities, such as protecting the water and soil.

If the responsibility of making all data available is passed on to researchers such as myself, I do not see how we can realistically meet the requirements of this regulation. My colleagues have previously touched on the privacy and technical challenges. I would like to add that making these data available is not as simple as it sounds. For every study that I publish, I would need to pay for webhosting of research data, which is only getting larger in the age of big data. From what I understand, I would need to shoulder these costs indefinitely. Might I remind everyone that most research grants, including those from government agencies, typically last for 5 years at most and not a lifetime and beyond. Since my primary responsibilities are to conduct research and to train future scientists, I would need to hire a new data manager to prepare, deidentify, and manage the hosting of these data. This person would also need to know the ins-and-outs of internet and network security, since a platform of scientific data could be a prime target for cyber-criminals from around the world.

In summary, the costs of proposed legislation would prevent the EPA from reviewing the "best available science," ultimately putting American's lives at risk. At most research institutions, we the scientists do not have the financial resources, time, expertise, or personnel to fulfill the requirements of this legislation.

(Epa 2020 budget proposal https://www.epa.gov/sites/production/files/2019-03/documents/fy-2020-epa-bib.pdf)

4. **Dr. Abiodun Oluyomi of Baylor College of Medicine, will discuss the fact that this regulation is not needed**.

Thank you again for the opportunity to address the OMB today. To follow up on what my colleagues have described so far, I will address the fact that there are already several existing solutions to the concern that the EPA Transparency rule aims to address. The concern according to the EPA is primarily about **study validation and analysis**.

I present three specific existing solutions here:

1. There is already a peer review system in place for many aspects of what researchers do before their work is published. Starting with grant writing, many grants are peer-reviewed and scrutinized for their data analysis approaches even before these grants get funded. Also, and perhaps more importantly, there is already a healthy and robust peer review system in place for all published research findings – where researchers across the globe get the chance to assess and validate study findings to determine whether such findings may be allowed to proceed to publication, or in fact be rejected.

2. The success of the academic research enterprise at large is based on the replication of studies by independent researchers and among different populations. This process (of study replication) ends up creating a library of research findings on any given topic, including topics related to environmental protection. Because of the near universal access to published work in the just mentioned library, parties

representing all aspects of civic, business, and government endeavors may access and review the compendium of study findings produced by disparate research groups on the same topic. This process inherently offers all parties the opportunity to conduct "systematic reviews" of published work and examine (or validate) findings over time.

3. Additionally, when specific and credible research quality concerns are raised, a phenomenon that is quite uncommon, an assessment can be conducted by a qualified independent body with domain knowledge about the topic of interest. A notable example in this regard is the independent review of past studies (the 6-Cities Study and the ACS Study) that The Health Effects Institute (HEI) conducted to address concerns by the public. HEI is a nonprofit independent research organization that could be called on again to address future concerns, especially those related to air quality. In fact, instead of investing in new rulemaking, the EPA may decide to increase support for organizations already able to re-assess research findings when credible concerns are raised by the public.

5. **Dr. Tracey Woodruff, UC San Francisco School of Medicine, discusses the deficiencies of the regulation in practice**, and that it is "Not ready for prime time" because so many needed procedural details are missing or undecided.

EPA should not promulgate significant new regulatory and scientific practices based on untested approaches for securing private information.

In the supplemental proposal, EPA indicates that it "*…is currently conducting a pilot study using the RDC's secure data enclave to host EPA datasets in a restricted use environment. Development of standard data repositories is still ongoing."*[1] In essence, EPA is proposing to change the scientific basis of their rulemaking based on an untested and unverified method for data security enclaves. There has been a nascent but growing body of research on data security. Researchers in 2015 were able to reidentify 90 percent of their study population as unique individuals and to uncover their records, knowing just four random pieces of information, due to the uniqueness of human behavior.[2] A more recent study in 2019 modeled the likelihood of individual reidentification in a "heavily incomplete" anonymized dataset, and found "that 99.98% of Americans would be correctly re-identified in **any** dataset using 15 demographic attributes."[3](emphasis ours This research shows how, with only a few pieces of information, current technology can already deidentify data that scientists currently consider sufficiently anonymized. Therefore, EPA needs to put out a rule that has a tested security method in place, not a 'plan' for a system that has not been tested and evaluated. EPA should not be basing regulations on pilots, but rather should focus on implementing EPA's own 2016 *Plan to increase access to results of EPA-funded scientific research*.[4]

EPA should focus on implementing its 2016 EPA Plan to increase access to results of EPA-funded scientific research.

[1] EPA (2020). Strengthening Transparency in Regulatory Science. Pg 15402. Available: https://www.regulations.gov/document?D=EPA-HQ-OA-2018-0259-9322

[2] Montjoye, Y.-A. D., Radaelli, L., Singh, V. K., & Pentland, A. S. (2015). Unique in the shopping mall: On the reidentifiability of credit card metadata. *Science*, *347*(6221), 536–539. Available: https://doi.org/10.1126/science.1256297

[3] Rocher, L., Hendrickx, J.M. & de Montjoye, Y. (2019). Estimating the success of re-identifications in incomplete datasets using generative models. *Nat Commun* **10,** 3069. Available: https://doi.org/10.1038/s41467-019-10933-3

[4] EPA (2016) Plan to increase access to results of EPA-funded scientific research. Available: https://www.epa.gov/sites/production/files/2016-12/documents/epascientificresearchtransperancyplan.pdf

A 2013 memo from the Office of Science and Technology Policy discusses policy principles and the development of federal agency plans to increase public access to federally funded research.[5] The objectives were developed in consultation with the National Science and Technology Council and with input from the public. In response, EPA developed the 2016 *Plan to increase access to results of EPA-funded scientific research*.[6] (2016 EPA Plan) The 2016 EPA Plan differs from the supplementary proposal in three critically important ways.

First, the 2016 EPA Plan's scope appropriately "prospectively covers peer-reviewed scientific research publications in scholarly journals and digital research data that result from EPA-funded research. The Plan does not apply to research publications or research data generated from scientific research conducted prior to the implementation of the Plan."[7] The 2016 EPA Plan does not apply retroactively, and thus would not impact research underpinning regulations which come up for renewal.

Second, the 2016 EPA Plan emphasizes that data availability does not affect the validity or usability of science, noting:

"Whether research data are fully available to the public or available to researchers through other means does not affect the validity of the scientific conclusions from peer-reviewed research publications."[8]

"The validity of scientific conclusions drawn from research publications or their associated research data, or EPA's ability to consider those conclusions and data in its actions, does not depend on compliance with this Plan."[9]

In stark contrast, the supplemental proposal indicates that:

"…when promulgating significant regulatory decisions or finalizing influential scientific information, the Agency will, other things equal, ***give greater consideration to studies where the underlying data and models are available*** *in a manner sufficient for independent validation either because the information is publicly available or available through tiered access when the data include CBI, proprietary data, or PII and* ***appropriate techniques have been used to reduce the risk of re-identification.****"* [10] (Emphasis ours)

EPA is proposing to downgrade or exclude science that is not made publicly available or "*cannot be sufficiently de-identified to protect the data subjects,*" [11] This de facto results in a rule that will equate study quality with availability of underlying data. Availability of underlying data is not a measure of study quality or validity – as EPA itself wrote in the 2016 EPA Plan. Further, study quality is dependent on the methods used in the study (further discussed below in Point 7).

Finally, the 2016 EPA Plan is in compliance with EO 12291 (Point 2) and acknowledges the significant costs to researchers that data access may impose, noting "Inclusion of costs for data management

---

[5] Executive Office of the President, Office of Science Technology and Policy. Memorandum for the Heads of Executive Departments and Agencies: Increasing Access to the Results of Federally Funded Scientific Research. February 22, 2013. Available: https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

[6] EPA (2016) Plan to increase access to results of EPA-funded scientific research. Available: https://www.epa.gov/sites/production/files/2016-12/documents/epascientificresearchtransperancyplan.pdf

[7] Id. pg. 5.

[8] Id. pg. 4-5

[9] Id. pg. 6

[10] EPA (2020). Strengthening Transparency in Regulatory Science. Pg 15403. Available: https://www.regulations.gov/document?D=EPA-HQ-OA-2018-0259-9322

[11] Id. Pg 15402.

and public access may be included in intramural and extramural research proposals."[12] Thus, the 2016 EPA Plan both references and sets up a mechanism for addressing the costs it imposes. The 2016 EPA Plan is scientifically and technically sound; thus, EPA should abandon the flawed proposed rule and its associated supplemental proposal and focus on implementing the 2016 EPA Plan.

6. Dr. George Thurston, of the NYU School of Medicine, will **discuss the potential for abuse by Vested Interests of this proposed data sharing program.**

I want to briefly point out the grave potential for vested Interests and their consultants to abuse the data that would be made publicly available under this regulation. In the past, they have accessed such data in order to inappropriately attempt to undermine the credibility of specific studies, and environmental health science in general. This is made more likely because of the fact that the EPA proposed rule does not address likely Conflicts of Interest (and in particular financial conflicts of interest) by those requesting access to data, opening the possibility for research data that are to be used for regulatory purposes to be accessed by anyone, regardless of their affiliation. Thus, it will not prevent agents of vested interests from obtaining the publicly available data demanded by this proposed rule, and then running their own studies with the intention of discrediting valid science. There is already empirically based evidence that results of studies conducted by the pharmaceutical and tobacco industry favor those industries; indicating that we can anticipate self-serving biased analyses of EPA provided data by vested interest funded analysts under this proposed EPA approach (White and Bero, 2010; Bero, 2017; Lundh et al., 2017). Indeed, it has been previously documented in an extensive review I wrote, back in 1998, of a similar proposal made by a vested interest (a tobacco company), upon which this EPA proposal is based: when such data have been accessed using a similar provision in the state of Georgia, the data were abused in an attempt to discredit the research they did not like, but the original research was later replicated by other researchers (Thurston, 1998). This new attempt to implement the tobacco company-inspired data release proposal will undoubtedly have similar inappropriate and untoward damage to scientific research and government decision making if this so-called "Transparency" rule is implemented.

We therefore recommend that this regulation not be approved, but instead sent back to the EPA for reconsideration. Thank you for your attention to our concerns.

[12] EPA (2016) Plan to increase access to results of EPA-funded scientific research. pg. 11 Available: https://www.epa.gov/sites/production/files/2016-12/documents/epascientificresearchtransperancyplan.pdf