

Comments for the Public Meeting of the Chartered Science Advisory Board on Strengthening Transparency in Regulatory Science

These comments on the EPA's proposed "Transparency" rule are submitted on behalf of the International Society in Environmental Epidemiology (ISEE) North American Chapter. ISEE represents scientists who study environmental determinants of health, the studies of whom have supplied a substantial part of the research that is the basis of regulatory standards on environmental agents, such as the National Ambient Air Quality Standards and standards for heavy metals, pesticides, drinking water, and other environmental contaminants. The North American Chapter of the ISEE is writing in strong opposition to the proposed Transparency rule that will greatly affect which studies can be considered in setting such standards. The adoption of such of rule, in our view, would severely weaken environmental policies and increase risks to human health.

One of the arguments proffered for promoting the Transparency rule is that researchers will thereby have the opportunity to reanalyze published studies and evaluate consistency in results. However, the gold standard in drawing conclusions about any environmental impact on health is not the reanalysis of the same dataset, but rather replication in other settings, i.e., other populations, other locations, and using other study designs and statistical methods. Any one single study can be susceptible to different issues: faulty study designs, lack of information on important variables, such as potential confounders, and other sources of bias. Therefore, it is only with the collective body of evidence across different study populations and locations that we assemble conclusive epidemiological evidence for the effect of environmental contaminants on adverse health outcomes. For example, Di et al.¹ in Figure S6 in the paper's Supplemental Material provide published effect estimates of long-term exposure to fine particulate matter air pollution on mortality conducted by numerous studies covering different time periods, study areas and countries, populations, study designs, and analytical methods. All these studies reported harmful effect estimates similar in magnitude. Similarly, the 2019 Integrated Science Assessment (ISA) for Particulate Matter² included a similar figure for short-term effects of fine particles on mortality in multi-city studies, most of which also reported adverse health effect estimates. Alone, each individual study may be susceptible to a different set of potential biases, depending on how the data were collected, the primary reason for data collection, the available information, etc. Reanalysis, thus, of the raw data of any single study would then also be susceptible to the biases that are not necessarily related to the statistical methods used for analysis, and the EPA's proposed rule will not provide the purported benefits. This means that reanalysis of single biased datasets does not help solve the overarching goal to estimate the causal relationship between environmental exposures and adverse health outcomes to inform regulatory policy. Furthermore, with any single analysis or reanalysis it is easier to cherry pick or manipulate the one finding to suit one's view, which is not possible when viewing the evidence in its totality.

Furthermore, all published studies have gone through a rigorous peer review process. This means that multiple independent experts on the study topic (e.g., air pollution epidemiology, lead effects on neurodevelopment, pesticide exposure and Parkinson's disease, etc.), unrelated to the researchers that conducted the study, carefully review the study design and analytic methods, and provide comments back to the editor and the authors. The authors subsequently revise their analyses and manuscripts accordingly or provide justification on their choices and why they believe their initial analysis/interpretation/etc. is appropriate. A paper is only accepted for publication after reviewers and editors are convinced that the study is scientifically sound and all potential biases are explicitly and clearly discussed. This process further ensures the validity of the multiple replication studies of any single relationship between environmental exposures and adverse health. The findings of any reanalysis of raw data should not be considered unless they too are peer-reviewed by those with the requisite, independent, and objective expertise. Finally, it should be noted that peer-reviewed published papers already include all key information on the studies, i.e., study protocols, recruitment and exclusion criteria, outcome and exposure assessment approaches, statistical modeling, including adjustment variables

and how they were assessed, and descriptive statistics. Thus, the EPA proposal seeks to address a problem that does not exist, is not needed, and can ultimately produce findings that can readily be manipulated.

An additional concern of the proposed rule is the protection of the information provided by study participants. Most cohort studies use information on people who have agreed to provide personal data to study researchers and have signed consent forms on how their data can be used for science. Consent forms are a mandatory component of conducting human research, and are specifically designed to protect rights of participants. To the best of our knowledge, most of these consent forms do not include making the data publicly available. Furthermore, many of these consent forms were signed many years ago. It would thus be infeasible to try to re-consent most study participants, even if they were to agree to make their medical histories publicly available, which is unlikely. Additionally, contacting subjects down the road could well likely introduce important sources of bias that cannot be readily quantified (those who agree vs. do not to allow subsequent use of data). For most studies, therefore, reanalysis would not be possible. Moreover, the Health Insurance Portability and Accountability Act (HIPAA) Privacy and Security Rules have established national standards for the protection of certain health information and how this information can be held or transferred in electronic form. These rules protect all individually identifiable health information, including the individual's past, present, or future physical or mental health or condition. Similarly, Institutional Review Boards that review all proposed research by universities and other government-funded research organizations require the protection of data from study participants. Europe has just tightened its data privacy laws with the General Data Protection Regulation (GDPR) and Canadian privacy laws also reject the idea that personal information from participants in research studies could ever be made public. GDPR, specifically, limits movement of private data outside of the EU and states that the data controller must "demonstrate that the data subject has consented to processing of his or her personal data" (article 7). If the data were made public, even (if possible) in unidentifiable ways, the data controller could still not demonstrate that the data were only being used for the purposes and by the people to whom the participant in the study consented. Requiring all study data be made public, thus, would automatically disqualify most cohort studies from inclusion in informing regulatory action, if the Transparency rule were implemented. Cohort studies are thought to represent the "gold standard" or the best type of data available for public health research. This is because cohort studies collect data systematically across study sites, using the same questionnaires, scales, and technologies. In addition, the detail of data collected in cohort studies is available nowhere else. If the Transparency rule were implemented, the majority of the existing studies, the best studies we have, conducted all around the world, historically used to inform regulations to protect the health and lives of Americans, will no longer be eligible to continue doing so. This is a huge disservice to the public health of Americans.

There have been discussions about potentially making de-identifiable data publicly available as a part of this proposed rule, but this suggestion is also not satisfactory for two main reasons. First, some key variables for analysis may be excluded prior to dissemination of the dataset to ensure de-identifiability. If these variables are critical for the study question, e.g., information on potential confounders, then any reanalysis would be biased and, of course, the results would not match those of the original analysis. Second, and most importantly, it is near-impossible to ensure de-identifiability in the era of Big Data. A recently published paper³ showed that with only 15 demographic attributes >99.9% of Americans can be correctly identified in any dataset. The authors concluded 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."³ Several other studies have also reported similar findings,⁴⁻⁶ challenging the presumed protection of the identity of study participants when "de-identified" datasets are made publicly available. Thus, the proposed rule would undermine the privacy of past and future participants in environmental health research.

Even if existing studies are exempt from the proposed rule, the Transparency rule poses significant questions on how we will conduct future studies. First, it would be much harder to recruit study participants. These participants would have to agree that their demographic information and medical history become publicly available, resulting in

much smaller study populations and, consequently, severely reducing the statistical power to detect effects. This is particularly problematic given known disparities in who participates in cohort studies to begin with (generally healthier, wealthier people and women). Individuals from stigmatized groups, those with risk behaviors like smoking, drinking, and drug use, and those of lower socioeconomic status are already less likely to participate.⁷ Requiring these individuals to make their data publicly available would further reduce their representation in studies and perpetuate unequal representation in scientific research. This issue relates to the second point: because most people would not agree to make their information publicly available, results from studies that depend on participants who do will not easily be generalizable to the general population, further limiting the ability of these studies to inform regulatory action.

In summary, the Transparency rule will pose severe threats to the privacy of study participants, exclude well-conducted, peer-reviewed published studies from informing policymaking, and, ultimately, lead to inadequate protection of the health and lives of millions of Americans. We urge EPA to withdraw the proposal.

Sincerely,

Joan A. Casey

ISEE North American Chapter Co-Chair

Assistant Professor

Columbia University Mailman School of Public Health

Department of Environmental Health Sciences

Marianthi-Anna Kioumourtoglou

ISEE North American Chapter Co-Chair

Assistant Professor

Columbia University Mailman School of Public Health

Department of Environmental Health Sciences

On behalf of the ISEE North American Chapter



References

1. Di Q, Wang Y, Zanobetti A, et al. Air pollution and mortality in the medicare population. *N Engl J Med*. 2017. doi:10.1056/NEJMoa1702747.
2. USEPA. *Integrated Science Assessment for Particulate Matter (Final Report)*. Vol EPA/600/R-. Research Triangle Park, NC: Center for Public Health and Environmental Assessment Office of Research and Development; 2019.
3. Rocher L, Hendrickx JM, de Montjoye YA. Estimating the success of re-identifications in incomplete datasets using generative models. *Nat Commun*. 2019;10(3069). doi:10.1038/s41467-019-10933-3.
4. Sweeney L. Matching Known Patients to Health Records in Washington State Data. *SSRN Electron J*. 2013. doi:10.2139/ssrn.2289850.
5. De Montjoye YA, Radaelli L, Singh VK, Pentland AS. Unique in the shopping mall: On the reidentifiability of credit card metadata. *Science (80-)*. 2015. doi:10.1126/science.1256297.
6. Narayanan A, Shmatikov V. Robust de-anonymization of large sparse datasets. In: *Proceedings - IEEE Symposium on Security and Privacy*. Vol ; 2008. doi:10.1109/SP.2008.33.
7. Galea S, Tracy M. Participation Rates in Epidemiologic Studies. *Ann Epidemiol*. 2007;17(9):643-653. doi:10.1016/j.annepidem.2007.03.013.