



OFFICE OF GOVERNMENT AND PUBLIC AFFAIRS
1436 Duke Street • Alexandria, VA 22314
www.populationassociation.org • www.popcenters.org • 301-565-6710 x 1006



POPULATION ASSOCIATION OF AMERICA
President

Dr. Sonalde Desai
University of Maryland

Vice President
Dr. Jennifer Van Hook
Pennsylvania State University

President-Elect
Dr. Lisa Berkman
Harvard University

Vice President-Elect
Dr. James Raymo
Princeton University

Secretary-Treasurer
Dr. Michelle Frisco
Pennsylvania State University

Past President
Dr. Robert Hummer
University of North Carolina-
Chapel Hill

Board of Directors
Dr. Jennifer Barber
Indiana University

Dr. Tyson Brown
Duke University

Dr. Deborah Carr
Boston University

Dr. Elizabeth Cooksey
Ohio State University

Dr. Jennifer Johnson-Hanks
University of California, Berkeley

Dr. Hedwig Lee
Washington University in St. Louis

Dr. M. Giovanna Merli
Duke University

Dr. Jennifer Karas Montez
Syracuse University

Dr. Emilio Parrado
University of Pennsylvania

Dr. Rogelio Saenz
University of Texas at San Antonio

Dr. Christine Schwartz
University of Wisconsin-Madison

Dr. Florencia Torche
Stanford University

ASSOCIATION OF POPULATION CENTERS
President

Dr. Sara Curran
University of Washington

Vice President
Dr. William Dow
University of California, Berkeley

Treasurer
Dr. Jennifer Karas Montez
Syracuse University

Secretary
Dr. Jennifer Glick
Pennsylvania State University

June 27, 2022

National Institutes of Health
Office of Science Policy
Building One
Bethesda, Maryland 20892

Delivered to: <https://osp.od.nih.gov/rfc-draft-supplemental-information-to-the-nih-policy-for-dms/>

To whom it may concern:

On behalf of the organizations we lead, the Population Association of America and Association of Population Centers, we are pleased to submit comments in response to “Request for Public Comments on DRAFT Supplemental Information to the NIH Policy for Data Management and Sharing: Protecting Privacy When Sharing Human Research Participant Data” (NOT-OD-22-131).

Our organizations jointly represent about 3,000 individual population scientists—including demographers, economists, sociologists, and epidemiologists—as well as approximately 40 federally funded interdisciplinary research centers. Population scientists study the individual, societal, and environmental implications of population change—and thus contribute key findings that help inform evidence-based policy making in the public and private sectors. Population research centers facilitate interdisciplinary research on a range of topics including mortality, morbidity, fertility, adolescent health, aging, population forecasting, immigration, labor and workforce policies, family dynamics, and human-environmental interactions. They also train emerging and early career scientists. In addition, they have unique expertise regarding data collection, management, and archiving, and the field embraces data sharing as a central principle.

The National Institutes of Health is the primary source of competitive, discretionary grant funding supporting the population sciences. To that end, our organizations serve on NIH study sections, working groups, and advisory councils, and provide comments, both formal and informal, on a variety of NIH proposals, including its evolving data management and sharing policy.

We appreciate the opportunity to respond to this latest request for comments that NIH has issued to inform its goal of developing a set of principles and best practices for protecting the privacy of research participants when sharing data under the NIH Policy for Data Management and Sharing.

The community of researchers we represent often employ data that are highly complex and socially relational. They might be longitudinal studies that are underway for years. They might be data that integrates layers of information to add social context or to identify the geographic location of individuals or where their behaviors take place, or many other layers at multiple scales of observation and aggregation. In data science terms, our data tables are very long and very wide, and we have many data tables with linkage identifiers to facilitate the analyses of critical health outcomes. Sometimes our scientists integrate existing data with newly found data to create entirely new datasets.

Consequently, our scientific community is sensitive to the confidentiality and privacy protections that should be employed to avoid deductive disclosure or any inadvertent disclosure that might be in violation of consent agreements or harmful to individuals or vulnerable groups. Importantly, not all our data are collected through consent from individuals. Instead, many of our scientists create innovative data by integrating multiple kinds of information, which in turn may make the data even more sensitive to privacy risks than before the integration. Thus, since the data we generate in our scholarship often has important value to a much wider community, we have developed professional practices and expertise in data sharing that preserves the integrity and value of the data, while protecting privacy. Our scientists, therefore, can be outstanding resources for NIH as the Office of the Director continues to develop policies and procedures around Data Management and Sharing.

Given our scientists experience and expertise, we reviewed the *NOT-OD-22-131: Protecting Privacy When Sharing Human Research Participant Data* with a great deal of interest and appreciation for its intent. Overall, the set of proposed principles, best practices, and points to consider for creating a robust framework for protecting participant privacy is good, but not complete. We would urge the NIH OD to additionally consider:

- That de-identification can undermine data quality and fitness for use for maintaining the highest quality research products. This point is partially addressed by the NIH OD under ‘DRAFT Operational Principles for Protecting Participant Privacy When Sharing Scientific Data’ item #1 bullet point three. However, we urge NIH OD to change the language from normative to possible – e.g. from ‘may be’ to ‘can be’ or ‘will be.’ Doing so will ensure that there will be support for the availability, accessibility, and secure protection of identifiable information for the highest quality data products for research purposes to advance health science and knowledge.
- Elaborating further the specific concerns related to collection of data from non-traditional research settings under ‘DRAFT Operational Principles for Protecting Participant Privacy When Sharing Scientific Data’ item #5.
 - The OD should specifically elaborate how these data were not collected with consent for the purposes of research and often include extremely sensitive information about individual behavior and identifying information.
 - Furthermore, OD should explain how data use agreements from providers are often meant to protect the proprietary interests of the commercial data collector or the agency interests and not the individual’s whose data is being shared. Thus, researchers using these data should not assume that a data use

agreement from a third-party provider is fully considerate of the privacy considerations of the individuals whose information is being shared.

Researchers should assess whether additional strict privacy considerations are warranted.

- Deductive disclosure risks are not sufficiently discussed in the current draft. While de-identified data are well-discussed, the risk of deductive disclosure requires more elaboration so that researchers are sensitized to the risks and prepared to limit and manage them. Deductive disclosure arises when highly relational data (e.g. as a result of longitudinal study designs or complexly integrated data) makes it possible to inadvertently discover the identities of an individual or vulnerable group within the data. While the current draft describes re-identification following de-identification, there are several concerns with the limited discussion of these concerns under (a) ‘DRAFT Best Practices for Protecting Participant Privacy’ items #1 and #3 and (b) ‘DRAFT Points to Consider for Designating Scientific Data for Controlled Access’.
 - For the former (a), the discussion appears to be limited to a relatively narrow range of data types and not fully considerate of the possible risks and options for protecting privacy, while advancing scientific knowledge.
 - For the latter (b), there is no mention of points to consider when the data were collected without consent of the individuals (e.g. non-traditional sources or at risk of deductive disclosure). This is a significant omission and should be redressed. For both the former (a) and latter (b), we recommend that NIH OD convene a small group of social and behavioral scientists who work with such highly sensitive data to advise on elaborating additional best practices and points to consider for these research cases.
- Finally, we strongly recommend that NIH OD include specific points in each draft section that address comprehensive data documentation as a policy requirement. Data management and sharing principles are important for increasing accessibility to data, but without data documentation principles, standards, and best practices, accessibility will be limited. The time and effort required for high quality data documentation is especially important for junior investigators with less experience in secondary data analysis. The DMS Policy should consistently acknowledge the costs and value of these efforts. In relation to privacy considerations, we recommend that NIH OD include:
 - A principle of comprehensive data documentation to accompany all data sharing and use agreements;
 - A practice of data documentation that fully elaborates the structure and content of the data and alerts the researcher to potential risks and appropriate data management for the particular datasets; and,
 - Points to consider for data documentation when managing controlled access.

Finally, we would like to reiterate that the scientists represented by our associations have particularly valuable insights about data management and sharing when research involves integrating data from multiple sources, including administrative or government sources, mobile devices, geo-located information, social media, and biomarkers collected in community settings (e.g. outside clinical settings). Their expertise, particularly in relation to

major population health studies, can be called up to support the development of NIH OD's Data Management and Sharing policies.

Sincerely,

Dr. Sonalde Desai



President
Population Association of America

Dr. Sara R. Curran



President
Association of Population Centers