

October 25, 2011

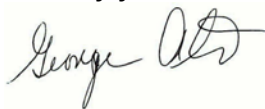
Dr. Jerry Menikoff
OHRP
1101 Wootton Parkway, Suite 200
Rockville, MD 20852

Dear Dr. Menikoff:

On behalf of the Inter-university Consortium for Political and Social Research, I am pleased to submit the attached responses to the Advance notice of proposed rulemaking, "Human Subjects Research Protections: Enhancing Protections for Research Subjects and Reducing Burden, Delay, and Ambiguity for Investigators"

Please feel free to contact me if there is any way that we can assist your office in this important process.

Sincerely yours,



George Alter
Director, ICPSR



ICPSR COUNCIL

Ann Wolpert, Chair • Massachusetts Institute of Technology

Francine Berman
Rensselaer Polytechnic Institute

G. Sayeed Choudhury
Johns Hopkins University

Paul N. Courant
University of Michigan

Catherine A. Fitch
University of Minnesota

Thomas LaVeist
Johns Hopkins University

Jeffrey Moon
Queen's University

Gregory N. Price
Morehouse College

Rogelio Saenz
Texas A&M University

Barbara Schneider
Michigan State University

Lori M. Weber
California State University at Chico

Christopher Zorn
Pennsylvania State University

Aletha C. Huston, Past Chair
University of Texas at Austin

Response to Advance notice of proposed rulemaking, "Human Subjects Research Protections: Enhancing Protections for Research Subjects and Reducing Burden, Delay, and Ambiguity for Investigators"

Inter-university Consortium for Political and Social Research
Institute for Social Research
University of Michigan
October 25, 2011

The Inter-university Consortium for Political and Social Research (ICPSR) strongly supports the goals of the ANPRM, and we wish to see changes that will reduce the burden of IRB review on types of research that involve minimal risks and develop a new approach to informational risk. However, we fear that the data security recommendations in the ANPRM risk imposing a new and unnecessary burden by not distinguishing between data that can be made available for public use with little or no risk and a much smaller group of data sets that do require strict protections. We suggest instead that HHS carefully consider a confidentiality protection system described and recommended by the National Research Council (NRC) that can meet the challenge of sharing of research data safely.

Key Recommendations

- We strongly support the ANPRM's goal of streamlining procedures for research involving minimal risk to subjects.
- We encourage HHS to adopt the ANPRM's suggestion that re-identification of subjects who have been promised confidentiality is a violation of research ethics and subject to appropriate sanctions.
- We recommend that HHS recognize organizations with the expertise needed to determine which data collections pose "informational risk," so that stringent data security measures are focused on data that need such protections.
- We recommend that HHS recognize that secondary analysis of data that are publicly available be designated as "not regulated" and not subject to IRB review.

About ICPSR

The Inter-university Consortium for Political and Social Research (ICPSR) is the world's largest archive of social science data. More than 100,000 users download data from ICPSR every year. Since our creation in 1962, we have expanded to

provide quantitative data across all social science disciplines. The Consortium includes more than 700 universities and research organizations located around the world, and we disseminate data for a range of government agencies and other groups, including the Bureau of Justice Statistics, the National Institute on Aging, the Substance Abuse and Mental Health Services Administration, and the National Collegiate Athletic Association. Our archive has more than 8000 research collections, some of which include hundreds of datasets. The highly regarded ICPSR Summer Program in Quantitative Methods offers more than fifty courses every summer, and almost 900 participants attended in 2011. ICPSR was also one of the founding members of the Data Documentation Initiative (DDI), which has become an international standard for metadata in the social sciences, and we provide the home office for the DDI Alliance.

Data Security

ICPSR supports the intent of the ANPRM to simplify the IRB process and increase the protection of sensitive information about subjects. The current situation imposes on IRBs the responsibility for evaluating “informational risk,” an area in which few IRBs have sufficient expertise. This is a growing problem, because research designs in the social sciences increasingly involve methods that make subjects easier to re-identify. However, we find that the approach to data security outlined in the ANPRM is poorly suited to the research environment in the social, behavioral, and economic sciences.

We support an alternative approach described by previous reports, such as the National Research Council’s 2003 report *Protecting Participants and Facilitating Social and Behavioral Sciences Research* (1) and reiterated in the 2005 NRC report *Expanding access to research data: reconciling risks and opportunities* (2). Their recommendations will be much easier to implement and will provide greater assurance of data security. In addition, we call your attention to several important aspects of data use in the social sciences that are not captured in the ANPRM.

First, applying a single standard for data security on all forms of data would waste effort and weaken controls on the small number of data collections that do pose risks to subjects. NIH, NSF, and other sponsors finance many important data collections that are widely disseminated in public-use form. Most researchers in the social sciences perform secondary analysis of data that are publicly available. Of the 8,000 data collections that ICPSR distributes, only 10 percent include data files designated as restricted-use. Standard data security measures, like isolating machines with sensitive data from the Internet, would impose unnecessary costs on

a very large number of analysts. Researchers are likely to resent and undermine data security procedures that lack a clear justification in the protection of research subjects.

Second, most researchers in the social sciences work alone or in small groups with minimal information technology support. The cost of imposing and monitoring secure data practices in the highly decentralized environment in which social science research is conducted would be very high, and it would have its most serious impact on students and researchers in smaller institutions.

Third, procedures already exist and are widely used by both federal agencies and research organizations to distinguish between data that can be safely released as public-use files and restricted-use files that require additional protections. Restricted-use data are routinely shared under data use agreements similar to those prescribed by HIPAA.

Fourth, several technologies are now available for controlling access to high risk or re-identifiable data. These facilities, such as physical data enclaves and remote execution and analysis systems, create safe environments in which sensitive or re-identifiable data can be analyzed. Providing these facilities for the small proportion of researchers who need to use data that pose informational risks will be much more efficient than trying to impose higher standards of data security on users of data that pose little or no risk.

Fifth, the HIPAA approach to protection of confidentiality is inappropriate and ineffective when applied to data in the social and behavioral sciences. The HIPAA approach, in which the potential for re-identifying subjects is associated with specific types of information, is not feasible in current social research. Complex research designs, like multi-level surveys (e.g. student/teacher/school) and longitudinal studies with repeated interviews, produce data that can be re-identified even when all of the HIPAA identifiers have been removed. As new types of studies are developed, the risk of re-identification must be carefully evaluated on a case by case basis.

Furthermore, HIPAA imposes restrictions that serve no purpose in social research. For example, Common Rule protection ends with the death of the subject, but HIPAA protection does not. Researchers in population studies and epidemiology often use data derived from administrative records on subjects that died more than a century ago. We see no value in applying strict data protection measures to such data.

Recommendations

ICPSR strongly favors the principles described in the 2003 NRC report *Protecting Participants and Facilitating Social and Behavioral Sciences Research*. That committee recommended that OHRP “should establish a new confidentiality protection system for these data. The new system should build upon existing and new data archives and statistical agencies.” (p. 5) We call attention to a few key dimensions of this system, where public-use data with minimal risk to subjects are clearly distinguished from restricted-use datasets with moderate or higher risks.

1. As recommended by the NRC, the determination of which datasets should be public-use and the restrictions imposed on data licensed under a data use agreement should be made by a survey research center, data archive, or disclosure review board, rather than by an IRB. As the ANPRM recognizes, IRBs rarely have expertise in disclosure risk analysis required to evaluate the risk that subjects can be re-identified. In our experience, data producers are strongly motivated to protect subjects’ identities, and they are most likely to err on the side of safety. After all, survey research centers depend on their interactions with subjects, and they are acutely aware that a major breach of confidential data will make it even more difficult to get subjects to respond to their questions. Decisions that designate data as public- or restricted-use should be documented for review by a competent authority.

2. Secondary analysis of de-identified public-use data should not be subject to IRB review. A number of IRBs already have policies allowing researchers to analyze public-use data from designated sources without IRB approval or certification of exemption. For example, the UCLA IRB has determined that:

UCLA investigators’ access to specified “public use” data sets does not constitute research with human subjects (as it does not involve access to identifiable private information about the persons from/about whom the data were collected) and therefore is not subject to UCLA IRB review and approval or Certification of Exemption from UCLA IRB review.

(University of California, Los Angeles, Office for Protection of Research Subjects, Standard Operating Procedures, "Research Involving Public Use Data Files," Policy Number: 42, Date of Last Revision: June 17, 2008, <http://ohrpp.research.ucla.edu/documents/pdf/42.pdf>.)

OHRP should provide clear national guidance on types of data that are outside the scope of IRB regulation.

3. We agree with the ANPRM that subjects should be given “a standard, brief general consent form allowing for broad, future research” when data are collected, and we welcome HHS guidance that would lead to a “standardized general consent form.” However, we note that it has been common practice for IRBs to recommend wording that unnecessarily restricts access to data for secondary analysis. For example, subjects are often informed that their responses will be accessed “only by members of the research team,” which limits the availability of data to other researchers. We believe that protecting the confidentiality of subjects should not preclude analysis by other researchers, and we encourage HHS to offer guidance for providing access to existing data for secondary analysis.

4. High levels of data security should be applied to restricted-use data. Wherever possible, secondary analysis of restricted-use data should take place in secure facilities such as data enclaves and remote execution and analysis systems. The development of these facilities will increase data security and reduce the burden of protecting confidential information security on individual researchers.

5. ICPSR strongly endorses the ANPRM proposal to classify intentional re-identification of subjects who have been promised confidentiality as research misconduct. This is a basic principle that should be a standard for all researchers. We would emphasize that this principle applies only to information provided with a promise of confidentiality, and it should not apply to information that is already public or to responses to interviews in which no confidentiality is promised or implied.

Responses to ANPRM questions:

Question 13.

Measures leading to greater consistency among IRBs will be welcomed by the research community.

Question 14.

We welcome creation of an “Excused” category that would streamline IRB registration of studies involving minimal risk to subjects. We are concerned, however, that including both “Exempt” and minimal risk studies under the new heading of “Excused” will result in intrusive oversight of studies that are outside the scope of IRBs. For example, personally identifiable data derived from public sources (such as newspapers and archives) are exempted from IRB oversight under the Common Rule, but we fear that classifying these studies as “Excused” will create the

incorrect impression that data protection is required for such data. As noted above, we also believe that research with public-use data should not be required to register with an IRB, which is current practice in many places.

Question 16.

The application of these principles will be clearer if distinctions are made (a) between data collection and secondary analysis and (b) between data classified as public-use and restricted-use by a competent body. Data collection involving questions that hold the potential for emotional, reputational, or other harm should be subject to IRB review. Secondary analysis of data that has been deemed public-use should be Excused or Not Regulated. Secondary analysis of data that has been deemed restricted-use (because of the sensitivity of the information or the risk of re-identifying subjects) should be held to research standards that will protect respondents' confidentiality in both data management and publication of results. If clear standards for data protection are available, research with restricted-use data may qualify for Excused review.

Question 20.

In accordance with our view that secondary analysis of public-use data should be "Not Regulated," we believe that "Registered" is more appropriate than "Excused." "Excused" is not sufficiently different from "Exempt" to avoid the confusion created by that term.

Question 23.

We believe that the ANPRM introduces potential for confusion by discussing "data" including survey data and biospecimens under the same heading. Social science data can often be successfully anonymized to allow distribution as public-use files. To the extent that biospecimens may contain DNA, de-identification may not be possible. (On the other hand, quantitative measures derived from biospecimens by various laboratory procedures may be anonymous.)

IRB review and documentation of informed consent should not be required for existing data when:

- subjects are no longer alive
- data are derived from public information, such as public archives
- information was provided without an expectation of confidentiality
- a competent body has judged the data to be suitable for release as public-use.

Question 25.

Some fields (e.g. journalism) should be covered by ethical codes that differ from the Common Rule, because they involve procedures and benefits to society that are quite different from scientific research. However, we also believe that much confusion could be avoided if HHS more clearly distinguished between “Not Regulated” and “Registered” research and abandoned the confusions associated with the usage of “Exempt.” For example, classical literature should be “Not Regulated,” because it fails three tests for research on “human subjects” defined by §46.102(f): (1) The subjects are not alive. (2) There is no interaction with subjects. (3) The information used is identifiable but it is not private.

Question 29.

It may be appropriate for an IRB to perform random audits to assure that researchers correctly understand the scope of IRB review and the meaning of exemptions. A reporting system that assures transparency and identifies activities not required by regulations will be welcome.

Question 46.

The social sciences would benefit greatly from a well-designed process to allow for the analysis of data in ways not anticipated at the time of data collection. Unfortunately, much data has been collected under informed consent statements with language that was unnecessarily limiting, such as “Data will only be shared within the research team.” In most cases, there is no reason to apply these restrictions to data that has been prepared for public use or shared under appropriate data protections.

Question 54.

As explained above, reliance on removal of the 18 specified identifiers in HIPAA Privacy Rules does not provide assurance that subjects cannot be re-identified in most social science data sets. Thus, researchers will continue to rely on “a formal determination by a qualified expert,” as those rules also require. These issues have been discussed by NRC panels, and we endorse their recommendations.

Question 59.

See response to Question 54.

Question 62.

Researchers who collect sensitive information (such as questions about illicit activities) through surveys and other methods are dedicated to the protection of their subjects. We believe that they will feel an obligation to their subjects to

require every user to obtain a data use agreement to assure that every user of restricted-use (or limited) data has agreed to necessary precautions. If this approach is adopted by HHS, the social science research community is likely to continue to operate under stricter standards.

Question 63.

Yes. Intentional re-identification of subjects in de-identified data should be prohibited and designated as research misconduct.

Question 64.

Researchers should be prohibited from disclosing de-identified data to parties who are not subject to the same rules of conduct governing use of the data. The standard practice at ICPSR is to prohibit re-distribution of data, including public-use files. This assures that all those who receive the data agree to our terms of use, which include agreement not to identify subjects.

Question 65.

Researchers should not be required to register with the institution for de-identified data that has been provided for public-use. As we noted above, some IRBs currently identify sources of data that can be used without registration.

References

1. National Research Council. Protecting participants and facilitating social and behavioral sciences research. Washington, D.C.: National Academies Press; 2003.
2. National Research Council. Expanding access to research data: reconciling risks and opportunities. Washington, DC: National Academies Press; 2005.