

# American Statistical Association

*Promoting the Practice and Profession of Statistics*

732 North Washington Street, Alexandria, Virginia 22314 USA  
(703) 684-1221 • Fax: (703) 683-2307 • Email: [asainfo@amstat.org](mailto:asainfo@amstat.org)  
Web site: <http://www.amstat.org/>

October 19, 2011

Jerry Menikoff, M. D., J. D.

OHRP

1101 Wootton Parkway, Suite 200

Rockville, MD 20852

Re: Comments in response to ANPRM for Common Rule

Dear Dr. Menikoff,

The American Statistical Association and its Committee on Privacy and Confidentiality appreciate the opportunity to offer comments on the proposed revisions to the Common Rule. We believe that the Common Rule can be revised to improve protection of data subjects' confidentiality and facilitate access to data for research purposes, and we applaud the efforts of OHRP in reconsidering the Rule. We have appended our comments to this letter about Section V of the ANPRM, which asks for comments on questions related to privacy and confidentiality.

As background, the American Statistical Association (ASA) is the world's largest statistical society, with over 18,000 members in some 90 countries (though most are in the US). One of its core missions is to advise government on matters related to data-centric research and policy-making. The Committee on Privacy and Confidentiality is an appointed group of ASA members with expertise in the technical methods and policy issues related to data access and confidentiality. The list of current and incoming Committee members who endorse the comments in this letter includes:

Jerome Reiter, PhD

Julia Lane, PhD

Tapan Nayak, PhD

Lance Waller, PhD

Simon Woodcock, PhD

Alan Zaslavsky, PhD

Jacob Bournazian, MA, JD

Aleksandra Slavkovic, PhD

Duke University (Committee Chair)

National Science Foundation (Committee Vice-Chair)

George Washington University

Emory University

Simon Fraser University

Harvard University

Energy Information Administration

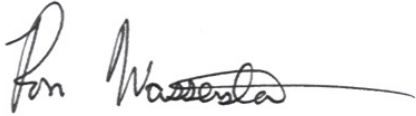
Pennsylvania State University

The Committee, and the ASA more broadly, would be delighted to share our expertise with OHRP as it considers revisions to the Common Rule.

Thank you,

A handwritten signature in black ink that reads "Jerome Reiter". The letters are cursive and fluid.

Jerome Reiter  
Chair, ASA Committee on Privacy and Confidentiality

A handwritten signature in black ink that reads "Ron Wasserstein". The signature is cursive and includes a long horizontal flourish at the end.

Ron Wasserstein  
Executive Director, ASA

**Subsection A, General comments:**

- We applaud the specific attention given to informational risks in the Notice. Informational risks have different consequences, methods of evaluation, and methods of abatement than other types of potential harms from research, and addressing these specifically will improve the management of informational risks.
- We believe that most researchers, if properly trained and equipped, can be motivated to protect research subjects by preventing improper disclosure. The approach taken to compliance with the rules should first emphasize education and adequacy of resources, which should be backed up with oversight procedures including audits and penalties.
- The overall level of risk to research subjects imposed by research use of data is a function of the content of the data (including the variables included, the sensitivity of the information, and the characteristics of the sample or population covered by the dataset), its relationship with other available exogenous sources of data, and the controls established on access to the data (technical and institutional). Given that the informational risk of research involves a combination of these factors, development of a data security plan should also consider the effects on research utility of the data and the resulting impact on the usefulness of the research, within the range of options that provide adequate privacy protection. (For example, if legitimate research objectives require use of sensitive or potentially identifying data elements, this might be made acceptable by maintaining a level of data security commensurate with the content of the dataset.) While we understand the necessity of separating these factors in drafting the ANPRM, nonetheless a framework is needed for identifying combinations of the factors that correspond to levels of risk acceptable at the various levels of review that might be established in a revised Rule. Ultimately this framework should be embodied in a document that provides adequate guidance to the work of human studies review boards, drafted by a combination of agency and external advisors. Our committee would be pleased to contribute to this process.
- Specialized expertise is required for evaluation of informational risks. Since most Institutional Review Boards most likely do not possess this expertise, some structures must be developed to facilitate review of informational risk. This might be some combination of creation of written guidelines to support triage of simple cases, development of expertise local to the research institution, and referral of proposals to qualified external efforts (for whom certification might be required) when evaluation of informational risks requires additional expertise. A higher level of expertise on nondisclosure will be required when access to the data is less restricted, as in dissemination of public-use datasets.

- Training requirements for human studies certification should include adequate instruction on the nature of informational risks and the rationale for and content of rules governing management of restricted data. Content standards for such training should be developed. Our committee would be pleased to participate in this area as well.
- Sanctions should be established for improper disclosure of data subject to restrictions, especially when this is done for pecuniary gain or with disregard for potential harms to research subjects. Failure to provide adequate training, resources and oversight leading to disclosure may also entail sanctions and should be addressed with a remediation plan at the appropriate level (institution or research group, as the case may be).
- Any revision of the Common Rule should be cognizant of requirements imposed on researchers by funding agencies for release of public-use versions of research data, and should include suitable provisions to ensure that these requirements do not conflict with appropriate Common Rule nondisclosure obligations.
- A revised rule should be applied to ongoing research, with appropriate phase-in respecting existing data use agreements (DUAs).

**Subsection A comments, Question 54:**

- We note that the HIPAA Privacy Rule is directed in the first instance at the Covered Entity which is the provider of the data, while the Common Rule is directed at the researcher who may be the recipient of data from a HIPAA Covered Entity. Thus the changes suggested by the Notice involve substantial changes even for researchers using HIPAA-covered data.
- We support the extension of standards defining levels of de-identification – the objective of the HIPAA Privacy Rule’s standards – to personally identifiable data in domains other than health. The lines between health data and social or behavioral data are not at all sharp; indeed these categories overlap substantially. While different issues arise from different datasets, these differences are not primarily related to these broad categories. However, the generic approach taken by HIPAA, which does not distinguish the potential harms to the research subject of disclosure of various data elements, is problematic for research use of health data and even less appropriate when applied to the still broader range of data included in social and behavior research.
- We find serious deficiencies in the approach to and standards of de-identification in the HIPAA Privacy Rule and therefore cannot support extension of the current Rule to data in non-health areas of research. The current Rule relies on two elements – expert certification of an acceptable level of risk and exclusion of a specific list of identifiers from the dataset – neither of which is adequately specified to meet the demands that will be placed on the Rule as guidance for protection against informational risks.

- The provisions for expert certification of nondisclosure do not indicate the nature of the required expertise or the means by which it may be certified, nor do they indicate what standards of nondisclosure should be applied. Guidelines should be developed outlining the required qualifications. Funding should be provided to foster development of centers of expertise on nondisclosure that can provide training on nondisclosure and thereby develop the required capacity. IRB review should assess the specification of expertise used commensurate with sensitivity and use of the data.
- The “Safe Harbor” provision for removal of 18 specific identifiers addresses only one of the factors determining the informational risk of a dataset. Deleting these identifiers might in many cases still leave an unacceptable level of risk, depending on other factors such as the characteristics of the population or sample covered, the coverage of the dataset, other possibly identifying information contained in the data, the potential harms of disclosure and the degree of motivation of a potential intruder. In other cases, this level of disclosure protection might be adequate.
- Even if IRBs do not possess technical skills in assessing the risk of disclosure, they are generally qualified to assess the potential harm to research subjects of disclosure (that is, the sensitivity) of particular data elements, and it is reasonable to include assessing sensitivity of the information within their scope of review.
- IRBs should be asked to assess the impact of any restrictions they may impose to prevent disclosure on the quality of the research that can be undertaken.

**Subsection A comments, Question 55:**

- As noted in the response to Q54, we reject the assumption that a de-identification standard can be framed solely by listing identifiers to be deleted from the data.
- Nonetheless we strongly support the principle that de-identification standards should be reviewed annually and updated when needed to remain current with developments that affect informational risk such as availability of new external sources of data, advances in computing methods, and research findings that support reassessment of risk. The occurrence of rare events such as worker accidental deaths, transportation crashes, and epidemic diseases creates challenges for data stewards to maintain the same of risk of re-identification in an information release because of extensive media coverage and other exogenous sources of information that become public. These rare events are triggering events that should be monitored and considered along with developments of new methodology and technology when reviewing de-identification standards on an annual basis. We note, however, that rare events also can include important information for policy, and we support methodologies that promote analysis of those events.

- A suitable mechanism for such a review might consist of a committee of experts from within and outside federal government, meeting at least annually, and a staff equipped to support the committee's work and to receive confidential communications concerning incidents of informational harms. More guidance documents need to be developed for public use so that users of disclosure limitation methodologies are aware of the benefits and limitations of each method.
- Suitable channels of communication should be established to facilitate rapid communication of changes in standards to researchers, consulting experts, and relevant institutions.

**Subsection A comments, Question 56, 57:**

- No comment.

**Subsection B general comments:**

- Even if most breaches of data security are due to lapses in basic computer security and system management, we still maintain that controlling these lapses involves training of staff with access to data and establishment of appropriate procedures for data management. The emphasis on technical aspects of computer security seems somewhat one-sided if in fact these lapses involve errors such as carrying data on unsecured media, sharing passwords, or mixing sensitive identifiable data with otherwise insensitive data in analytical files. Minimal standards of training should be established to guard against these errors.
- We question whether retrospective audits are a suitable mechanism for enforcing data security across the many institutions and research groups that are involved in covered research. Who would conduct these audits, and what number of them would be feasible with available resources? Audits by a federal agency might be useful as an information-gathering exercise to assess the overall competence of data collection, but are unlikely to be intensive enough to serve as an adequate mechanism of enforcement or system improvement. The retrospective audit is not an adequate safeguard against current and future violations and should be used in conjunction with other controls and sanctions for a violation. For example, if a researcher violates his/her obligations to safeguard sensitive information and causes an unauthorized disclosure, then the authorizing institution should hold an ethics hearing with written findings and future restrictions on that researcher's access to sensitive data for defined time periods.
- We recommend that attention be paid to the substantial advances made in protecting confidential financial and national security information, funded by, among others, the National Science Foundation and the Department of Defense.

- Implementation of data security requirements might be best conducted by relying on the institutions sponsoring research, tiered through the appropriate subunits. For example, an institution might establish its own standards for system management and training for groups (departments, labs, or research teams) handling data at various levels of sensitivity. It could certify such groups as having established adequate safeguards to handle data up to a certain level of sensitivity. External audit then could focus on the adequacy of the institutional procedures rather than the individual research projects. (Question 66 seems to assume this approach although it is not explicit in the narrative text.)

**Subsection B comments, Question 58:**

- If data security is regarded as largely an institutional function as suggested in the general comments above, the data security and information protection standards should apply to all data and biospecimens now held, not only new data collections. Researchers should certify compliance with these new standards in any research proposal after the standards take effect.

**Subsection B comments, Question 59:**

- The HIPAA rules are primarily designed to assure the security of data being transmitted for healthcare operations and billing. While the rules specify the categories of data that might be released from the healthcare system to researchers, they do not deal with how those data are protected once they are released. We do not think that rules designed for operational data management will necessarily translate well to the much different and more diverse world of research.
- The relevant categories for establishment of data security standards involve the sensitivity of the data, not whether they are health data. As noted above, there is no sharp line between health data and social/behavioral data.

**Subsection B comments, Question 60:**

- General guidance and standards for data security should address the special issues arising during field collection of sensitive data, during which some controls applicable to analytic data files are not feasible and identifying information may inevitably be linked to sensitive responses.

**Subsection B comments, Question 61:**

- The NIST standards, designed for the types of systems used in federal government, are too prescriptive to be generally usable across the more diverse IT environments used across the non-federal research enterprise.
- A system of levels of data security should be defined in more general terms, focusing on requirements rather than specific procedures.
- A process should be established through which institutions could develop their own data security plans consistent with these requirements. Voluntary development and dissemination of a few standardized plans appropriate to the resources of various research centers should be encouraged.

**Subsection B comments, Question 62:**

- Data Use Agreements (DUAs) are an important tool for ensuring compliance with data protection standards. Limited data sets carry some risk of re-identification, and the use of a DUA provides a flexible mechanism for controlling the risk of re-identification by controlling not only the limited data set and who has access but other data files that may be used in the research project.
- Training of researchers is another important tool for ensuring compliance with data protection standards.
- Both approaches can be used in combination to provide access to higher quality data than simply relying on technical means alone.

**Subsection B comments, Question 63:**

- In general, a prohibition on re-identification of de-identified datasets would be a reasonable requirement. Exceptions could be allowed to permit adding essential content to the dataset. When permitted under the applicable DUAs and part of a research plan approved by an IRB, the revised rule should allow for linking datasets by using variables that might uniquely identify individuals. However, such re-identification should only be temporary and limited to specific research purposes. The security levels for management of datasets that have been temporarily re-identified in this manner and for the final linked datasets should be appropriate to their content, taking into account any identifiers that are temporarily attached, the potential for increased disclosure risk of any added characteristics, and the possibly greater sensitivity of added data elements.

**Subsection B comments, Question 64:**



ASA Committee on Privacy and Confidentiality, comment on ANPRM, 19 October 2011.

- Obviously the prohibition on re-identification, and indeed any Common Rule protection against informational risk, is useless if the researcher is allowed to redistribute data to users who are not under Common Rule restrictions, especially since DUAs only cover the researcher and would not apply to third parties. Training will help ensure that researchers understand the strictures against redistribution.

**Subsection B comments, Question 65:**

- This seems like an appropriate safeguard for Excused research on sensitive data, and could be linked to the institution's local procedures for assessing adequacy of data security. The registration should require updated contact information and re-certification of the terms and conditions on an annual basis.

**Subsection B comments, Question 66:**

- Institutional Review Boards should have the authority to delegate this responsibility to appropriate entities possessing the authority and technical resources to conduct audits.