COMMENT ON PROPOSED CHANGES TO OMB CIRCULAR A110, SECTION §.36

THE COMMITTEE ON PRIVACY AND CONFIDENTIALITY

The Committee on Privacy and Confidentiality of the American Statistical Association recommends additions to the draft language in *Federal Register* (64/23 5684-5685 February 14, 1999).

At the outset we wish to make clear that we strongly support the principle that the public should have access to data sufficient to allow independent estimation and scientific criticism of published findings from research funded with Federal grants. However, a particularly damaging effect of some forms of access would be to violate promises already given to respondents to maintain confidentiality for their answers. We believe careful legal drafting of terms in the proposed rule will minimize potential damage from the new rule.

Our concerns relate to ambiguity in the terms *data*, *reproduce .. and otherwise use data*, and *publication.*

The current language would make it possible for persons (and organizations) to require that the grant-receiving organization disclose identities of parties who supplied data (respondents, patients, or organizations). The language could subvert established practice to protect the rights of human subjects. Responsible researchers strongly oppose and resist any efforts to force release of information for identifiable individuals or other research subjects[(1)]

These undesirable consequences arise because identifiable data collected by most recipients of Federal grants are not records exempted from FOIA. They do not have protections afforded by statute to persons and organizations who supply data to the Bureau of the Census, the Bureau of Labor Statistics, or the National Center for Health Statistics. Nor do the researchers necessarily collect data that is considered a "sensitive record" and exempted under FOIA.

Standard practice of scientific professionals calls for informed consent and non-disclosure of individual or organizational identities. Data are collected from persons and organizations who agree to provide information for the uses described by the researcher. Researchers promise to use data in ways that do not reveal the identity of the data supplier. Scientific data are shared in ways that limit the end product of study to statistical presentations. Those presentations do not disclose, directly or by inference, facts about individuals or organizations that are not public.

For these reasons we urge the OMB to make the language of *A110, Section §.36* much more specific.

Specific Recommendations

1. *Data* should be carefully defined in the circular.

a. *Scope of the data* to be included should be defined. When the data are not directly and materially related to the language of the regulation, they should not come under the control of FOIA.
b. *Data* has many meanings. Information about particular individuals or organizations, an estimate of a statistic computed from probability samples, and an aggregate obtained from public documents are all referred to as data.
c. Data to be covered by the rule should be limited to micro-data, estimates, and aggregates for which the risk of disclosing individual identities is small. "Small" should be defined by standards that are deemed acceptable for non-disclosing release in major Federal statistical agencies.

2. *Reproducing .. and otherwise using data* should be limited to procedures that retain the integrity of data collectors' promises of confidentiality to their respondents.
3. *Publication of findings* should also be denoted by a positive enumeration.
    a. Acceptance of papers by scientific journals and the incorporation of a report (as an appendix or a reference) in a proposed rule should constitute publication.
    b. Circulation of preliminary scientific work prior to peer review should be excluded from the concept of publication, even though multiple copies or web-based copies are available.
    c. The use of any report by a regulatory agency in rule-making or publication in peer-reviewed journals creates a presumption that underlying data can be accessed by the public. Third parties need to be able to replicate estimates, and challenge the validity of assertions supported by the data. However, this testing and replication can not require reidentification of data suppliers or other activity that compromises the consent given.
4. Cross-references. Section *§*.36 needs to cross reference statutes and rules that protect the confidentiality of research records, interviews, and other data collected from organizations and human subjects. The cross references should be exhaustive, as various degrees of protection apply to seemingly similar data collections. These cross-references are needed to assure that protection of confidential material is the general intent of the new rule".

General Comments

The National Academy of Sciences is forceful in supporting data sharing as an important positive goal for science (S.E. Fienberg, et al. 1985. *Sharing Research Data*; and G.T. Duncan, et al. *Private Lives and Public Policies).* The goal of minimizing harm to data suppliers while achieving release of information required for good social policy is supported in these works. Use of data is reserved to statistical analysis unless the providers give consent.

> The proposed rule does not adequately implement balance between need for social information and harm to data suppliers.

The scientific community and the Federal grant-giving agencies have already implemented access to data for scientific purposes. Those modes of access should be seen as the primary mechanism for openness in the process of Federal rule-making.

> The proposed rule does not indicate that access to data through established channels is the primary mechanism that should inform debate on proposed rules.

When the OMB rule is final, the research community will have two options: To plan for release of the kind of data enumerated above, or to avoid the scope of the ruling by funding research without Federal grants and in some instances censoring viable research topics because the impacts on data quality and researcher time are seen as extremely deleterious.

> The consequences of the proposed rule will be a decrease in the quality and quantity of research with corresponding losses in the future. Minimizing this negative impact is an important objective for redrafting.

The kind of research that will be inhibited will be research that uses confidential records, matching with confidential records, interviews in which illegal or sensitive matters are discussed, and a variety of research activities in which analysis of data by persons other than the collecting team is viewed as invasion of privacy, or risking the release of proprietary knowledge.

Conflicts between state open records legislation and the new rule may make it impossible for researchers to gain access to assemble valuable information that is needed for epidemiology, criminology, economic, and demographic research. The problems of assembling information on families and absent parent indicate the need to assure data suppliers that the results will be used for research and not for litigation.

In summary, the scope of what must be disclosed through FOIA should be narrowed, the meaning of data, publication, and reproduction should be carefully constructed. Language used should be guided by the knowledge of data librarians who control access to datasets generated under Federal grant programs. The policies of Federal agencies for data-sharing, such as the National Science Foundation and major private data libraries should guide releases under the new rule. The rule should cross-reference Federal statutes that protect privacy and confidential data from disclosure.

> 1. Procedures recommended by the Council of American Survey Research Organizations (www.casro.org):
>
> "The use of survey results in a legal proceeding does not relieve the Survey Research Organization of its ethical obligation to maintain in confidence all Respondent-identifiable information or lessen the importance of Respondent anonymity. Consequently, Survey Research firms confronted with a subpoena or other legal process requesting the disclosure of Respondent-identifiable information should take all reasonable steps to oppose such requests, including informing the court or other decision-maker involved of the factors justifying

confidentiality and Respondent anonymity and interposing all appropriate defenses to the request for disclosure."