Comment On
Second notice of Proposed Chances To OMB Circular A110
Section ____.36

THE COMMITTEE ON PRIVACY AND CONFIDENTIALITY

AMERICAN STATISTICAL ASSOCIATION

Boldface is used for language that we propose be incorporated into the rule.

RECOMMENDATIONS

1. Explicit language about disclosure risk must appear in the rule. We propose adding to the rule the language that appears below in items a, b, and c.

Scope of data and manner of release that may be employed under this rule is governed by three preemptive principles:

a. Disclosure risks for data supplied under this rule must be no greater than the risks attaching to data collected by the federal government where confidentiality of responses is guaranteed to the data suppliers.

Discussion:

The proposed language fails to provide a workable standard for what constitutes a disclosure risk, nor does it provide a process for access to data that must be restricted to avoid disclosure risk. These omissions leave the rule with, at most, partial and, very likely, ambiguous protection under the Privacy Act. In addition, the censoring of individual records is informative so that problem cases can not simply be deleted from a data file.

Comments by principal investigators who have commissioned surveys under federal grants, directors of university-based survey organizations and state agencies who arrange access to confidential records by grant-funded university-based researchers is consistent - risk of abusive access to data under the proposed A-110 __ .36 will (i) discourage proposals from universities to undertake surveys using federal grant resources and (ii) discourage many survey research organizations from using current best ethical survey techniques that promise confidentiality of responses. This will result in inability to collect certain kinds of data from certain kinds of respondents, degrading the survey research tool for all scientific research.

The inability of reputable private survey organizations to promise confidentiality for their respondents will be widely perceived as a liability by the public at large. Public distaste for survey research will degrade response to all surveys including government voluntary and mandatory surveys. Further, if the confidentiality of responses is under question for surveys financed by federal grants, then the response rates of other nonfederally funded surveys made by the same organization may be reduced. The public does not strongly differentiate among

alternative survey practices but has a broad cognitive understanding of the reputation of survey organizations in general.

b. The scope of data to which A-110 __ .36 applies is limited to the variables, the time period, and subpopulations that are referenced in the scientific work published.

Discussion:

i. Many data collections extend for a long period and obtain diverse measurements on the population studied during that period. Orderly scientific inquiry will be disrupted and possibly obstructed if data other than the empirical support for the finding published (or used in rule making) are released through the FOIA procedure.

Additional considerations such as those given below make it essential to define the scope of a request in the proposed rule.

ii. It is almost impossible to avoid all disclosure risks with a panel of data that deals with human populations. Restricted access to such data is the preferred scientific procedure in this instance.

iii. The scope of large datasets implies that independent lines of research and phased research objectives are required to mine the data for their scientific content.

c. Rules that affect disclosure of scientific data collection by the federal government statistical agencies should be cross-referenced and explained in the body of rule A-110 __ .36.

Discussion:

The federal government has strong legislative protection for its administrative records and data (viz., Revenue Act of 1976, Title XIII of the Census Act). Federal government agencies have standards for releasing for public use data that have minimal risks. Finally they have mechanisms for giving restricted access to data with disclosure risks. Each of these areas of law and rules should be cross-referenced so that applicants for data under the rule are cognizant of the kind of protections that will need to be in force in any release.

2. Primary data should be the target of the regulation. We propose adding the following language to the proposed rule:

a. This regulation applies only to such primary data collectors, individuals or consortia, who are responsible for direct measurement, quality control in measurement, and control of sample and experimental designs.

b. Where scientific data are already documented and disseminated through a data archive, principal investigators may demonstrate compliance with A-110 __ .36 by citing availability of the data through the data archiving organization.

c. Investigators who pursue analysis on preexisting data and undertake no "measurement" are responsible for citing primary data sources and providing computational algorithms used to manipulate those data.

Discussion:

The proposed rule need not create new mechanisms for the release of data when that release has already been anticipated by the grantees. Existing mechanisms for securing access to data should be used in place of FOIA, if those mechanisms have already been established by private contracts between grantees and data archiving organizations. Grantees should be able to obtain a waiver for their legal responsibility when the intent of the rule is met by a non-FOIA procedure.

3. Cost reimbursement. We propose adding the following language to the proposed rule:

Cost reimbursement entails two parts, (i) financing of planning for data release and (ii) additional costs precipitated by individual requests under the proposed rule. This will entail a two-part charge for requests under this rule.

(i) Financing for planning will be required for every grantee after the effective date for this rule. OMB should request sum-sufficient appropriations from the Congress to finance a revolving fund for compliance with this rule. Funds obtained will be distributed to grantees at the time of awards for activities relevant to potential future dissemination of data under this rule.

(ii) At the time of a request for data, requesters should be asked to place funds in escrow sufficient to pay (1) a pro rata share of the revolving fund, (2) the cost of establishing the $100 million economic impact estimate, and (3) out-of-pocket costs for the grantee (or organizations with whom the grantee had previously contracted) for the dissemination of data.

Discussion:

No grantee will be immune from the risk of being required to supply data under this rule. Therefore, immediate financing is needed for planning compliance with this rule. That financing cannot come from the grantees or their host institutions that have not been funded for this activity.

Scientific data collection is decentralized, and the control of data rests with thousands of independent investigators. Those investigators typically are not funded adequately and are often forced to terminate lines of investigation for lack of financial resources. The rule being proposed effectively is asking for a large number of "bankrupt investigators" to provide working capital required to attend to the provision of data.

The problem of resource availability is compounded by the continuing shift of investigators from one institution to another and a corresponding flux in research teams that makes it difficult to identify persons responsible for data.

It appears clear that additional resources must be provided from the effective date of this rule if compliance is to be obtained. To ask grantees to absorb the costs of the planning required is to create litigation, delay, obfuscation of responsible parties, unnecessary disruption of scientific endeavor, and a substantial diminution of U.S. leadership in scientific investigation.