The following are the steps used by the Substance Abuse and Mental Health Data Archive (SAMHDA) at ICPSR in conducting disclosure analysis.

1. Data sets identified as needing review undergo preliminary review by a senior data processor, supplemented by meetings with the project director and other staff. The preliminary review includes, but is not limited to, the items below. The review process is iterative.

   a. Setup needed documents for the review, and later for study release
      i. Ensure all documentation for study has been sent by PI or agency
      ii. Coordinate with ICPSR staff and PI to obtain bibliographic citations
      iii. Review journal articles, particularly those that discuss sampling
      iv. Summarize study:
         1. Background
         2. Purpose
         3. Methods
            a. Survey instrument
            b. Variable groups, types of variables to describe study;
               specific examples of variables that might pose disclosure risk
            c. Sampling design
            d. The type of data collection, e.g. longitudinal (panel or time series), cross sectional
            e. Weighting
   b. Clean data as necessary (e.g., add labels) to produce usable frequencies
   c. Concurrent with (a) and (b), identify and document potential problem variables, matching databases, or other disclosure risks for the study
   d. Contact PI as necessary to complete needed information
   e. Draft a report for the initial Disclosure Review Committee (DRC) meeting; the report includes the documentation specified in (a (iv)) and (c), outlines possible disclosure protection methods, and lists discussion items for the meeting.

---

1 Deb Schild of ICPSR assisted with the first draft.
2. The DRC reviews the SAMHDA report, discusses the identified disclosure risks, as well as any additional risks. The committee then discusses solutions to address the risk, keeping in mind the balance between disclosure risk protection and analytic utility. In general, the committee addresses the following issues, as well as others that may be unique to a particular study.
   a. Do the data pose unacceptable disclosure risk for a public-use file?
      i. If the committee determines the risk is acceptable for a public release, the data are processed according to SAMHDA and ICPSR standards, including the basic risk reduction methods, such as renumbering cases, removing direct identifiers, removing or recoding specific dates, etc.
      ii. If the committee determines that the disclosure risk is unacceptable, the review continues.
   b. Is additional information required? Risk may be further defined or assessed based on the initial document prepared by staff. Additional analyses may be run or potentially matching databases investigated.
   c. What actions can be taken to reduce risk? Several approaches may be taken depending on the nature of the data and the risk (e.g., coarsening the data, introducing noise). In rare instances, and depending on what has been published, including earlier versions of the data, the committee may determine that a safe public-use file may not be possible.

3. Next steps.
   a. Staff continues disclosure analysis and/or gathers more information, as required.
   b. Staff applies agreed-upon disclosure risk protection techniques and assesses results.
   c. The committee is reconvened to discuss new information and/or results.

4. Staff documents DRC meetings.
   a. Meeting summaries are written and circulated to all who attended each meeting.
   b. Summaries are revised as necessary and finalized.

5. Staff produces a final report summarizing the disclosure analysis. Committee members may write some sections of the report. The DRC reads and comments on report drafts before the final report is sent to SAMHSA. The report takes the following format.
   a. Study summary and purpose
   b. Purpose of the disclosure analysis (e.g., source of the request)
   c. Description of the files and/or data
d. Summary of disclosure risk

e. Existing protections in the data file (disclosure protection techniques applied as well as factors that add protection, such as recall error, missing data, imputation)

f. Recommended disclosure protection approach and rationale

g. Documentation of additional (other) techniques discussed and rejected

h. Summary of the actions taken and how these actions reduce risk, including analytic proof of risk reduction

i. Analysis and proof of analytic impact to the data

Steps (a) – (g) are typically put into a preliminary report, which is sent to SAMHSA as a disclosure protection plan for approval. The final report includes (a) – (i).

6. If new information about the study arises and/or the first disclosure protection techniques do not sufficiently reduce risk, analysis is resumed until risk reduction and analytic utility of the data are acceptable.