Purpose of Matching

Matching of the survey data against publicly available databases has been the original and primary NCES statistical standard to ensure data confidentiality. Other confidentiality tools have been added and the matching approach has undergone major changes over time based on newer technologies and statistical approaches (see history of matching).

Matching provides the clearest basis for determining whether data intended for dissemination present a disclosure risk. When performed properly, schools identified as disclosure risk are masked sufficiently to eliminate the disclosure risk.

Currently, the required disclosure analyses for public release data focuses on the two DRB-required components for the confidentialization of data. They include (1) the identification and masking of potential at-risk schools by comparing the study variables with CCD and PSS data, using **probabilistic matching and deterministic swapping** on school level data, and (2) starting in 1999, the implementation of an additional measure of uncertainty of school, student, and teacher identification with the **controlled random swapping** of data elements within all levels of the school data (e.g., teacher, student, school, and principal files). The design and procedures for the identification of disclosure-risk schools and the masking of these data should follow the procedures conducted and approved by NCES for the most of their studies. Some additional statistical confidentiality measures include the suppression or collapsing of variables that may be identifying.

Both deterministic swapping and random swapping are best conducted in a way that maintains the overall breakouts and statistics. Close partners are used for the swapping to minimize any data distortion or noise.

NCES Statistical Standards 4-2 describe the requirement for data collection and dissemination to maintain confidentiality of the respondents. The full description of the current standards can be found at:

https://nces.ed.gov/statprog/2012/pdf/Chapter4.pdf

NCES STANDARD: 4-2

PURPOSE: To protect the confidentiality of NCES data that contain information about individuals (individually identifiable information). For this reason, staff must be cognizant of the requirements of the law and must monitor the confidentiality of individually identifiable information in their daily activities and in the release of information to the public.

STANDARD 4-2-8: For public-use data files, NCES minimizes the possibility of a user matching outliers or unique cases on the file with external (or auxiliary) data sources. Because public-use files allow direct access to individual records, perturbation and coarsening disclosure limitation techniques may both be required. The perturbation disclosure limitation techniques by definition, include the techniques applied in a confidentiality edit (if one is performed) and may include additional perturbation disclosure limitation techniques as well.

Methods for Protecting Individually Identifiable Data

Type of Protection	Methods	
	Perturbation	**Coarsening**
Confidentiality Edit	Yes	Yes
Disclosure Limitation Techniques	Yes	Yes

All public-use files (i.e., the edited restricted-use files) that contain any potentially individually identifiable information must undergo a disclosure risk analysis in preparation for release to the public. The steps are as follows:

1. At an early stage in designing and conducting this analysis, staff must consult the Disclosure Review Board (DRB) for guidance on disclosure risk analysis and on the use of NCES disclosure risk software. Any modifications that are necessary as a result of the analysis must be made, and the entire process must be documented.

2. The documentation of the disclosure risk analysis must be submitted to the DRB. The documentation must include descriptions of the risk of disclosure and the types of edits used to avoid disclosure. Decisions over the type of confidentiality edits must take into account the procedures needed to avoid disclosure of individually identifiable information, age of the data, accessibility of external files, detail and specificity of the data, and reliability and completeness of any external files. The documentation should also include the results demonstrating the disclosure risk after adjustments to the data.

3. The DRB will review the disclosure risk analysis report and make a recommendation to the Commissioner of NCES about the file release.

4. The Commissioner then rules on the release of the data file.