

MiCDA Enclave Rules for Pre-Export Disclosure Review

Introduction

A contractual obligation of researchers who qualify for access to restricted data from the Health and Retirement Study is to maintain respondent anonymity. This document is designed to assist those researchers in meeting this requirement by providing them with guidelines for implementing their own disclosure limitation review process.

Goals of the Disclosure Limitation Review Process

- Prevent disclosure of confidential information
- Reduce the likelihood of respondent re-identification
- Provide useful data resources to researchers
- Ensure that the results of the review process are acceptable to both the researcher and the provider(s) of the restricted data.

Methods Used to Protect Confidentiality: Analysis Inputs

- All HRS microdata products (public and restricted) held by the MiCDA Data Enclave are directly or indirectly based on sample survey methodology.¹
- Public file variables containing indirect identifiers such as industry, occupation, and geographic information have been bracketed.²
- Microdata files derived from SSA administrative data (W2-SE Income and Earnings-Benefits Data) have been subjected to rounding and top-coding in accordance with the governing Memorandum of Understanding.³
- Direct respondent identifiers such as name, address, SSN, Medicare/Medicaid identifier, place of birth, etc. have been removed from all public microdata products, and limitations have been placed on access to geographic detail information⁴
- Data items at the respondent level related to sample design, such as PSU, segment, and line, are not available to users.

Methods Used to Protect Confidentiality: Analysis Output

- Data Enclave users may only export statistical summary information (frequency tabulations, magnitude tabulations, means, variances, regression coefficients, and correlation coefficients) that does not permit the identification of any individual person, family, household, employer or benefit provider.
- Export of microdata files or analysis output containing information at the respondent level is not allowed unless specific permission has been obtained from the Michigan Center on the Demography of Aging and all relevant restricted data providers.
- Users may not remove any analysis output that can potentially identify respondents, sampling information, or geographic areas below the level of Census Region, either directly or inferentially. Under certain circumstances Data Enclave Users with access to state-level geographic information may wish to report state-level summary information. In such cases, analysis results must be submitted to the Enclave Data Confidentiality Committee and the dataset provider for review and approval prior to export.
- Tabulations may be exported, but are subject to the following rules:
 - Magnitude Data: Ensure that no cells/strata with $N < 3$ are produced.⁵
 - Frequency Data: Apply a marginal threshold of $N \geq 5$ and cell threshold of $N \geq 3$ to all tabulations.⁶
- Cross-Category merges (e.g., State with Earning and Benefits Data) are only allowed under special circumstances (see Merge Rule Cross-Reference Table, below).
- Analysis results containing merged area data based on geographic information may be exported if there is no direct identification of geographic areas, or if geographic areas are reported using the same grouping characteristics as public files, or if special approval has been granted by the Enclave Data Confidentiality

¹ *Report on Statistical Disclosure Limitation Methodology (Working Paper 22)*, Federal Committee on Statistical Methodology, Office of Management and Budget, May 1994, Chapter II.C.1.

² *ibid.*, Chapter II.C.2

³ *ibid.*, Chapter II.E.2.a

⁴ *ibid.*, Chapter II.E.1

⁵ *ibid.*, Chapter III.B.1, *Summary of Agency Practices*.

⁶ *ibid.*, Chapter III.B.1, *Summary of Agency Practices*.

Committee. Any analysis results that directly or indirectly identify geographic areas below the Census Region level as a row or column heading may not be exported. Users are reminded that geographic information may not be used in conjunction with files derived from Social Security Administration administrative data without written permission from the Social Security Administration. Users with questions about the use of geographic information may wish to contact MicCDA Data Enclave technical staff for further details.

- High visibility variables such as certain Cause of Death and Medical Condition codes must be recoded or collapsed before being exported⁷
- Disclosure review rules for Class 3 merges (see Merge Rule Cross-Reference Table, below) are based on negotiations among HRS, restricted data provider(s) and researcher using the review principles stated above as a starting point.
- All analysis output is subject to disclosure review by Enclave staff members who, in consultation with restricted data providers, have ultimate authority over whether a given set of analysis results may be exported.

Merge Rule Cross-Reference Table

	HRS Public Data	HRS Restricted Geographic Information	HRS Restricted Industry and Occupation	SSA Administrative Data Sets	Medicare Summary and Detail Records	National Death Index	Other HRS Restricted Data Sets
HRS Public Data	Unrestricted	Class 1	Class 1	Class 2	Class 2	Class 2	Class 1
Geographic Information	Class 1	n.a.	Class 1	Class 3	Class 2	Class 2	Class 1
Industry and Occupation	Class 1	Class 1	n.a.	Class 2	Class 2	Class 2	Class 1
SSA Administrative Data Sets⁸	Class 2	Class 3	Class 2	n.a.	Class 3	Class 2	Class 2
Medicare Summary and Detail Records	Class 2	Class 2	Class 2	Class 3	n.a.	Class 2	Class 2
National Death Index	Class 2	Class 2	Class 2	Class 2	Class 2	n.a.	Class 2
Other HRS Restricted Data Sets	Class 1	Class 1	Class 1	Class 2	Class 2	Class 2	n.a.
<p>Unrestricted: Public data sets (including Sensitive Health data products) provided by the Health and Retirement Study that may be merged with any restricted data set. Sensitive Health Data products are provided to researchers under terms of a data use agreement signed by the researcher and HRS.</p> <p>Class 1: Restricted data sets provided by the Health and Retirement Study for merging with HRS public data sets under terms of HRS Data Confidentiality Rules.</p> <p>Class 2: Restricted data sets provided to HRS for merging with HRS public data sets under terms of a negotiated MOU or DUA</p> <p>Class 3: Special case – merging procedures and disclosure review rules are based on negotiations among HRS, restricted data provider(s) and researcher.</p>							

⁷ *ibid.*, Chapter II.E.2

⁸ HRS cohort (respondents): Covered Earnings, RDSI Benefits, SSI Benefits, Summary of Earnings and Projected Benefits file.
 AHEAD cohort (respondents and deceased spouses): Covered Earnings, RDSI Benefits, SSI Benefits, and Wage and Self-Employment Income.
 CODA/War Baby cohorts (respondents and deceased spouses): Covered Earnings, RDSI Benefits, Payment History, and Wage and Self-Employment Income.