



CJARS Virtual Data Enclave Disclosure Avoidance Handbook

Contents

1. Importance of Disclosure Review	3
2. Requesting Disclosure Review	3
2.1 How to Request Disclosure Review.....	3
2.2 Document Formatting Requirements of Research Output.....	3
2.3 Supporting Documentation Requirements.....	4
2.4 Review of Graphical Output and Code.....	5
2.5 Other Non-standard Review Requests.....	5
3. Disclosure Avoidance Requirements.....	6
3.1 Cell Size and Suppression.....	6
3.2 Top and Bottom Coding.....	8
3.3 Summary Statistics	9
3.4 Saturated Models.....	9
3.5 Geographic Aggregation.....	10

1. Importance of Disclosure Review

Researchers with access to the Criminal Justice Administrative Records System (CJARS) Virtual Data Enclave (VDE) are contractually obligated to protect sensitive information on the VDE. This handbook describes the requirements that must be followed to receive approval for research output to be removed from the CJARS VDE. Under no circumstances are CJARS VDE users permitted to remove or transcribe information from the CJARS VDE. Only research output reviewed and approved by CJARS VDE staff can be removed from the data environment. CJARS VDE staff have the sole authority to remove research output from the data environment.

This handbook defines the requirements that research output must meet to receive disclosure approval. The requirements are designed to prevent the re-identification of individual identities. The remainder of this handbook describes how to request disclosure review in Section 2 and disclosure avoidance requirements in Section 3.

2. Requesting Disclosure Review

2.1 How to Request Disclosure Review

Researchers must fill out a [Disclosure Review Request Form](#) to initiate their review request. A high-level overview of the disclosure request process is illustrated in the figure below.

Disclosure review occurs on an ongoing basis. Each project space can request *one disclosure review per month*. Requests for more than one disclosure review in a month may be accepted at the discretion of CJARS staff and will depend on staff workload at the time of the request. Each disclosure request can include up to 1,000 point estimates. Statistics in supporting documentation do not count against the count of 1,000 point estimates.

Notification of approval, denial, or requests for additional information will be sent within 10 business days of form submission. Requests for more information require additional review time.

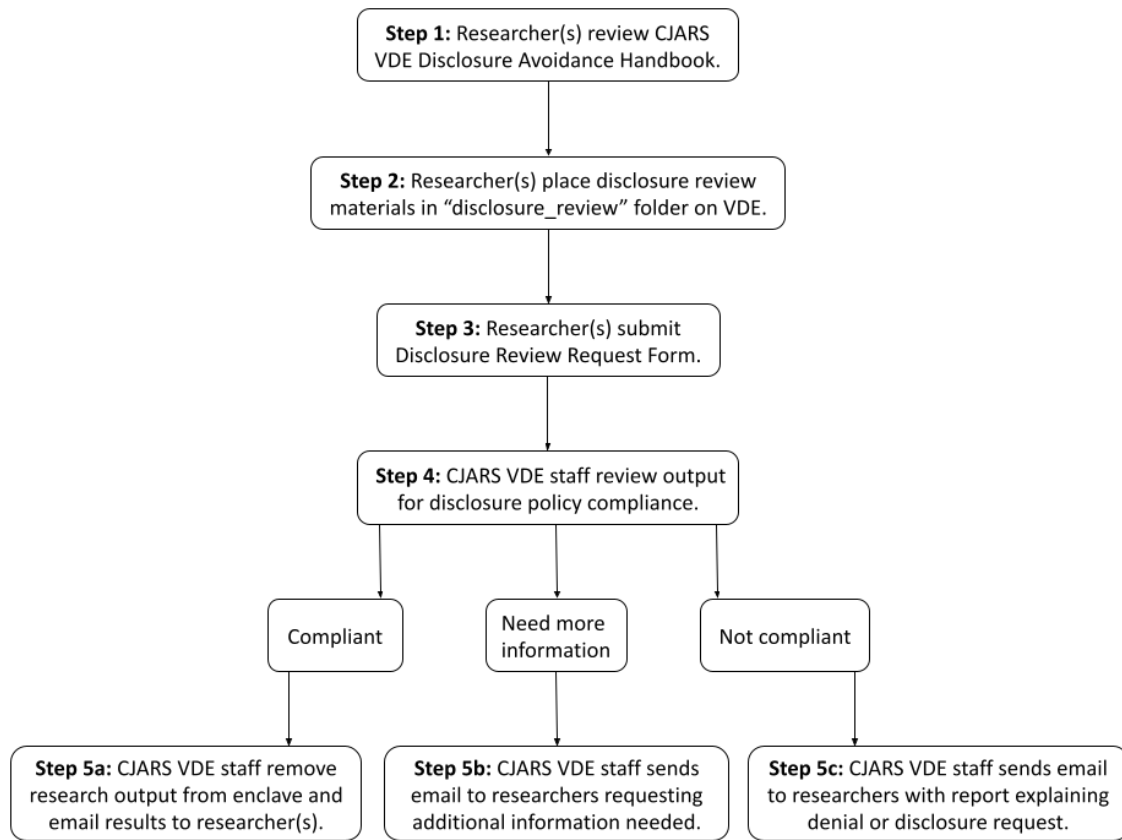
2.2 Document Formatting Requirements of Research Output

Researchers must create a folder named “disclosure_review” in their project folder on the VDE. Within that folder, research output being requested for disclosure and supporting documents

should be placed in a folder named with the date the review request is submitted “YYYYMMDD”. Only research output in the folder dated the date of the review request will be reviewed.

Each table of statistical output should be saved as a separate file and named as: “research_output_1”, “research_output_2”, etc. Excel format is strongly preferred.

Disclosure Review Process Flowchart



2.3 Supporting Documentation Requirements

Each table of statistical output must include two supporting documents.

1. *Narrative description.* Each table of research output must have an accompanying narrative description that describes (1) the sample, (2) analytic method used, (3) and how the output falls within the scope of the project as described in the Project Proposal Form. Each

narrative description should be saved as a separate file and named: “narrative_description_1”, “narrative_description_2”, etc. with the number corresponding to the associated research output table. Text files or Word documents are preferred.

2. *Supporting tables.* Each table of research output must include an accompanying table of supporting statistics. The supporting statistics will include any statistics not already included in the research output table that is required to confirm conformance with the disclosure avoidance requirements described in Section 3 of this document (e.g., cell sizes, degrees of freedom). Each supporting table should be saved as a separate file and named: “supporting_table_1”, “supporting_table_2”, etc. with the number corresponding to the associated research output table. Supporting table(s) will not be disclosed. All statistics for disclosure must be in the research output table(s). Excel format is preferred.

2.4 Review of Graphical Output and Code

Graphs and figures can be reviewed. However, it must be accompanied by a tabular version of the statistical information that underlies the graph or figure and a text file describing the sample. All disclosure requirements described here apply to the tabular information that underlies a graph or figure.

Code files can be reviewed. Researchers should ensure that no identifiable information is contained in the code. Additionally, code should be parsimonious. Researchers must include a text file accompanying the code that explains why it must be disclosed (e.g., journal requires submitting code for replication).

2.5 Other Non-standard Review Requests

Requests for a large amount of research output (1,000+ point estimates) may require additional review time and fees. Please reach out to cjars-vde-users@umich.edu in advance to discuss review of large requests.

CJARS acknowledges that there are rare circumstances which might require additional disclosure avoidance considerations not covered here. CJARS reserves the right to require researchers to take additional precautions to ensure the protection of individual identities. These decisions will be made on a case-by-case basis.

Review of data that has been linked using CJARS Custom Roster Integration may also require additional considerations. Researchers should consult with CJARS VDE staff prior to analysis.

3. Disclosure Avoidance Requirements

This section describes all requirements that research output must meet to receive disclosure approval. Research output is anything derived from data available on the CJARS VDE. Researchers should ensure all requirements described in this section are met to avoid denials or prevent delays in receiving disclosure approval.

3.1 Cell Size and Suppression

All statistics must be derived from a minimum sample of 5 individuals. This includes all samples and any subgroups within tabular output and modeled parameters. Statistics from samples below this threshold must be suppressed. The table below illustrates a fictitious example where the two cells shaded gray would need to be suppressed because they do not meet the minimum threshold of 5 individuals.

Average Number of Felony Convictions of Incarcerated Individuals by Sex and Age Category

		Age Category					
		18-24	25-34	35-44	45-54	55-64	65+
Males		1.5	1.9	2.9	3.1	3.2	3.2
		n = 3,368	n = 5,426	n = 1,215	n = 511	n = 62	n = 17
Females		1.6	1.8	2.6	2.7	2.8	2.9
		n = 321	n = 608	n = 145	n = 48	n = 4	n = 3

Researchers must also consider the existence of complementary cell(s). A complementary cell is one that can be used to derive information about a suppressed cell. A simple example of this is a table that presents output from a binary variable and information about the full sample. The table below illustrates an example where the orange cell could be derived from the information

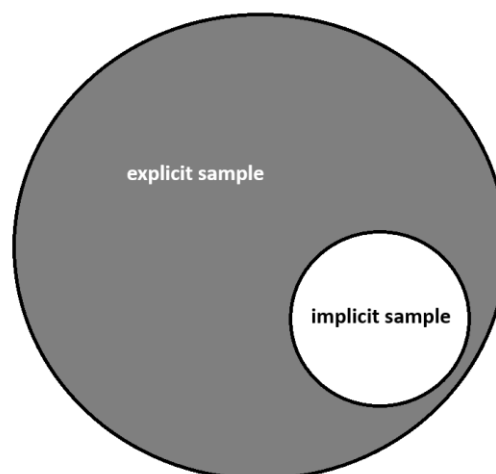
available on males and the full sample. Additional suppression is required to prevent the risk posed by complimentary cells, such as suppression of the male sample size.

Average Number of Felony Convictions of Prison Population by Sex for Individuals 65+

Males	3.2 n = 17
Females	2.9 n = 3
Full Sample	3.0 n = 20

Similar to complementary cells, implicit samples pose risk of disclosure but are less obviously problematic than a complimentary cell. An implicit sample is one not directly identified in research output, but can be assumed through deduction. For example, a statistic derived from a population with an exclusion criteria could pose disclosure risk for the excluded subsample if details about the population are publicly known. The figure below illustrates this situation. Researchers should describe implicit samples in their disclosure materials (i.e., text file that describes the sample) when known and should suppress information necessary to eliminate disclosure risk (e.g., rounding sample size).

Example of Implicit Sample



3.2 Top and Bottom Coding

Percentages that are 0% or 100% must be bottom or top coded, respectively. Alternatively, they could be masked. Percentages that round to 0% or 100% must also be bottom or top coded. The same rules apply to proportions. Percentages and proportions with a numerator fewer than five individuals must suppress the numerator and denominator, but not the percentage/proportion. The table below lists these requirements.

Top or Bottom Coding of Percentages and Proportions

Situation	Solution		
	Numerator	Denominator	Percent/Proportion
Percent is 100%, or rounds to it	Suppress	Suppress	Top code or suppress
Percent is 0%, or rounds to it	Suppress	Suppress	Bottom code or suppress
Difference between numerator and denominator is fewer than 5 individuals	Suppress	Suppress	Allowable
Numerator is fewer than 5 individuals	Suppress	Suppress	Allowable

The rules in the table below can be used to top or bottom code percentages and proportions if a researcher does not want to suppress a percentage/proportion. For example, a statistic of 100% derived from a sample of 500 individuals would be top coded to $\geq 99\%$.

Top and Bottom Coding Requirements

Denominator	Solution	Top Code	Bottom Code
n < 5	Suppress	Suppress	Suppress
5 < n < 15	Change by 10%	≥90%	≤10%
15 < n < 20	Change by 7%	≥93%	≤7%
20 < n < 30	Change by 4%	≥96%	≤4%
30 < n < 50	Change by 3%	≥97%	≤3%
50 < n < 300	Change by 2%	≥98%	≤2%
n > 300	Change by 1%	≥99%	≤1%

3.3 Summary Statistics

In some circumstances, summary statistics can pose disclosure risk. The list below includes, but is not limited to, examples of summary statistics that cannot be released. Research must include in their text file describing the sample the existence of any observations that could contribute to disclosure risk.

- *Median* – cannot be disclosed when median is a unique value
- *Minimum and maximum values* – cannot disclose minimum or maximum values
- *Outliers* – cannot disclose outliers
- *Influential observations* – cannot disclose statistics where one observation is overly influential (e.g., a sample consisting of the values: 1, 2, 1, 2, 3, 500)

3.4 Saturated Models

Researchers must report the degrees of freedom with model output. A minimum of 5 degrees of freedom (i.e., the number of contributing observations minus the number of variables used in the model) is required for model output.

3.5 Geographic Aggregation

Statistics can only be disclosed that are geographically aggregated in a manner that aligns with the project's Project Proposal Form. This allows CJARS VDE staff to ensure that the output complies with the data use agreements it holds with its data providers.