

**National Center for Health Statistics
Research Data Center**

**Disclosure Manual
*Preventing Disclosure:
Rules for Researchers***



Preventing Disclosure: Rules for Researchers

Contents

[Part 1: Confidentiality and the Research Data Center](#)

- Introduction
- Preventing Disclosure: RDC Procedures and Researcher Rules
- Confidentiality and the Law
- Sanctions for Violating Rules

[Part 2: Approved Proposals: Next Steps](#)

- Amending Your Approved Proposal
- Confidentiality Training
- Confidentiality Paperwork
- Creating Your Analytic Data Set
- Providing Public-use and Non-NCHS Data
- Important Notes About Submitting Public Data
- Merging Data
- Policies to Protect Geographical Information
- Policies to Protect Temporal Information
- Policies to Protect Perturbed and Masked Information

[Part 3: Working With Restricted-use Data](#)

- Working on Site at an NCHS RDC
- Working on Site at an FSRDC

[Part 4: Disclosure Review Policies and Procedures](#)

- General Output Policies
- Output Policies Specific to the NCHS RDC and FSRDCs

[Part 5: Publishing Your Research](#)

- Your Disclosure Responsibilities
- Disclosure Concerns to Consider
- How to Cite the RDC in Your Publication

Preventing Disclosure: Rules for Researchers

Part 1: Confidentiality and the Research Data Center

Introduction

The National Center for Health Statistics (NCHS) Research Data Center (RDC) provides the opportunity to conduct research using confidential respondent-level data. This privilege comes with special requirements. These rules and procedures are based on legal requirements associated with your appointment as a researcher in the RDC. Researchers at the RDC are legally obligated to maintain the confidentiality of respondent identities.

Your proposed research and the output that will be generated will be subject to an in-depth review. You will conduct your work in a secure computer environment with unique rules and procedures, which are outlined in this document.

Preventing Disclosure: RDC Procedures and Researcher Rules

The NCHS RDC strives to maintain confidentiality and prevent disclosure of confidential information. Researchers with approved projects also have important responsibilities in preventing disclosure. This document highlights both, with an emphasis on the rules and procedures that researchers must follow. **Failure to comply with RDC rules and procedures will result in cancellation of your research project and potential disbarment from future research activities in the NCHS RDC.**

Preventing Disclosure: Rules for Researchers

RDC Procedures to Prevent Disclosure	Your Responsibilities
<p>1. Limit access: Restricted-use data cannot leave the NCHS RDC and the Federal Statistical RDC (FSRDC) network. Not all projects are suitable for all RDC sites.</p>	<ul style="list-style-type: none"> • Do NOT attempt to remove data sets from RDC workstations. • Do NOT attempt to save output, files, or programs to transportable electronic media. RDC Analysts will email approved output or programs to you after a disclosure review.
<p>2. Require a research proposal: The Review Committee carefully examines the variables requested, the plan of analysis, and the desired output. The approved proposal becomes a contract between the RDC and the researcher.</p>	<ul style="list-style-type: none"> • Follow the proposal process instructions and instructions for amending your proposal. • Conduct ONLY the analyses for which you have received approval.
<p>3. Require confidentiality training and completion of confidentiality paperwork: The RDC has an online confidentiality training that reviews policies and procedures related to maintaining confidentiality. Additionally, there are legal documents that give researchers Designated Agent status and bind them to the laws that govern the confidential data. These forms are specific to the project and must be completed each time a researcher has a new approved proposal.</p>	<ul style="list-style-type: none"> • Comply with all requirements, rules, and instructions outlined in the confidentiality training and those described in the paperwork. • Hold in strictest confidence the identity of any establishment or individual in any documents, discussion, or analysis. • If there is an inadvertent identification of an establishment or individual, immediately bring it to the attention of RDC staff.
<p>4. Data access policies, procedures, and rules: The RDC must follow rules set by the Department of Health and Human Services (HHS), NCHS, and the data owner for data management and output release to prevent disclosure. The most common rules are highlighted in this document. However, special rules may be put in place upon approval of your proposal or as required by the data owner.</p>	<ul style="list-style-type: none"> • Follow all data access rules, procedures, and policies. Violation of these rules can result in suspension of your research project. • Read data documentation for all years and cycles of data before you visit the RDC.
<p>5. Analytic data sets for researchers: RDC Analysts will merge restricted-use, public-use, and non-NCHS data for the researcher. Whenever possible, restricted-use merge variables will be removed or may be randomized or coarsened versions of restricted-use variables.</p>	<ul style="list-style-type: none"> • Follow the guidelines for providing the public-use and non-NCHS data to the RDC Analyst. • Do NOT use ANY technique to learn the identity of any person, establishment, or sampling unit in any restricted-use file.
<p>6. Perform disclosure review: An RDC Analyst must review all output before it can be released to the researcher.</p>	<ul style="list-style-type: none"> • Follow the Output Review Policies and the Publishing Guidelines. • Only request review of output that is listed in your approved proposal. • Do NOT remove any output or notes until an RDC Analyst has reviewed them.

Preventing Disclosure: Rules for Researchers

Confidentiality and the Law

There are two laws that govern the NCHS RDC: Section 308(d) of the Public Health Service Act and the Confidential Information Protection and Statistical Efficiency Act (CIPSEA). The Public Health Service Act protects confidentiality and states that the only people who can access confidential data are NCHS staff and Designated Agents. Therefore, researchers wishing to access confidential data must become Designated Agents. CIPSEA stipulates that the penalty for willfully violating confidentiality is a class E felony with up to 5 years in prison or a \$250,000 fine or both. The Freedom of Information Act does not apply to data collected under CIPSEA.

Section 308(d) of the Public Health Service Act 42 U.S.C. 242m(d) states:

“No information, if an establishment or person supplying the information or described in it is identifiable, obtained in the course of activities undertaken or supported under section 242b, 242k, or 242l of this title may be used for any purpose other than the purpose for which it was supplied unless such establishment or person has consented (as determined under regulations of the Secretary) to its use for such other purpose; and in the case of information obtained in the course of health statistical or epidemiological activities under section 242b or 242k of this title, such information may not be published or released in other form if the particular establishment or person supplying the information or described in it is identifiable unless such establishment or person has consented (as determined under regulations of the Secretary) to its publication or release in other form.”

<https://www.cdc.gov/nchs/data/misc/staffmanual2004.pdf> (page 1)

CIPSEA states:

“Whoever, being an officer, employee, or agent of an agency acquiring information for exclusively statistical purposes, . . . comes into possession of such information by reason of his or her being an officer, employee, or agent and, knowing that the disclosure of the specific information is prohibited under the provisions of this title, willfully discloses the information in any manner to a person or agency not entitled to receive it, shall be guilty of a class E felony and imprisoned for not more than 5 years, or fined not more than \$250,000, or both.”

<https://www.cdc.gov/nchs/data/misc/staffmanual2004.pdf> (page 2)

Sanctions for Violating Rules

In addition to the potential legal ramifications specified above by CIPSEA, researchers who violate RDC rules will lose access to the RDC and related data and their sponsors and institutions will be notified. Researchers who are suspected of violating the rules may be prevented from continuing their research. If this happens, access to the data is denied for all members of the research team. Research team members interested in continuing the project must submit a new proposal. If approved, a new data file will be created and the project will be charged additional RDC fees.

Preventing Disclosure: Rules for Researchers

Part 2: Approved Proposals: Next Steps

Approval of a proposal does not mean that NCHS endorses the merit of your research or its substantive, methodological, theoretical, or policy relevance. NCHS approval reflects the judgment that the research, as described in the proposal and conducted in a secure RDC, can proceed to data analysis after the additional procedures outlined on the RDC website are completed. Approval of a proposal does not explicitly or implicitly guarantee that all output generated by the researcher will be released.

Amending Your Approved Proposal

Research evolves and your analysis plan may change after you have submitted the proposal. As these changes may affect the disclosure risk, inform the RDC Analyst of any changes that occur throughout the process. Update all relevant sections of the research proposal to reflect these changes. Highlight all changes in the proposal, include a revision date, and submit the revised proposal to your RDC Analyst.

The following are examples of common changes and the standard responses. Contact your RDC Analyst if you have specific questions about a possible amendment.

New Variables: New variables related to the original research question must be reviewed. Add the variables to the data dictionary in your proposal and explain how they will be used.

Adding public variables requires the RDC Analyst's approval and may require the entire Review Committee's approval.

Adding restricted-use variables or non-NCHS variables or both always requires the RDC Review Committee's approval. This process typically takes 2 weeks.

Adding variables unrelated to the original research will most likely require a new proposal and a complete review.

New Methods or Types of Output: If the analysis starts to take a different direction, inform the RDC Analyst, as this may significantly change the disclosure risk. Discuss the changes with your RDC Analyst. They may decide additional review by the committee is required. Highlight these changes in the proposal.

New Researchers: If the Primary Investigator or individuals in direct contact with the data change during the research project, the new researcher must:

- 1) Complete the Confidentiality Training and related paperwork.
- 2) Update the proposal to reflect the change and include the new researcher's CV, resume or bio-sketch.

Change to the Location of Data Access: Changes to the location of data access require additional logistics and may incur additional costs. Follow these steps to update your location of access:

- 1) Update the proposal to reflect the change.
- 2) Discuss the logistics with your RDC Analyst.

Confidentiality Training

All research team members must take the confidentiality training and complete the confidentiality forms. It should take about 20 minutes to complete the training. At the end of the training, you will take a quiz. You must score 100% on this quiz. Confidentiality Training is at: <https://www.cdc.gov/training/products/ConfidentialityOrientation/>.

Preventing Disclosure: Rules for Researchers

Confidentiality Paperwork

Because the forms are specific to the proposal, complete them every time you have a new approved proposal with the RDC.

Forms:

- Designated Agent Form (must be notarized)
- Agreement Regarding Conditions of Access

Provide the completed forms and your confidentiality training certificate to your RDC Analyst as soon as possible. Send forms via email or mail; if emailing, make sure the notary seal is visible. You cannot begin your research until the RDC receives your completed forms.

Creating Your Analytic Data Set

There are several steps involved in creating your analytic data set; some responsibilities are yours and some are the RDC Analyst's. Communicate regularly with your Analyst during this time. RDC staff will merge the public, restricted-use, and any non-NCHS data to create your analytic data set. Follow the guidelines below when you provide public-use and non-NCHS data.

Unless otherwise specified, data sets will be made accessible as SAS data sets.

Providing Public-use and Non-NCHS Data

Researchers are responsible for providing the NCHS public-use data set and any non-NCHS data. Compiling the public-use data set allows you to become familiar with the data and helps expedite the data creation process. There are a few exceptions; see the website or contact your RDC Analyst for details. Follow these steps when providing your RDC Analyst the non-restricted-use data for merging:

1. Create a public-use data set that includes only the variables specified in your approved proposal. If you want to include additional variables, update your proposal and discuss the changes with your RDC Analyst. Additional variables will likely require additional review.
2. If you would like to rename variables, include the original variable name in the variable description.
3. If you choose to create derived variables before working with the data, clearly define these variables in the proposal. The variable description should include the original variable name and an explanation of any arithmetic manipulation. Save the code you used to create these variables, as your RDC Analyst may request it.
4. If you received approval to merge NCHS data and non-NCHS data (e.g., Census data), the non-NCHS data set should only include the variables specified in your proposal.
5. Email the data files and a list of the variables to your RDC Analyst. If your data files are too large to email, discuss other delivery options with your RDC Analyst.

Important Notes About Submitting Public Data

Any attempt to include variables that may lead to re-identification of study participants or establishments is a disclosure violation and will result in the cancellation of your project and possible legal actions.

Preventing Disclosure: Rules for Researchers

If you are requesting access to the restricted-use Mortality files, you cannot include any public-use mortality variables or any variables derived from the public-use mortality data.

Merging Data

The RDC Analyst will merge all of the files based on the specifications in the merge section of the approved proposal. Discuss the merge with your RDC Analyst throughout the process to ensure that your analytic data set is created to your specifications.

RDC Analysts will follow the policies to protect geographical, temporal, and perturbed or masked data described in the section below.

After the merge is complete, the RDC Analyst will send you the PROC CONTENTS for review. Check the number of observations, variables, and the list of variables to make sure everything has been included.

Policies to Protect Geographical Information

The RDC Analyst will apply the following techniques when creating analytic data sets:

Remove Geography: Geographical variables are removed when the variable is only needed to merge NCHS data to another source.

- **Randomize Geography**: A random version of the geographical variable will be provided for analysis.
- **Coarsen Geography**: When lower levels of geography can be grouped into larger areas, the RDC Analyst will create the coarsened variable and will not provide access to the underlying lower level of geography. Analysts may request that the researcher write the code for how they want the variables created; researchers are always welcome to provide their own code and to review the code for created variables.
- **Provide Geographical Identifiers**: True geography will remain in the data set when used to make estimates at a lower level of geography as described in the approved proposal. However, most NCHS data are not representative at geographical levels not available on the public-use files; therefore, this is the least common use of geography in the RDC.

Policies to Protect Temporal Information

Similar safeguards are also necessary for temporal variables not included in public-use files:

- **Coarsen Dates**: The RDC Analyst will create the coarsened variable (e.g., year, month, quarter) and will not provide access to the underlying exact date unless there is strong justification. Analysts may ask the researcher to write the code for how they want the variables created; researchers are always welcome to provide their own code and to review the code for created variables.
- **Creating Variables of Time From Dates**: If exact dates are used to calculate time (e.g., exact length of life calculated based on date of birth (DOB) and date of death (DOD), the exact dates should only be used for data management and the resulting created variables (e.g., length of time) should be used in the analysis.

Preventing Disclosure: Rules for Researchers

Policies to Protect Perturbed and Masked Information

Perturbation and masking, common disclosure limitation methods, allow potentially sensitive information to be changed in a way that allows it to be made public. If the method were to be revealed, the public-use files could be compromised. The following are examples of public-use variables that cannot be included in researcher data sets:

- Public-use and restricted-use mortality variables
- Pseudo, masked, and true primary sampling units and Strata variables for the National Health and Nutrition Examination Survey

Part 3: Working With Restricted-use Data

Working on Site at an NCHS RDC

Getting Ready for Your Visit

1. At least five (5) business days before your visit, email your statistical programming code and reference documentation (codebooks, statistical reference materials, etc.) to your RDC Analyst. The RDC Analyst will review the documents and upload them to a workstation.
2. RDC computers do not have access to the Internet, so you will need to bring all resources with you. Flash drives are not allowed in the RDC.

On Site NCHS RDC Rules

When you come on site to an NCHS RDC, you will experience restrictions that the NCHS RDC has developed to prevent the disclosure of confidential information.

1. An RDC Analyst will provide your analytic data file on a workstation.
2. If you have approval for multiple research projects, you are only allowed to work on one project on one computer at a time.
3. Absolutely NO individual respondent-level data are allowed to leave the RDC facilities.
4. The RDC reserves the right to search all materials brought into the facility and all materials that will leave the facility.
5. Cell phones, pagers, laptops, and other communication devices are not permitted in the RDC.
6. RDC computers do not provide access to the local network or the Internet.
7. Researchers may derive variables from the analytic data set to create new variables but they cannot introduce new data using their statistical code. Do NOT attempt to put any content in your code that would facilitate re-identification of a study participant. This is a disclosure violation and will result in the cancellation of your project and possible legal actions.
8. You are not allowed to bring items into the RDC that may enable you to identify study participants (e.g., a national hospital directory, information from voter registration).
9. All output and all notes generated in the RDC are subject to disclosure review by RDC Analysts before removal from the RDC.
10. Allow for up to 3 weeks for your output to be reviewed. If your output is approved, it will be emailed to you.
11. If you would like a copy of your updated code after your visit, ask your RDC Analyst to review the code. If approved, the RDC Analyst will email the program code to you.
12. Before you submit your request for output review, review it for potential disclosure concerns and unnecessary information.

Preventing Disclosure: Rules for Researchers

13. Populate the actual tables that appear in your approved proposal. Output that does not appear in your approved proposal will not be reviewed or released.
14. Submit your output in a plain text file (.e.g. .lst files from SAS) tab delimited text [.txt] files or comma-separated [.csv] files so files can be opened and read in Windows Notepad. Release of additional formats for output is at the discretion of the RDC Analyst.

Working on Site at an FSRDC

Getting Ready for Your Visit

NCHS RDC Analysts and FSRDC Administrators have unique roles and responsibilities. Contact your RDC Analyst and FSRDC Administrator before your first visit to ensure that everything will be ready when you get there.

NCHS RDC Analyst:

- Facilitates review of your proposal
- Creates your analytic data set
- Accepts payments
- Accepts and processes required NCHS Confidentiality paperwork
- Transfers your data set to Census
- Reviews your output for disclosure risk

FSRDC Administrator or Designee:

- Answers questions about the FSRDC RDC
- Ensures you have completed all of the Census Bureau Security requirements
- Transfers your output to NCHS

To work in an FSRDC you must become an NCHS Designated Agent and have Special Sworn Status (SSS) from Census. Contact your FSRDC Administrator for information about obtaining Special Sworn Status. Keep your RDC Analyst up to date on where you are in the SSS approval process. After your SSS is approved and your account is created, your RDC Analyst will transfer your analytic data set to Census.

On Site FSRDC RDC Rules

While working at a FSRDC RDC, you are subject to all the same rules and restrictions as in NCHS RDCs. An additional FSRDC rule is:

1. You must follow FSRDC procedures to transfer files to the RDC. If you need help, ask your FSRDC Administrator.

Email all documents and reference materials to your RDC Analyst at least five (5) business days before your visit to an FSRDC. No notes, papers, or hard copies of any kind are permitted.

Part 4: Disclosure Review Policies and Procedures

RDC disclosure review policies and procedures exist to protect the identity of study participants.

General Output Policies

1. A data set is not appropriate output and will not be released.

Preventing Disclosure: Rules for Researchers

2. Absolutely NO output will leave the RDC facilities without first being reviewed by an RDC Analyst for possible disclosure of confidential information.
3. Output for review MUST match the research questions and the output described in your approved proposal.
4. RDC Analysts may apply cell suppression criteria: Guidelines may differ by data system and possibly by survey year because of sample size, sample design, and content. Sometimes specific projects pose additional disclosure concerns and additional or more stringent cell suppression may be applied.
5. Approved output is returned to the researcher via email approximately 3 weeks after the date of request. Plan ahead to allow this amount of time to receive your approved output.
6. Although RDC Analysts review the output files for disclosure concerns, it is your responsibility to use the output and statistics in a way that will not pose additional disclosure risks to study participants.

Output Policies Specific to the NCHS RDC and FSRDCs

All output requests must match what was approved in your proposal. Follow the steps outlined here and in the Disclosure Manual. Otherwise, you may be asked to return to the RDC to redo your output so that it conforms to RDC policy and your approved proposal.

Before submitting output to your RDC Analyst for review, you must:

- Make sure it is in a form that can be released by the RDC. Output will be released in a plain text file (i.e., files that can be opened and are readable in Windows Notepad, such as tab delimited text [.txt] files or comma-separated values [.csv]).
- Populate the actual tables that will appear in your publication They must match what is in your approved proposal.
- Remove any output that you feel could lead to the identification of an individual or institution. If you have questions, ask your RDC Analyst.
- Output that has individual record-level information is not permitted. Remove any individual-level data from your output.
- Extreme values or values representing an individual must be removed. Examples include minima, maxima, medians, and modes. If a procedure such as Proc Univariate creates extreme observations, (e.g. 0, 1, 99, and 100 percentiles) those extreme values must also be removed.
- Re-categorize variables with a frequency less than 5. If you are unable to do so, add an asterisk to all cells with a frequency less than 5 before submitting them to your RDC Analyst.

Make sure output submitted for review includes:

- A description of the output (e.g., A regression of...).
- Descriptions of (sub)samples in the analysis and output (e.g., black males aged 20–29).

Intermediate output (output that will not appear in publication) can be created and used on site, but the RDC generally does not allow its release.

Preventing Disclosure: Rules for Researchers

- Examples of intermediate output include: unweighted n's, tables of preliminary descriptive statistics, large volumes of numbers, and early regression models.
- Reasons intermediate output increases disclosure risk: Similar tables based on different subsamples may cause complementary disclosure problems because comparison of tables could reveal information about the sample and individual characteristics.

Preventing Disclosure: Rules for Researchers

Part 5: Publishing Your Research

Use the statistics and results generated through your work at the RDC in a way that does not pose any disclosure risk for the study participants. To ensure that there are no disclosure risks in your proposed published results, your product (e.g., journal article, manuscript, white paper, slides, conference presentation, poster, other product) must be reviewed by the RDC before submission for publication. An additional review may be needed if substantive changes are made to how the output from the RDC is reported in the tables or text in subsequent versions of your product (e.g., a revise and resubmit from a journal article).

Email your product to your RDC Analyst to start this process. Put “Product Review” in the Subject Line of your email.

When conducting these reviews, the RDC does not comment on scientific merit or impose any merit-based publishing guidelines.

Your Disclosure Responsibilities

Before publishing, you must:

1. Adhere to any specific requirements in your approved proposal and any requirements in emails received from the RDC regarding your output.
2. Not include data in your publication. Journals can ask that data be included in your article and made available. This is not appropriate for the data used in the RDC. We recommend that you share the data dictionary and merge procedures that were included in your approved proposal. It is RDC policy not to share information about proposals, so all information regarding the data set must be provided by your research team.
3. Not reveal any information:
 - a. That could identify an individual or establishment.
 - b. That could be used to identify geographical areas where respondents live or were sampled unless you specifically have permission to make estimates for those areas in your approved proposal.
 - c. That provides exact dates.
 - d. That could unintentionally reveal information about any of the above items. For example, publishing information on external sources of data after they have been merged with NCHS data could facilitate the identification of areas, dates, or individuals.
4. Follow the RDC citation requirements.
5. Include the disclaimer.
6. Ask the RDC Analyst for help or clarification any time you have a disclosure concern or question.

Preventing Disclosure: Rules for Researchers

Disclosure Concerns to Consider

Ask yourself the following questions as you write a publication based on restricted-use data:

Question	Action
1. Are you mentioning anything in the publication that you learned while looking at the restricted-use data?	<ul style="list-style-type: none"> • If you intend to discuss something that you observed while in the RDC but this information did not undergo disclosure review, share this information with your RDC Analyst to assess if it poses a disclosure risk.
2. Were restricted-use geographical variables used in your project?	<ul style="list-style-type: none"> • Do not mention geographical information not available in the public files (e.g., specific states, counties), unless it was part of your approved proposal. • Do not discuss the number of geographical units represented in your output without written permission. • Do not present unweighted sample sizes related to lower levels of geography.
3. Did your project use any restricted-use temporal information/variables?	<ul style="list-style-type: none"> • Do not mention exact dates. • Do not mention coarsened dates (e.g., season, year) without written permission. • Do not present unweighted sample sizes related to temporal components (e.g., dates).
4. Did your final analytic data set include data from another data source that was merged on geographical or temporal components?	<ul style="list-style-type: none"> • Do not reveal any information that could facilitate the identification of areas, dates, or individuals who participated in NCHS or non-NCHS surveys.
5. Could an individual or establishment possibly be identified by any information in your publication?	<ul style="list-style-type: none"> • Discuss inclusion and exclusion criteria for subpopulation analysis in a way that does not inadvertently identify small cells or extreme cases. This includes information on individual(s) or establishment(s). Review for: <ul style="list-style-type: none"> ○ Fewer cases than those specified in the required cell suppression criteria. ○ Extreme cases or outliers that were identified during analysis. ○ Unweighted counts that could be subtracted from totals to reveal a sample size less than that specified in the required cell suppression criteria.

How to Cite the RDC in Your Publication

When describing where the restricted-use data for your study came from, specify that restricted-use data were accessed through the NCHS Research Data Center. Also indicate why these data were essential to your research study. Here are two examples:

Preventing Disclosure: Rules for Researchers

- Geographical variables including state, county, and tract were used to merge Census variables that provided neighborhood contextual information. State, county, and tract are restricted-use variables, and these data were accessed through the NCHS Research Data Center.
- To merge the patient and agency files, we needed the restricted facility identification variable. Because the analyses required restricted-use data, the data were accessed through the NCHS Research Data Center.

Do not place these kinds of statements in your products:

- Analyses of these data were approved by the NCHS Research Data Center.
- The NCHS Research Data Center reviewed and approved our results.

When displaying output in your products, provide the Data Source. The content of a Data Source citation is usually provided by the data owner in their data documentation.

Include the following disclaimer in your product:

- The findings and conclusions in this paper are those of the author(s) and do not necessarily represent the views of the Research Data Center, the National Center for Health Statistics, or the Centers for Disease Control and Prevention.