

**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 Methods and Researcher Responsibilities
- Confidentiality and the Law
- Sanctions for Violating Rules

[Part 2: The Proposal Process](#)

- Proposal Process
- Restricted Variables
- Restricted Merge Variables
- Amending Your Approved Proposal

[Part 3: Approved Proposals: Next Steps](#)

- Confidentiality Orientation
- Confidentiality Paperwork
- Creating the Analytic Data Set
- Providing the Public Use Data Set
- Merging the Data
- Policies to Protect Geographic Information
- Policies to Protect Temporal Information
- Policies to Protect Perturbed and Masked Information

[Part 4: Working with Restricted Data](#)

- Working On Site at an NCHS RDC
- Working On Site at a Census RDC
- Working through the Remote Access System

[Part 5: Disclosure Review Policies and Procedures](#)

- General Output Policies
- Output Review Policies Specific to
 - NCHS or Census RDCs
 - Remote Access System
- Cell Suppression

[Part 6: Publishing Your Research](#)

- Your Disclosure Responsibilities
- Disclosure Concerns to Consider
- Providing Your Publication Citations to the RDC
- Citing the RDC in Your Publishing

Preventing Disclosure: Rules for Researchers

Part 1: Confidentiality and the Research Data Center

Introduction

Working through the National Center for Health Statistics' (NCHS) Research Data Center (RDC) allows you the opportunity to conduct research using confidential line-level data. This is a special privilege and it comes with special requirements. Your research will be subject to an in depth review at two stages: during the proposal process and during output review. Additionally, you will work in environments with additional rules and limitations to which you may not be accustomed. The intention of this document is to outline these constraints for you. The constraints stem from the legal requirements to meet the conditions of your appointment as a researcher at the RDC- particularly the requirement to maintain confidentiality of the underlying data sets to which you have been granted access.

Preventing Disclosure: RDC Procedures and Researcher Rules

It is important to note that the NCHS RDC's primary concern is with disclosure. The NCHS RDC strives to maintain confidentiality and prevent disclosure in several ways. However, researchers with approved projects, also have important responsibilities in preventing disclosure. This document highlights both, with an emphasis on the rules researchers must follow.

| RDC Procedures to Prevent Disclosure | Rules for Researchers: Your Responsibility |
|---|--|
| <p>1. Limit access: Restricted data cannot leave the secure access modes which include: on-site at an NCHS RDC (Hyattsville, MD or Atlanta, GA), the remote access system, and the Census RDC network. Not all projects are suitable for all sites.</p> | <ul style="list-style-type: none"> • Do NOT attempt to remove data sets from the on-site workstations. • Do NOT attempt to save output, files, or programs to transportable electronic media. RDC Analysts will email approved output or programs to you. |
| <p>2. Require a research proposal: The Review Committee (RDC Analyst, RDC Director, Confidentiality Officer, and representation from the data system(s)) carefully examine the variables requested, the plan of analysis, and the desired output. This approved document 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. Failure to comply will result in cancellation of the research activity and potential disbarment from future research activities in the RDC. |
| <p>3. Provide confidentiality training and require researchers to complete confidentiality paperwork: The RDC has an online confidentiality orientation that reviews policies and procedures related to maintaining confidentiality. Additionally, there are three legal documents that when completed give researchers Designated Agent status and hold them accountable to the laws. These forms are specific to the project and must be</p> | <ul style="list-style-type: none"> • Comply with all requirements, rules, and instructions outlined in the confidentiality orientation and paperwork. • Do NOT use ANY technique in an attempt to learn the identity of any person, establishment, or sampling unit not identified on public use data files. • Hold in strictest confidence the identification of any establishment or individual that may be inadvertently revealed in any documents, discussion, or |

Preventing Disclosure: Rules for Researchers

| | |
|---|--|
| <p>completed each time a researcher has a new approved proposal.</p> | <p>analysis.</p> <ul style="list-style-type: none"> • Immediately bring any inadvertent identification of any establishment or individual to the attention of RDC staff. |
| <p>4. Design access mode policies, procedures, and rules: each mode of access has specific policies, procedures, and rules designed to prevent disclosure. These rules must be adhered to: violation of these rules can result in suspension of the research project.</p> | <ul style="list-style-type: none"> • Follow all access mode policies. |
| <p>5. Create analytic data sets for researchers: NCHS staff will merge restricted, public, and non-NCHS data for the researcher. Whenever possible, restricted merge variables will be removed or randomized/coarsened versions of restricted variables will be substituted.</p> | <ul style="list-style-type: none"> • Follow the guidelines for providing the public use and non-NCHS data. |
| <p>6. Perform disclosure review: All output must be reviewed by the remote access system or an RDC Analyst before it can be released to the researcher.</p> | <ul style="list-style-type: none"> • Follow the Output Review Policies and the Publishing Guidelines • Limit the amount of output requested for review. • Do NOT remove any output or notes until they have been reviewed for disclosure risk by RDC staff. • Ask for help anytime you have questions related to disclosure or if you do not feel that you have a complete understanding of what may be a concern. |

Confidentiality and the Law

There are two laws that govern the NCHS RDC: Section 308(d) of the Public Health Service Act and Confidential Information Protection and Statistical Efficiency Act (CIPSEA). In short, the Public Health Service Act asserts the importance of protecting confidentiality and 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 the penalties for violating confidentiality as up to 5 years in prison and/or a \$250,000 fine. We are often asked about how the Freedom of Information Act (FOIA) applies: the short answer is that records collected under the Public Health Service Act are not subject to FOIA.

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

“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

Preventing Disclosure: Rules for Researchers

(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.”

<http://www.cdc.gov/nchs/data/misc/staffmanual2004.pdf>, page 1

Confidential Information Protection and Statistical Efficiency Act (CIPSEA)

“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.”

<http://www.cdc.gov/nchs/data/misc/staffmanual2004.pdf>, page 2

Sanctions for Violating Rules:

In addition to the potential legal ramifications specified above by the Confidential Information Protection and Statistical Efficiency Act (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.

Preventing Disclosure: Rules for Researchers

Part 2: The RDC Research Process

Proposal Process

Researchers are required to follow the proposal process instructions and use the proposal format provided on the RDC website: <http://www.cdc.gov/rdc/B3Prosal/PP300.htm>. There are also instructions for submitting an amendment: <http://www.cdc.gov/rdc/B3Prosal/PP350.htm>. The research proposal is a legally-binding agreement between your research team and NCHS. It is important that all research team members be familiar with this document regardless of their role in writing it.

Restricted Variables

The “data requirements” section of the proposal format requires you to explain each of the variables you need and how you will use them. Please be as detailed as possible. If approved, you are being approved to use the variables as you stated in the proposal. If you plan to deviate from the approved use of the variables, you will need to inform your RDC Analyst and submit an amendment.

Restricted Merge Variables

If restricted merge variables (i.e. state, county, dates) can be removed, coarsened or randomized versions of can be substituted, please state so in your proposal. During the review process, if you haven’t specified removal, randomization, or coarsening for these types of variables, nor have you provided a need for the true versions of these variables, the Review Committee will require you to do so. If you have questions about the best choice for your research question, please contact your RDC Analyst: s/he is available to assist.

- Example of removing merge variables: state and county are used to merge urban/rural status to a data set. State and county are removed after the merge.
- Example of coarsening variables: Month of interview is requested to control for season. Instead of providing you the month variable, the RDC Analyst uses month to create a new variable called season, and that variable is provided in the analytic data set.
- Example of randomizing variables: NCHS and non-NCHS data are merged based on state and county, but you want to control for individuals who live in the same area, so you need an indicator of what county the person lives in. Randomized versions of variables are substituted for the true state and county and serve as that indicator.

Amending Your Approved Proposal

We understand that research evolves and your analysis plan may change from the day you submit the proposal to the end of your analysis. As these changes may affect the disclosure risk, it is important that your RDC Analyst be made aware of changes throughout the process. It is also important that the research proposal be updated to represent these changes. The following are examples of common changes and the standard responses. Please contact your RDC Analyst if you have specific questions about a possible amendment.

- New Variables: Addition of new variables related to the original research question must undergo a brief review.
 1. Please add the variables to the data dictionary in your proposal and an explanation of how they will be used: update all relevant sections of the proposal.
 2. Highlight all changes including a revision date and submit the updated proposal to your RDC Analyst.

Preventing Disclosure: Rules for Researchers

- Adding public variables requires the RDC Analyst's approval and may or may not require the entire Review Committee's approval.
- Adding restricted variables and/or non-NCHS variables 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 we must be informed as this may significantly change the disclosure risk.
 1. Discuss the changes with your RDC Analyst. He/she may decide additional review by the committee is required.
 2. 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, two steps must be taken:
 1. Complete the Confidentiality Training and related paperwork.
 2. Update the proposal to reflect the change, including submission of the CV.
- Change to the Mode of Access: Not only do changes to the mode of access require additional logistics and possibly cost, they also affect disclosure risk. Not all projects are suitable for all modes of access. Follow these steps to update you mode of access:
 1. Update the proposal to reflect the change.
 2. If you adding remote access, the proposal may require additional review.
 3. Discuss logistical requirements with your RDC Analyst.

Preventing Disclosure: Rules for Researchers

Part 3: Approved Proposals: Next Steps

Please keep in mind that approval of a proposal does not mean that NCHS endorses the merit of the proposed research or its substantive, methodological, theoretical, or policy relevance. NCHS approval reflects the judgment that this research, as described in the proposal and conducted in a secure RDC access mode, 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 analysis will be released.

Researchers are required to follow the instructions for next steps provided on the RDC website: <http://www.cdc.gov/rdc/B5AprovProj/AP500.htm>.

Confidentiality Orientation

The principal investigator and all research team members who come in contact with the data must take the confidentiality orientation and complete the confidentiality forms. It should take 10-20 minutes to complete the orientation. At the conclusion of the orientation, you will be asked to take a quiz. You need to score a 100%. Confidentiality Orientation: <http://www.cdc.gov/training/products/ConfidentialityOrientation/>

Confidentiality Paperwork

Because the forms are specific to the proposal, they will need to be completed each time you have a new approved proposal with the RDC. <http://www.cdc.gov/rdc/B4ConfiDisc/CfD400.htm>

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

Creating Your Analytic Data Set

There are several steps involved in creating your analytic data set: some responsibilities are yours and some are the responsibility of the RDC Analyst. It is important to communicate with your Analyst during this time. NCHS staff will merge the public, restricted, and any non-NCHS data to create your analytic data set. We provide access to data needed to answer your research questions, but will limit any excess variables. We take several steps to protect especially sensitive variables. We also ask you to follow these guidelines when you provide the public data.

Providing the Public Use Data

Researchers are responsible for providing the NCHS public dataset as well as any non-NCHS data. Compiling the public use dataset provides you the opportunity to become familiar with the data and expedite the data creation process. There are a few exceptions: please see the website or contact your RDC Analyst for details. Please follow these steps when providing your RDC Analyst the non-restricted data for merging.

1. Create a public data set that includes only the variables specified in your proposal.
2. Original NCHS variables must retain the name they are given in the public data set. If you would like to rename these variables, include the original variable name in the variable description.
3. If you choose to create derived variables prior to working with the data onsite or via remote access, make sure these variables are clearly defined. The variable description should include the original variable name(s) from which it was derived and any arithmetic manipulation must be explained. Please save the code you used to create these variables as your RDC Analyst may request it.

Preventing Disclosure: Rules for Researchers

4. If you are also sending another source of data, for example Census data, this data set should only include the variables specified in your proposal.
5. Discuss with your RDC Analyst the preferred format for any merge variables. This is especially important for complex merges that involve multiple data sets and multiple merge variables. Create the variables as your RDC Analyst requests. This helps expedite the merge and improves data quality.
6. If your RDC Analyst is deriving any variables for you from the restricted data, discuss how you would like them created in advance. For example, if you have requested a season variable to be created, you will need to provide a definition of season. You can also send the code you would like used to create the new variables.
7. Email the data files along with a list of the variables to your RDC Analyst. If your data files are too large to be emailed, please discuss other options with your RDC Analyst.

Important Notes about Submitting Public Data

- Any attempt to include variables that may lead to re-identification of subjects/establishments is considered a disclosure violation and will result in the cessation of your project and possible legal actions.
- If you are requesting access to the restricted Mortality files, you cannot include any public use mortality variables, or variables derived from the public use mortality data.
- Do not include variables that are not listed in your approved proposal without first updating your proposal and discussing the matter with your RDC Analyst. Additional variables will require additional review.

Merging the Data

1. The RDC Analyst will merge all of the files based on the specifications in the merge section of the proposal. We strongly encourage you to discuss the merge with your RDC Analyst throughout the process to ensure that the data set is created to your specifications.
2. RDC Analysts will follow the policies to protect geographic, temporal, and perturbed/masked data listed below.
3. Once the merge is complete, the RDC Analyst will send you a PROC CONTENTS for review. Please check the number of observations, variables, and the list of variables to make sure everything you are expecting has been included.
4. If you have questions about any of the restricted variables, do not hesitate to ask.
5. Data sets will be made accessible as SAS data sets unless otherwise specified in advance.

Policies to Protect Geographic Information

During the proposal process, decisions about how geographic variables are used will be specified. The RDC Analyst will apply the follow techniques when creating analytic data sets.

- **Geographic Identifiers Remain:** If the geographic variables are being used to make estimates at a lower level of geography than is publically available true geography can remain in the data set. However, it is important to note that little NCHS data is representative at geographic levels not available on the public files; therefore this is the least common use of geography in the RDC.
- **Coarsening Geography:** If lower levels of geography can be grouped into larger areas, the RDC Analyst will create the coarsened variable and 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 and review the code for created variables.

Preventing Disclosure: Rules for Researchers

- **Removing All Geography:** If the geographic variables are being used to merge NCHS data to another source of data and are not needed for analysis, the geographic variables are removed.
- **Randomizing Geography:** If the geographic variables are being used for the merge and for analytic purposes, random versions of the variables are substituted.

Policies to Protect Temporal Information

Similar challenges and safeguards are also necessary for temporal variables not included on the public files.

- **Coarsening Dates:** If coarsened dates (e.g. year, month, or quarter, not the exact date) are needed for merging or analysis, the RDC Analyst should create the coarsened variable and never provide access to the underlying exact date. Analysts may request that the researcher write the code for how they want the variables created: researchers are always welcome to provide and review the code for created variables.
- **Creating Variables of Time from Dates:** If exact dates are being used to calculate time (exact length of life calculated based on DOB and DOD), the exact dates should only be used for data management and the resulting variables (e.g. length of time) should be used in analysis. ANDRE users will be required to submit their data management code to create those values and the exact dates dropped prior to their data set being loaded on the system. This may also be enforced on site.

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 is revealed, the public files could be compromised. The following are examples of changed public variables that cannot be included in researcher data sets that include the more detailed restricted counterparts.

- Public and restricted mortality variables
- Pseudo/masked and True PSU and Strata variables for NHANES

Preventing Disclosure: Rules for Researchers

Part 4: Working With Restricted Data

Working On Site at an NCHS RDC

Getting Ready for Your Visit

1. Confirm with your RDC Analyst that you will be able to access the building. Make sure you have directions, information on parking, and any other logistical information you may need.
2. Email your statistical programming code, reference documentation (codebooks, statistical reference materials, etc.) to your RDC Analyst before your visit to the NCHS RDC. The RDC Analyst will review the documents and upload them to the workstation.
3. It is important to keep in mind that the RDC computers do not have access to the Internet, so you will need to bring all resources with you. Also, there is no wireless network anywhere in the RDC or in the government buildings where the NCHS RDCs are located. Please plan ahead to email your programs/documentation to your RDC staff member in advance of your visit. Email is the only method of transfer: we do not allow flash drives in the RDC.

On Site NCHS RDC Rules

1. When you come on site to NCHS RDCs, you will experience restrictions that have been developed to decrease the likelihood of a disclosure.
2. RDC Analysts will merge the approved confidential variables to the public-use data and other data sources (if applicable).
3. If you have approval for multiple research projects, you are only allowed to work on one at a time. The RDC lab computers are designed for one project to be worked on at a time. Do not attempt to access multiple projects or data from multiple projects at the same time.
4. Absolutely NO individual level data will leave the RDC facilities
5. The RDC reserves the right to search all materials brought into the facility and all those leaving the facility.
6. Cell phones, pagers, laptops, or other communication devices are not permitted in the RDC.
7. RDC computers do not provide access to the local network, the mainframe, or the Internet.
8. Researchers may derive variables from the data set provided 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 subject/establishment. This is considered a disclosure violation and will result in the cessation of the project and possible legal actions.
9. You are not allowed to bring items into the RDC that may enable you to identify individuals and/or establishments (e.g. a national hospital directory, information from voter registration).
10. All output and all notes written in the RDC are subject to disclosure review by RDC Analysts before removal from the RDC.
11. If you would like your updated code after your visit, please ask your RDC Analyst to review the code. If approved, the RDC Analyst will email the programs to you.
12. Before requesting that your output be reviewed, you must first have reviewed it yourself for potential disclosure concerns and unnecessary information.
13. Submit your output in a human-readable plain text file (i.e., .lst files from SAS, files that can be opened and that are readable in notepad such as tab delimited text .txt files or comma-separated values .csv files). Release of additional formats for output is at the discretion of the RDC Analyst. We strongly encourage you to populate the actual tables that will appear in your publication and submit those for review.

Preventing Disclosure: Rules for Researchers

Working On Site at a Census RDC

Getting Ready for Your Visit

It is important to remember the NCHS RDC Analyst and Census RDC Administrators have different and unique roles and responsibilities. Be in contact with both leading up to 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 the set up fee payment
- Accepts your NCHS Confidentiality requirements
- Transfers your data set to Census
- Reviews your output for disclosure risk

Census RDC Administrators or Designee

- Answer logistical questions about the Census RDC
- Ensure you have completed all the Census Bureau Security requirements
- Transfers your output to NCHS

Several steps need to occur before you can begin your work:

1. To work in a Census RDC you must become an NCHS Designated Agent and have Special Sworn Status (SSS) from Census. Contact your Census RDC Administrator for information about obtaining Special Sworn Status.
2. Your merged data set needs to be transferred to Census. Once your SSS is approved and your Census account created, your RDC Analyst will transfer your data. It is important to keep your RDC Analyst aware of where you are in the process.

On Site Census RDC Rules

While working at a Census RDC you are subject to all the same rules and restrictions as NCHS on site users in addition to following rules:

1. If you have approval for multiple research projects, you are only allowed to work on one at a time. Do not attempt to access multiple projects or data from multiple projects at the same time.
2. Before requesting that your output be reviewed, you must first have reviewed it yourself for potential disclosure concerns and unnecessary information.
3. You must request output review through the Census Tracking System. In addition, please send your RDC Analyst an email indicating that you requested output review and provide a description about what output you submitted and its purpose. Please allow three weeks to receive your approved output.
4. Submit your output in a human-readable plain text file (i.e., .lst files from SAS, files that can be opened and that are readable in notepad such as tab delimited text .txt files or comma-separated values .csv files). Release of additional formats for output is at the discretion of the RDC Analyst. We strongly encourage you to populate the actual tables that will appear in your publication and submit those for review.

Preventing Disclosure: Rules for Researchers

Working through the Remote Access System

Getting Ready to Use the Remote Access System

1. The remote access system (ANDRE) does NOT allow researchers to view any data at the individual record level.
2. Certain SAS procedures are not permitted on the remote access system. Remote access users, please read the complete list <http://www.cdc.gov/rdc/Data/B2/SASSUDAANRestrictions.pdf>.
3. You cannot create permanent data sets on the remote access system. You may write intermediate results/values/temporary datasets in the “work” directory. All the values/info stored in the “work” directory will disappear as soon as SAS finishes execution of your job. Any attempt to write to the hard drive will abort your run and the system will inform you of the violation.

Remote Access System Rules

4. Remote access users can only submit statistical code related to the analysis plan outlined in the research proposal. If additional/different analyses are desired, they must first be discussed with your RDC Analyst and may require a new or amended proposal.
5. Remote access rights are only granted to one person. Only that person can submit data requests and receive output from the remote access system (ANDRE).
6. Any attempt to circumvent ANDRE to obtain prohibited information that may result in a disclosure will result in an immediate suspension of the account and possible legal actions.
7. Output results that pose a disclosure risk will be suppressed. Additional cells or values may be suppressed to maintain confidentiality (complementary suppression).
8. Output will be returned in a text file via email.

Preventing Disclosure: Rules for Researchers

Part 5: Disclosure Review Policies and Procedures

Our disclosure review policies and procedures exist to protect the confidentiality of NCHS study participants.

General Output Policies

1. Data sets will not be released.
2. Output in the form of a data set will not be released.
3. Absolutely NO output will leave the RDC facilities without first being reviewed by an RDC Analyst or the remote access system.
4. Approved output is returned via email.
5. Although the output files are reviewed by an RDC staff member or the remote access system for disclosure concerns, it is your responsibility to use these statistics in a way that doesn't pose additional risk to the respondents. If you discover or can inadvertently deduce small cells or any individual-level information, it is your responsibility to not share that information with anyone or in any publication and to immediately bring it to the attention of your RDC staff member. If you have questions about any other concerns, it is your responsibility to ask your RDC staff member or the RDC Director for assistance.
6. Output MUST match the research questions/output suggested in the proposal.
7. If after disclosure review, the RDC has denied release of output, you have the choice to ask that the Data Review Board (DRB) consider the issue. Their decisions are final. You will have to write a one page justification explaining why the output can be released.

Output Policies Specific to NCHS or Census RDCs:

1. NCHS RDC Users: Output can be reviewed at the conclusion of the research project OR once per week. It will be returned within five business days unless a more extensive review is required.
2. Census RDC Users: Output will be reviewed and returned within three weeks of submission.
3. Make sure output submitted for review includes:
 - What is this output? (This can be a title: a regression of...)
 - Who is the (sub)sample in this analysis? (black males age 20-29)
 - How will output be used? (publication, presentation, etc.)
4. Before submitting output for review by your RDC Analyst, you must review your output for the following:
 - Make sure it is in a form that can be released by the RDC. Output will be released in the default format produced by the statistical package (i.e. .lst files for SAS) or in a human-readable plain text file (i.e., files that can be opened and that are readable in notepad such as tab delimited text .txt files or comma-separated values .csv files). We encourage you to populate the actual tables that will appear in your publication and submit those for review.
 - If there is any output that you feel could lead to the identification of an individual or institution, remove it. If you have questions, discuss this with your RDC Analyst.
 - Any procedure that produces output on an individual or institution must be removed.
 - Printing individual cases 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

Preventing Disclosure: Rules for Researchers

Univariate creates extreme observations, 0, 1, 99, and 100 percentiles, those extreme values must also be removed.

5. If you do not review your output before requesting that your RDC Analyst review it, you may be asked to return to the RDC to do so.
6. All cells with a frequency less than 5 should be asterisked before they are submitted to your RDC Analyst. If at all possible, we encourage you to consider re-categorizing the variables instead of requesting suppression. We will suppress small cells and an additional cell to prevent the determination of the unacceptable cell figure through subtraction (complementary suppression).
7. You must make a special request (which will take time to review) if you wish to have intermediate output (output that will not appear in publication) reviewed. This output can certainly be created and used on site, but we discourage its release.
 - Examples of intermediate output include: unweighted n's, tables of preliminary descriptive statistics, large numbers, early regression models
 - Reasons intermediate output increase disclosure risk: similar tables based on different sub-samples may cause complementary disclosure problems because comparison of tables could reveal information about the sample and individual characteristics.

Output Policies Specific to Remote Access:

1. Output should be received within a few hours. If it takes longer, please contact the Remote Access Team.
2. Although intermediate output is an evitable by-product of using the remote access system, it is important to note that this information must not be shared or published.
 - Examples of intermediate output include: unweighted n's, tables of preliminary descriptive statistics, large numbers of preliminary regression models
 - If you have questions about what constitutes intermediate output and what you cannot share in a publication, please contact your RDC Analyst.
3. Certain procedures are not possible using the remote access system because the output generated by them is deemed a disclosure concern. If you need to use any procedures listed on the SAS/SUDAAN Restricted Procedures list (<http://www.cdc.gov/rdc/Data/B2/SASSUDAANRestrictions.pdf>), you will need to consider coming on site or using the staff assisted mechanism (<http://www.cdc.gov/rdc/B2AccessMod/ACs240.htm>).
4. Output results are automatically reviewed by ANDRE and observations that pose a disclosure risk will be suppressed. Additional cells or values may be suppressed to maintain confidentiality (complementary suppression).
5. In addition to the automatic disclosure review/suppressions performed by the remote access system (ANDRE), outputs are randomly reviewed by an RDC staff member for disclosure concerns. It is your responsibility to use these statistics in a way that doesn't pose additional risk to the respondents. If you discover or can inadvertently deduce small cells (<5) or any individual-level information, it is your responsibility to not share that information with anyone or in any publication and to immediately bring it to the attention of your RDC staff member. If you have questions about any other concerns, it is your responsibility to ask your RDC staff member or the RDC Director for assistance.

Cell Suppression Criteria Employed by the RDC

Preventing Disclosure: Rules for Researchers

Guidelines may differ by data system and possibly by survey year because of sample size, sample design, and content. Sometimes specific projects pose additional concerns and will experience additional or more stringent suppression.

When one value is suppressed, the RDC will perform complementary suppression (in order to prevent determination of the unacceptable cell figure through subtraction, other cells will also be suppressed). ANDRE will automatically perform this operation. We strongly encourage you to re-categorize whenever possible, so as to avoid small cells.

Preventing Disclosure: Rules for Researchers

Part 6: Publishing Your Research

It is your responsibility to use the statistics you generate through an RDC mode of access in a way that doesn't pose additional risk to the respondents. These guidelines refer only to disclosure concerns related to publishing. The RDC does not comment on scientific merit or impose any merit-based publishing guidelines. You are entitled to ask RDC staff to review publication content for disclosure concerns. RDC staff will not comment on scientific merit. If RDC Staff deems language acceptable, that simply means that the language does not pose a disclosure concern, it is not an endorsement or approval of the research.

Your Disclosure Responsibilities

When publishing you must

1. adhere to any additional requirements specified in the approval email (During the proposal process, the Review Committee may request to review your publication for disclosure concerns prior to publication.)
2. not reveal any information that could identify an individual or establishment
3. not reveal any information that could be used to identify geographic areas where respondents live or were sampled unless you specifically have permission to make estimates for those areas
4. not reveal any information about specific dates
5. not reveal any information from external sources of data that have been merged to NCHS data based on temporal or geographic components that could facilitate the identification of areas, dates, or individuals
6. ask the RDC staff for help any time you have a disclosure concern

Disclosure Concerns to Consider

Ask yourself the following questions why writing a publication based on restricted data.

| Question | Action |
|--|---|
| 1. Are you mentioning anything in the publication that you learned looking at the restricted data? | If you are discussing something that you observed while in the RDC, but did not specifically include in output for review, share this information with your Analyst to assess if it poses additional risk. |
| 2. Were restricted geographic variables used to merge your data set? | Do not mention any specific geographic information not available on the public files (e.g. specific states, counties) unless your approved proposal specifically stated that you would. Limit the number of unweighted sample sizes you show related to lower levels of geography. If you are discussing the number of geographic units represented and didn't specifically state that you would in the approved proposal, ask your Analyst to review the statement |
| 3. Does your data set | Do not mention any exact dates. |

Preventing Disclosure: Rules for Researchers

| | |
|--|--|
| include any restricted temporal variables? | <p>Limit the number of unweighted sample sizes you show related to temporal components.</p> <p>If you want to mention coarsened dates (e.g. season, year), this must have been explained in your proposal. Discuss the appropriate language with your RDC Analyst before publishing.</p> |
| 4. Did your final analytic data set include data from another source that was merged on geographic or temporal components? | Follow the same actions when discussing the non-NCHS data as you would NCHS data. |
| 5. Could an individual or establishment possibly be identified by any information in your publication? | <p>Be mindful that the way you discuss inclusion and exclusion criteria could not inadvertently identify a small cell or extreme case. This includes information on an individual or a small group of individuals such as</p> <ul style="list-style-type: none"> • fewer cases than those specified cell suppression criteria • extreme cases or outliers that were identified during analysis • unweighted counts that could be subtracted to reveal a sample less than those specified cell suppression criteria. |

Providing Your Publication Citations to the RDC

Citations for all publications, presentations, and reports that refer to research conducted using the RDC must be emailed to rdca@cdc.gov and your RDC Analyst as soon as possible.

Citing the RDC in Your Publication

When publishing, we ask that your methods section specify which restricted variables you accessed through the Research Data Center and why they were essential to your research question. Here are two brief examples:

- Geographic variables including state, county, and tract were used to merge in Census variables which provided neighborhood contextual information. Randomized versions of the variables were substituted after the merge which allowed us to control for neighborhood effects. State, county, and tract are restricted variables therefore these data were accessed through the Research Data Center.
- To merge the patient and agency files, we needed the restricted facility identification variable. Because the analyses required restricted data, they were accessed through the Research Data Center.

Please avoid statements such as:

Preventing Disclosure: Rules for Researchers

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

Please add the following disclaimer at the conclusion of your publication:

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.