
DEPARTMENT OF HEALTH AND HUMAN SERVICES

Centers for Disease Control and Prevention

Procedures and Costs for Service at the Research Data Center

Last revised April 2016

AGENCY: National Center for Health Statistics, Centers for Disease Control and Prevention (CDC), Department of Health and Human Services (HHS).

ACTION: Notice and request for comments.

AUTHORITY: Section 306 of the Public Health Service Act, as amended (42 U.S.C. 242k) and Pub. Law 103-333.

SUMMARY: This notice provides information about the Research Data Center (RDC) operated by the National Center for Health Statistics (NCHS) within the Centers for Disease Control and Prevention (CDC). The Research Data Center was established in 1998 to provide a mechanism whereby researchers can access detailed data files in a secure environment, without jeopardizing the confidentiality of respondents.

Historically, the data files accessed in the RDC have consisted of NCHS survey data and vital statistics. RDC has recently begun accepting data files that were not produced from NCHS survey data. In order to assure that all data files are processed in a consistent manner, the original guidelines for accessing files in the RDC are being reviewed and revised as necessary.

As part of the revision process, potential users are being given the opportunity to provide input on how the procedures of the RDC can best

serve their research needs. This notice describes how to submit proposals requesting use of the data, mechanisms to access the RDC, requirements, use of outside data sets, costs for using the RDC, and other pertinent topics. We are seeking comments on these procedures and will post the final procedures on the NCHS Web site.

ADDRESSES: Send comments concerning this notice to Peter Meyer, National Center for Health Statistics, 3311 Toledo Road, Room 4113, Hyattsville, MD 20782, or e-mail to pmeyer1@cdc.gov

FOR FURTHER INFORMATION CONTACT: Peter Meyer 301-458-4375.

SUPPLEMENTARY INFORMATION:

Operational Procedures for Use of the Research Data Center; National

Center for Health Statistics; Centers for Disease Control and

Prevention

Table of Contents

National Center for Health Statistics Research Data Center Procedures

Background

The National Center for Health Statistics (NCHS) releases and hosts a range of statistical data products on the health and well-being of the nation and its health care system. Statistical tabulations (tables) present data in predetermined categories such as age, race, sex or geographic region that are important to describe health status and trends. In addition, statistical microdata containing health and related variables are published so that outside analysts may conduct original research and special studies to address issues of public health science and policy. Section 308 (d) of the Public Health Service Act, the Confidential Information Protection and Statistical Efficiency Act, and the NCHS Staff Manual on Confidentiality do not permit the release of data that are either identified or identifiable to persons that are not agents of NCHS. In order to preserve privacy and confidentiality, details that might identify or facilitate the identification of persons and entities participating in NCHS surveys and data systems either owned or hosted by NCHS are not released in published data products. Examples of data elements that might be abridged or suppressed to prevent reidentification are geographic identifiers, genetic data, details of sample design, and variables such as age or income that might exist in other databases.

Despite the wide dissemination of NCHS data through publications, web releases, etc., the inability to release files with these sensitive variables limits the utility of NCHS data for research, policy, and programmatic purposes and sets a boundary on one of the Department of Health and Human Service’s goals: to increase our capacity to provide state and local area estimates. In pursuit of this goal and in response to the public research community's interest in restricted data, NCHS established the Research Data Center (RDC). The physical facilities are places where researchers can access detailed data files in a secure environment without jeopardizing the confidentiality of respondents. Access is regulated by the Confidential Information Protection and Statistical Efficiency Act (CIPSEA) and other Federal statutes. The RDC provide restricted access to NCHS data and external data. Researchers function under the supervision of NCHS employees and agents of NCHS. The researchers and are subject to the same provisions of law with regard to confidentiality as NCHS employees.

Submission of Research Proposals Using NCHS Data

The Proposal Process outlines the primary steps leading up to the review committee decision. To access restricted data through the RDC, researchers must submit a proposal.

On receipt of a proposal, the RDC Director assigns an RDC Analyst. The RDC Analyst is the primary contact throughout researcher's project in the RDC. The RDC Analyst responsibilities include but are not limited to:

- Facilitating review of the project proposal
- Creating researchers' analytic data set
- Accepting payment
- Accepting researchers' NCHS Confidentiality requirements
- Transferring project dataset to the appropriate laboratory
- Reviewing output for disclosure risk

The Proposal Process

Researchers must:

Step 1: Determine a need for restricted data. Restricted Data

Step 2: Determine a preferred mode of access. Mode of Access

Step 3: Draft research proposal. The Proposal

Step 4: Submit proposal (using the Proposal Format) to rdca@cdc.gov .

Step 5: Wait for comments from the review committee and respond quickly to expedite review. Review Committee

Step 6: Update your proposal when there are changes. Amendment Procedures

Prospective researchers are encouraged to review data documentation and analytic guidelines specific to the data of interest. Researchers should develop proposals in a way that facilitates the ability of the RDC staff to assess disclosure risk and create the analytic files required by the project. Proposals should be explicit regarding the variables needed as well as any case selection required. Only those data items required to conduct the proposed analyses will be included in the analytic data file and the proposals should address why the requested data are needed for the proposed study. Overly large and complex projects, or poorly defined projects will require extensive communication between RDC staff and the researchers proposing the project, and this can cause the process to be delayed.

Proposal Review

The RDC Analyst will distribute the proposal to the other members of the Review Committee which consists of the Director of the NCHS RDC (or his designee), the RDC Analyst, the NCHS Confidentiality Officer, and a representative of the program producing the data.

The following criteria apply to proposal review:

1. Risk of disclosure of restricted information.
2. Appropriate use of the data and concurrence with the intended use for which it was collected. Including assurance that the use of the data is in accordance with the informed consent procedures associated with the collection of the data.
3. Scientific and technical feasibility of the project.
4. Availability of resources at the RDC.

The review usually takes 6-8 weeks. The exact amount of time is dependent on a number of factors including the complexity of the proposal, timely responses from the researcher, and the quality and clarity of the proposal. The Review Committee can make one of three decisions: approve, resubmit, or disapprove. Researchers should note that approval of their application does not constitute endorsement by NCHS of the substantive, methodological, theoretical, or policy relevance or merit of the proposed research. Rather, NCHS approval constitutes a judgment that this research, as described in the application, is not an illegal or unethical use (as determined by the informed consent and original reason for collecting the data) of the requested data file and does not jeopardize the confidentiality of the data. Approval of a proposal does not explicitly or implicitly guarantee that all output generated by the analysis will be released. Output that poses a disclosure risk will be suppressed.

Confidentiality and Human Subjects Protection

In order to access restricted data files in the RDC, researchers must sign an NCHS Designated Agent Agreement, the Agreement Regarding Conditions of Access to Confidential Data in the Research Data Center of the National Center for Health Statistics and complete the Confidentiality Orientation. The NCHS RDC expect that all researchers will adhere to established standards and principles for carrying out statistical research and analyses. Researchers must conduct only those analyses which received approval. Failure to comply with RDC rules and regulations will result in cancellation of the research activity and potential ban from future research activities in the RDC. In the case where Ethics Review Board (ERB) approval is required to conduct research, NCHS will notify relevant ERBs of infringements of protocol approvals.

As mentioned earlier, confidentiality protection at NCHS is governed by Section 308(d) of the Public Health Service Act, PHSA, and (42 USC 242m). Specifically, "No information, if an establishment or person supplying the information or described in it is identified, obtained in the course of activities undertaken or supported under Sections 304, 305, 306, 307, or 309 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 (1) in the case of information obtained in the course of health, statistical or epidemiological activities under Section 304 or 306, 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.

Any willful disclosure of confidential statistical information by the researcher is punishable under CIPSEA and carries a fine of up to $250,000 and up to 5 years in prison.

Methods to Access Data

Once the proposal is approved and all access requirements are met, researchers are allowed access, under strict supervision, to restricted statistical file(s). There are three modes of access:

1. NCHS RDC – The NCHS RDC are secure research facilities located at NCHS headquarters in Hyattsville, MD, the Centers for Disease Control and Prevention in Atlanta, GA and the Department of Health and Human Services in Washington, DC. Researchers are allowed access to approved restricted statistical microdata file(s). The workstations in NCHS RDC are “stand alone” and have no connection to NCHS network or the internet. Removable media drives, such as USB ports, on workstations are inaccessible. All workstations are installed with statistical programming languages (SAS, SAS-callable SUDAAN, and STATA) commonly used by statisticians.

2. Remote Access – When registering for the remote access system, Researchers are required to acknowledge and accept the terms of use for the remote access system. The terms includes procedures and programming limitations of the remote access system. Once researchers have accepted the terms, they send programs to, and receive output from, the remote system through a secure communication network. Their programs execute on a computer in the RDC. Both submitted programs and output are subjected to a programmed disclosure review and may also be subjected to a manual review. The .log file is also scanned with particular attention to certain types of error conditions that may spawn case listings. Some projects are not suitable for the remote access. Researchers should consider the programming limitations of the remote access system when choosing this method of access. However, the data stewards and RDC staff may also deem the project inappropriate for remote access during the review process. The remote access system is only available to programmers sufficient in SAS or SAS-callable SUDAAN.

3. Federal Statistical (formally Census)RDC – Analytic data sets are constructed at the NCHS RDC according to specifications included in the research proposal and are then securely transferred to the Census data processing facility. Users then view the data using “front end dumb terminals” at a Federal Statistical RDC. The data do not leave the processing facility. The researcher’s output is sent via a secure communication network to RDC staff for disclosure review. Once the output has been approved for release, it is sent via email to the researcher. A listing of available Census RDC locations can be found here:http://webserver02.ces.census.gov/index.php/ces/researchlocations

Researcher supplied Data

Many projects will require researchers to download public files from the internet and create an extract that includes only the variables required for the project. There are a few exceptions that the RDC Analyst will discuss as needed with the researcher.

The researcher may provide two types of data: 1) publically available NCHS data and 2) external sources of data. The RDC Analyst will accept researcher data files in SAS, STATA, or ASCII format (flat files) with variables either column delimited or column specific. The merging of researcher-supplied data with NCHS in-house data will be done by an NCHS RDC Analyst prior to researcher access.

Key points:

- The public-use file can only include those variables required for analysis. Do not send the entire public-use files.
- Original NCHS Variables must have the name they are given in the public-use data set. Researcher will have the opportunity to rename once they have been granted access to the dataset.
- Public-Use Mortality Variables: Do not include any public-use mortality variables or variables derived from the public-use mortality data if the project involves restricted mortality variables.
- Any attempt to include variables that may lead to re-identification of subjects or establishments is considered a disclosure violation and will result in the cessation of your project and possible legal actions.

The external data may consist of proprietary data collected and owned by the researcher or other publicly available data obtained such as Census data. Researchers are responsible for working with RDC Analyst to ensure that the data can be merged with the NCHS data and the format of the data is consistent with it. Researchers are also responsible for ensuring that they have proper consent for merging external data to data provided by the RDC.

Disclosure Review Process

All output will undergo disclosure review by an RDC Analyst and/or the remote access system. In general, disclosure review is consistent with the guidelines published in the NCHS Staff Manual on Confidentiality.

Output generated through RDC access mechanisms will be subject to a review that will include, but not be limited, to the following procedures:

1. No table should contain all cases of any line or column in a single cell.
2. In no case should the total figure for a line or column of a cross tabulation be less than 5.
3. In no case should a quantity figure be based upon fewer than five cases.
4. In no case should a quantity figure be released to the researcher if one case contributes more than 60 percent of the amount.
5. In no case should data on an identifiable case, nor any of the kinds of data listed in preceding items, be derivable through subtraction or other calculation from the combination of output on a given study.
6. Low level geography will not be included in output provided to the researchers.

The reviews will all be performed by an NCHS RDC Analyst who is trained in statistics and statistical disclosure limitation. For more information consult the Report on Statistical Disclosure Limitation Methodology: http://www.fcsm.gov/working-papers/wp22.html and the RDC Disclosure Manual.

Service Costs for Using the RDC

Each access modes has an associated cost which offset equipment, space rental, and staff overhead. The staff overhead includes the time and resources necessary for creating the analytical file, monitoring progress, setting up equipment and data files, disclosure limitation review, and file management. Since these reflect varying demands on resources, accurate cost estimates cannot be given without complete knowledge of the proposed research. The cost per project (or creation of an analytic file) is given below:

Management $750 per day
Additional fees charged as needed for file creation and for special handling, such as the merging of additional data or creating custom file formats, and transferring data between different modes.

On site $300 per day
Time on-site in the RDC can be scheduled in daily increments but the first visit to the RDC requires a 2 day minimum. Scheduling time at the RDC is on a first-come, first-served basis.

Staff- Assisted $750 per day

Remote $750 per month

Payment is expected in advance of the use of the RDC. Payments can be made in the form of a check, money order, electronic payment or Interagency Agreement.

Checks and money orders should be made payable to “DHHS Statistical Services” must be received 7 business days prior to the scheduled start date of use of the RDC.

Payments should be mailed to:

Research Data Center
Attn: Peter Meyer
National Center for Health Statistics
3311 Toledo Road
Hyattsville, MD 20782