

## NEW RESEARCH OPPORTUNITIES

Negasi Beyene, John Horm, Deanna Dick. National Center for Health Statistics.  
Negasi Beyene, NCHS, 6525 Belcrest Rd, Room 700, Hyattsville, MD 20782.

**Key Words: Survey, Research,  
Health Statistics.**

### Abstract

The National Center for Health Statistics (NCHS) has developed a Research Data Center (RDC) which allows researchers and data users to access internal data files from its numerous surveys which have not been available to the research community until now. The internal files contain lower levels of geography such as state, county, census tract, block-group, or blocks, depending on the survey. Examples of data system that are available through the RDC include the National Health Interview Survey, the National Health Nutrition Examination Survey, the National Hospital Discharge Survey, the National Survey of Family Growth (the NSFG contextual data files consist of the survey data and about 1,300 contextual variables and is available through the RDC) and others. Research may use the files to merge other contextual data from the Census Bureau, the Area Resource File, or other data collected or provided by the researcher ( air pollution data, state, county, or local or ordinances, reimbursement policies, medical facilities, etc.) to perform contextual analysis while maintaining respondent confidentiality. because of the confidentiality constraints NCHS has not been able to release survey data with lower levels of geography to its data users which has limited the amount and types of research projects that could be undertaken with its data systems. The development of RDC begins an exiting new era for NCHS and its data users.

### Introduction

The National Center for Health Statistics (NCHS) is part of the Centers for Disease Control and Prevention, US Department of Health and Human Services. NCHS is the Federal Government's principal vital and health statistics agency. NCHS works closely with other Federal agencies as well as researchers and academic institutions to develop data systems designed to address current and future health data needs. Since 1960, when the National

Office of Vital Statistics and the National Health Survey merged to form NCHS, the agency has provided a wide variety of data with which to monitor the nation's health. Since then, NCHS has received several legislative mandates and authorities. The NCHS mandate from Congresses the full spectrum of concern in health field from birth to death, including overall health status, life style and exposure to unhealthful influences, the onset and diagnosis of illness and disability, and the use and financing of health care and rehabilitation services. The key milestones in the evolution of NCHS authorities includes the NCHS confidentiality protections and a cooperative system for collecting national statistics were legislated in 1970. The Heart Disease, Cancer, Stroke, and Kidney Disease Amendments of 1970 established confidentiality provisions for national health surveys that remain ( with minor amendments) in current authority ( section 308d). These amendments also called for the design and implementation of a cooperative system for Federal and State level data collection. It also provided for the annual collection of data from records of birth, deaths, marriages, and divorces. In addition, the Act called for the Center to undertake and support research, demonstrations, and evaluations regarding survey methods and to provide technical assistance to State and local jurisdictions. These data are used by policymakers in congress and administration, by medical researchers, and by others in the health community.

NCHS is located in Hyattsville, Maryland, with offices in Research Triangle Park, North Carolina, and with a CDC-liaison office in Atlanta, Georgia. The continuing demand for analysis that require restricted data with lower level of geography such as states, counties and smaller areas, but without confidential identifiers such as names or social security numbers, has been the impetus for creation of Research Data Center (RDC) located at NCHS headquarters in Hyattsville, Maryland. Designed for the researchers outside NCHS, the RDC allows access to data that can not be released for public use of confidentiality (disclosure rules and regulations).

## NCHS Research Data Center

Prospective researchers must submit a research proposal that will be reviewed and approved by a committee, criteria for approval include consistency with the mission of NCHS; availability of RDC resources; and the feasibility of project. It is expected that the user will develop the research proposal in conjunction with RDC staff in order to minimize the time required for review. In addition to signing confidentiality agreements, researchers with approved projects must complete their work using the facilities located within the RDC. Researchers can supply their own data to be merged with NCHS data sets. Analytic files will be created by the RDC staff. Files, including merged researcher supplied data, will be only available to the originating researcher unless written permission is given to allow access to others.

### Data at the RDC

The RDC has a repository of data that includes the:

- National Health Interview Survey (NHIS), with state and county level identifiers.
- National Health Interview Survey polio data for 1994-1995.
- National Survey of Family Growth (NSFG) with contextual data added, 1995;
- Third National Health Examination Survey (NHANES III), with county level identifiers;
- National Hospital Discharge Survey (NHDS), with state identifiers.

However, any other NCHS data can be made available with restricted data included for example, the Longitudinal Study on Aging or Vital Statistics data can be utilized with restricted data such as lower levels of geography or sample design information incorporated into the data file.

*Example of data collected are:* the NHIS is a large sample of about 40,000 households( about 110,000 persons) and collects a wide range of information on both individuals and families including health status measures, acute and chronic conditions, health insurance, and family resources; The National Health and Nutrition Examination Survey (NHANES), a sample of over 30,000 persons which conducts both in-person interview and medical examination consisting of numerous test and measurements on its subjects; death certificate information on all deaths occurring in the United States each year; information on all births in the United States; the National Survey of Family Growth (NSFG) which

asks questions related to reproductive issues on about 10,000 women of childbearing age. Data from these and other NCHS data collection system are used for research, policy, and programmatic purposes at the national level.

The researchers can supply their own data which can be proprietary data collected and "owned" by the researcher. It can be other publically available information such as Census data. Whatever the case, early consultation with RDC staff is of utmost important to insure the data can be successfully merged with NCHS data and the format is consistent with NCHS data. All data manipulations are done prior to the official arrival of researcher in regards to the beginning of the analysis. Any intermediate files used in linking the data sets will be destroyed after the final data file is created and can not be made available to the researcher. The final, merged data set belongs solely to the researcher, who can give access to other with written consent only.

### Operations

Researcher have a choice of different modes for research in the RDC: on-site guest researcher, remote access, or subcontracting with RDC to perform the analysis. In all cases, the staff of the RDC will construct all the necessary data files, including merged NCHS and user-supplied data, before the researcher can begin analysis. As an on-site guest researcher, the user has access to the standard software (SAS and Stata) and other statistical packages at his/her request with sufficient lead time to acquire and load the software. The output from the statistical packages is subject to disclosure review based upon the guidelines published in the "NCHS Staff Manual on Confidentiality" essentially RDC staff will examine the output printed by the procedures in the statistical packages and will allow the printed material passing the disclosure review to be taken off site. All materials can be examined on site, but only that material that has passed the disclosure review can be taken away. Because the process requires staff oversight, RDC is open for researchers only during working hours whose time frames will vary relative to staff availability, but will usually mean from 8:00 AM to 5:00 PM during weekdays operations. The RDC is not open on weekends, holidays or on occasions when staff are not available. If remote access option is chosen, the researcher can submit analytical computer programs through the Internet using an email address for the RDC. After the programs execute, the output is returned to the user's registered email address. Only SAS statistical software package can be used with limitations that prevent the disclosure of restricted information. The limitations include procedures such as PROC TABULATE and PROC IML

which are not allowed; LIST and PRINT are not available along with any of listing individual cases; R\_\_\_, FIRST., and LAST., functions are not allowed again with no individual case listing. If any cell has less than 5 observations, that cell will be suppressed along with other, additional cells. The SAS job log will be scanned for error conditions that spawn case listings. The log will not display any case level information. Finally, if necessary, a manual disclosure review will be conducted to insure confidentiality. The last option for analysis, subcontracting with the RDC staff to perform programming tasks, relies upon accurate specification from the researcher. However, it is highly recommended that the specifications include input from RDC staff to minimize any potential misunderstandings regarding the specifications.

The RDC is fee-for-service entity, dependent on user support to remain viable. Fees charged the user pay for equipment, space rental, staff overhead and setup costs. A overall charge is estimated for each research project. The accuracy of the cost is dependent on estimating the resources required for each specific project. With the close collaboration between RDC and researcher, the estimated cost should be accurate. There is a setup of \$500 per day of effort for custom file creation that uses researcher data merged with NCHS data or creating custom file formats for any data set. The cost per project is as follows:

Onsite = \$1000 per week;

Remote Access : NSFG-CDF = \$500 per year,  
NHIS- Polio = \$500 per year,

Files less than 130,000 records = \$500 per month,

Files over than 130,000 records = \$1,000 per month.

The charges for onsite and remote access include equipment, work space and NCHS staff time for monitoring, providing technical assistance, disclosure limitation review and file management.

### **Research Proposal**

To reduce the time required for the review process, a meeting between the researcher and Director and staff of the Research Data Center should be made early in the application process. The development of the research proposal would be best accomplished in conjunction with RDC staff and hasten the acceptance of proposal. The research proposal may include portions of doctoral proposals or grant application, but should include the details as to:

- 1) Why publically available data is not sufficient to accomplish the research objectives.
- 2) Full personal and institutional affiliation.
- 3) Dates of proposed visits to the RDC.

- 4) And an estimation of total hours expected to be working in the RDC are required.

The researcher should state the specific NCHS data required including the years or year needed by the analysis and the variables necessary to conduct the research. If the researcher supplies their data, documentation of the data sets is necessary for the RDC staff to merge the researcher-supplied data with the NCHS data. In addition, software required for the analysis must be specified. If the software differs from standard packages SAS for windows or STATA, the RDC must be notified quickly to insure the availability of the software for analysis. The researcher is required to sign the “ Agreement Regarding Conditions of Access to Confidential Data in the RDC of the National Center for Health Statistics” and the “Research Affidavit of Confidentiality”. The acceptability of the proposal is based upon its consistency with the NCHS mission; the availability of RDC resource; and the feasibility of the project.

### **User Supplied Data**

The NCHS RDC allows users to supply their own data to be merged with NCHS data sets. The user-supplied data may consist proprietary data collected and ‘owned’ by the user or other publicly available data obtained by the user such as Census Data. Users expecting to use merged files are responsible for interacting with RDC staff to ensure that their data can be merged with the NCHS data and the format of the data is consistent with the NCHS data. The merging of user-supplied with NCHS data sets will be done by NCHS RDC staff prior to the arrival of user. Files and information used in linking the internal and external files will be destroyed after the merge and will not be made available to the users. The NCHS RDC, as a policy, will not release or give access to any user-supplied files to anyone without a written consent of the user who supplied the files. The NCHS RDC periodically creates and maintains backup copies of all computer files. The backup tapes are stored in a secure storage area accessible by NCHS RDC staff only. These backup files will contain user-supplied data as well as merged files and only be destroyed upon a written request of the user.

### **Examples of Jobs processed at RDC ( from January 1 to July 20, 1999)**

I. Topic : Do Doctors perform “Defensive Cesareans ? “

Overview : One topic of much current policy is

the extent to which physician's practice "defensive medicine", performing excessively expensive Procedures out of fear of malpractice suits. To re-examine this issue, by examining the effect of state tort reforms designed to limit malpractice risk on the use of cesarean section delivery. There was a wave of such reforms in the mid 1980.

NCHS Data Used? National Hospital Discharge Survey (NHDS).  
Years of Data Used: 1980 through 1992, inclusive.

User's Data Merge with NCHS's: Yes.

Method of Access to NCHS Data: Remote and onsite.

Statistical Software Used: SAS

## II. Topic : Economic Model to Explain the Incidence of Sexual Activity, Contraceptive Use, STD's, and Pregnancy Among Teenage Girls.

Overview: National Survey of Family Growth Data provides socio-demographic information and reports of the sexual histories of these women. Researcher focus on the effects of a number of policies measured at the state-level. These include:

- parental notification of consent laws.
- medicaid funding of abortion.
- welfare generosity.

In order to assess the impact of these policies, researcher have compiled a state-level data set containing these measures. Which must be merged with respondent data using a state of residence.

*Note:* To protect the confidentiality of the subjects, we assign randomized number for each state and county.

NCHS Data Used: National Survey of Family Growth (NSFG).

Years of Data Used: 1995.

User's Data Merge with NCHS's : Yes.

Method of Access to NCHS Data: Remote Access

Statistical Software Used: SAS

## III. Topic: Assessing the Impact of Family Planning Services on Individual Contraceptive Use and Pregnancy : A Multilevel Analysis.

Overview: The central question of this research has both policy and theoretical significance. In regard to policy, the primary issue is whether dollars spent on making family planning clinics available and improving their

service (for example, increasing the provider/client ratio). Of related social significance, examine the determinant of contraceptive use, undesired pregnancies and relative significance of family planning clinics and providers. Whose fertility is targeted by federally and state-funded clinics.

NCHS Data Used: National Survey of Family Growth.

Years of Data Used: 1995.

User's Data Merge with NCHS: Yes.

Method of Access to NCHS Data: by RDC Staff.

Statistical Software Used: HLM (Hierarchical Linear Modeling).

## IV. Topic: Cost Effectiveness of Providing Health to Children of Working Poor: A Look at Access, Utilization and Outcomes.

Overview: To describe the socio-economic and health status characteristics of the uninsured, Children's Health Insurance Program (CHIP) eligible population in the United State and Illinois; and to simulate expected changes, along with cost and benefits of those changes, in health care access, utilization, and selected outcomes based on enrollment in the CHIP program.

NCHS Data Used: The National Health Interview Survey (NHIS).

Years of Data Used: The last two quarters of 1993 and all of 1994.

User's Data Merge with NCHS: No.

Method of Access to NCHS Data: Remote Access.

Statistical Software Used: SAS.

## Conclusions

The Research Data Center at the National Center for Health Statistics has taken a number of new and creative measures to enhance access to type of data that have not until now been available to researchers outside the NCHS . It is our belief that as additional researchers take advantage of this new opportunity, new and exciting discoveries in the health and health status of populations will emerge.

The NCHS continues to develop, modify and enhance its research data center to take advantage of new technologies and statistical development in the fields of data access and release.

## References

- 1) National Center for Health Statistics Approaches  
Micro data. John Horm, National Center for  
Health Statistics, Hyattsville, MD, 1999.
- 2) Producing a Public Use File: A Case Study.  
Kenneth Rasinski, Jeffrey Timberlake, Lisa  
Lee, Javier Porras, National Opinion Research,  
Center and Mulrow, Jeri Ernst & Young.  
In press, 1997.
- 3) National Health Interview Survey State Files:  
Compromises, Tradeoffs, and Concessions.  
John Horm, Al Zarate, and Steve Botman.  
Presented at the Public Health Conference on  
Records and Statistics, Washington, DC., 1997.