Accessing Restricted National Center for Health Statistics Data through the Research Data Centers

Ajay Yesupriya, MPH
Center for Social and Demographic Analysis (CSDA)
3/23/2017
Outline

• Brief background on the National Center for Health Statistics (NCHS)
• Discussion of the NCHS Research Data Center (RDC)
  • Restricted data/variables
  • Proposal process
  • Other requirements
• Accessing restricted NCHS data at a Federal Statistical RDC (FSRDC)
  • Description of the FSRDC System
  • Process for accessing datasets and conducting analyses at NYRDC
• Questions
Background on NCHS
Mission of NCHS

• To provide data to identify and address health issues
  • Document the health status of the U.S. population
  • Monitor trends in health status, associated risk factors, and health care delivery
  • Guide public health and health policy decisions
NCHS Surveys

• Population Surveys
  • National Health Interview Survey
  • National Health Nutrition and Examination Survey
  • National Survey of Family Growth
• Vital Records
  • National Vital Statistics System
  • National Death Index
• Telephone Surveys
  • State and Local Area Integrated Telephone Survey
  • National Immunization Survey
• Health Care Surveys
  • National Ambulatory Medical Care Survey
  • National Hospital Discharge Survey
  • National Survey of Residential Care Facilities
  • National Home and Hospice Care Survey
  • National Nursing Home Survey
  • National Home Health Aide Survey
Types of Data and Data Access

• Public
  • Data files: collapsed categories, top and bottom coded
  • Reports

• Restricted: Indirect Identifiers
  • Items that when combined with other information (from NCHS or other sources) could lead to identification
  • Available through the RDC
NCHS Research Data Center
NCHS Research Data Center

• The NCHS RDC was created to provide a secure environment for researchers to access restricted data in order to answer important public health questions

• The RDC protects restricted data by requiring:
  • A research proposal
  • Analysis conducted within a secure environment
  • Confidentiality training and paperwork
  • Output review
Examples of Restricted Variables

• Geography, exact dates, more detailed versions of variables (race/ethnicity, industry/occupation)
  • NHIS: Preliminary Quarterly files
  • NHANES: Sensitive information for the youth population (e.g. conduct disorder, drug use, sexual behavior)
  • NSFG: Contextual data, Interviewer data
  • SLAITS: Links between the surveys
  • NHCS: Provider/facility characteristics, contextual information
Data Linkage

• Linked data enable broader analyses of factors that influence health and health outcomes
• Many surveys have been linked with administrative data
  • The National Death Index (NDI)
  • Claims data from the Centers for Medicare & Medicaid Services (CMS)
  • Benefit history data from the Social Security Administration (SSA)
  • Housing assistance information from the Department of Housing and Development (HUD)
Data Hosting

• The NCHS RDC hosts datasets from other groups in HHS:
  • United States Cancer Statistics
    • NPCR/SEER Restricted Access Datasets
    • Comparative Effectiveness Research (CER) Dataset
  • National Assisted Reproductive Technology Surveillance System (NASS)
  • National Sample Survey of Nurse Practitioners (NSSNP)
Reasons to Use Restricted Data

• To answer main research questions
• To control for confounders/examine effect modifiers
• To combine datasets/link in external data (e.g. ACS, researcher compiled state policy information)
• To calculate standard errors, you may need restricted information about the sampling design
RDC Proposal Format

• Abstract
• Research Question
• Background
• Public Health Benefit
• Data Requirements
  • Survey, Years, Files
  • Restricted Data
  • Non-NCHS Data
  • Merge variables
• Methodology

• Output
  • Table Shells
  • Presentation of Results
• Data Dictionary
  • Public Data
  • Restricted Data
  • Non-NCHS Data
• References
• Resumes/CVs
RDC Proposal Review Process

• Conducted by a Review Committee
  • RDC Analyst (primary point of contact at NCHS RDC)
  • RDC Director or designee
  • Data System Representatives (programs that produce the data)
  • Confidentiality Officer
• Generally takes 6-8 weeks
Writing the proposal

• Data Requirements
  • List every restricted variable (and any non-NCHS variables)
  • Explain how you will use these variables
  • Describe the merge procedures in detail

• Output
  • Must match research question
  • Must be representative of items that will appear in your publication

• Data Dictionary
  • Only include variables that you need
Output and Disclosure Review

• No individual level data can leave RDC facilities
• Some results may be suppressed due to disclosure risks (e.g. cells <5, minimum and/or maximum values)
• Output must match the research questions/output suggested in the proposal
• Avoid requesting release of intermediate output (results that will not appear in your publication)
• Adhere to any additional requirements specified in the approval email
• Approved output will be emailed to you
Accessing NCHS Data at the New York Federal Statistical RDCs
Description of FSRDCs

- Network of 24 secure facilities managed by the Census Bureau that provide access to restricted use datasets for statistical purposes
- Partnerships between federal statistical agencies and research institutions
  - Agency for Healthcare Research and Quality
  - Bureau of Labor Statistics
  - Census Bureau
  - National Center for Health Statistics
  - Many other agencies also allow access to surveys they co-sponsor
New York Research Data Center (NYRDC)

• Consortium of 12 research institutions (University at Albany is a member)
• Operates three facilities
  • Baruch College-CUNY (New York, NY)
  • Cornell Institute for Social and Economic Research (Ithaca, NY)
  • Yale University (New Haven, CT)
• Each facility has a primary point of contact who administers the location on behalf of the partner statistical agencies (FSRDC Administrator)
• Membership in the NYRDC Consortium covers access fees for faculty members, graduate students, and researchers affiliated with member institutions (usual project fee to non-member institutes is $20,000)
NYRDC NCHS Project Fee Grants

• NCHS has data management fees of $750 per day (creating the analytic data set and transferring file to FSRDC)

• Researchers at institutes that are members of the NYRDC are eligible for NCHS Project Fee Grants (maximum $3,000 per project)
  • Must have approved NCHS RDC project
  • Must have attempted but were unable to obtain grant or other funding to cover NCHS fees

• Researchers may receive a project grant once every two years

• Each member institution limited to two grants per year
NCHS RDC Analyst and FSRDC Administrator

RDC Analyst (NCHS)
• Answers questions about NCHS RDC process and datasets
• Facilitates review of your proposal
• Creates your analytic data set
• Accepts the set up fee payment
• Accepts your NCHS Confidentiality requirements
• Transfers your dataset to FSRDC
• Reviews your programs and output for disclosure risk

RDC Administrator (FSRDC)
• Answers logistical questions about the FSRDC
• Ensures you have completed all the Census Bureau Security requirements (Special Sworn Status)
• Facilitates transfer of output to NCHS
• Schedules time in the lab
Working in the FSRDC

• Only researchers with approved projects and Special Sworn Status are allowed in the RDC

• Researchers working on approved NCHS RDC projects can only work when the FSRDC Administrator is present

• Researchers work on thin-client workstations which provide access to statistical software (e.g., SAS, Stata, R) as well as word processing and spreadsheet programs

• No information can be removed without review for disclosure risks
Working with the NCHS analyst

• All materials must be reviewed by the NCHS RDC Analyst for disclosure risk
  • Programs, documentation, and other materials
  • Statistical output
  • Approved output will be emailed to you within 3 weeks
• Manuscripts must be reviewed by your NCHS RDC analyst before submission for publication
Links to resources

• NCHS RDC Website: https://www.cdc.gov/rdc/
  • Proposal process (includes template and example proposal): https://www.cdc.gov/rdc/b3prosal/pp300.htm
  • Fees: https://www.cdc.gov/rdc/b5aprovproj/ap540.htm

• FSRDC Website: https://www.census.gov/fsrdc
  • Available Data: https://www.census.gov/about/adrm/fsrdc/about/available_data.html

• NYRDC Website: https://nyrdc.cornell.edu/
  • Contacts: https://nyrdc.cornell.edu/contacts/
  • Policies and Procedures: https://nyrdc.cornell.edu/policies-and-procedures/