

# Data Release through Research Data Centers: The NCHS Experience

*Jennifer H. Madans*

*Associate Director For Science*

*NCHS, CDC*

# Why create a Research Data Center?

As a federal statistical agency, NCHS's mission has two primary components: to collect and disseminate data related to the health of the U.S. population

Data dissemination flexibility is maximized with the release of unrestricted, public use files

However, NCHS's authorizing legislation mandates data confidentiality

# NCHS 308(d) Confidentiality Statute

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 section 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 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in the course of health statistical or epidemiological activities under section 304 or 306, such information may not be published or released in other form if the particular establishment or person supplying the information or described in it is identifiable unless such establishment or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 to its publication or release in other form.

# NCHS 308(d) Confidentiality Statute

.... such information may not be published or released in other form if the particular establishment or person supplying the information or described in it is identifiable unless such establishment or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 to its publication or release in other form.

# What is a Research Data Center?

- A secure, controlled environment where researchers can be allowed access to data files that cannot be released as Public Use Files
- Confidentiality is protected because researchers cannot link files to other data sets that would allow subjects to be identified
- RDC staff primarily have stewardship functions

# History of the NCHS RDC

- The NCHS RDC was opened in 1998
- Modeled after the Census Bureau Centers
- Policies and procedures were developed to assure confidentiality with procedures posted on the NCHS website

# NCHS RDC Procedures

## Research Proposals

### Contents:

Contact related information

Key study questions/Public health benefits

Cases to be included in analytic file

Variables to be included in analytic file

Why publicly available data are insufficient

Analytic strategy/statistical methods/software

Output to be reviewed for non-disclosure

# NCHS RDC Procedures

Proposal evaluation:

Consistency with the NCHS mission

Feasibility

Requested data are needed to meet  
analytic objectives

Likelihood that results will be releasable  
outside the RDC

Scientific merit is not an evaluation criteria



# NCHS RDC Procedures

Proposals reviewed by:

- Director of the RDC
- RDC staff liaison
- NCHS Confidentiality Officer
- Director (or designee) of the Data System

# NCHS RDC Procedures

Why proposals are important:

- Guide the RDC staff in creating the analytic files – they need to be explicit
- Facilitate disclosure review and ensure confidentiality protection

# NCHS RDC Procedures

## Affidavit of confidentiality

All users must sign the Affidavit which includes a promise not to use any method to attempt to identify respondents.

# NCHS RDC Procedures

## Analytic File Creation

RDC staff create analytic files based on requirements set out in approved research proposals

User supplied data: Data files that are to be merged with NCHS data must be provided before visiting the RDC and all merging is done by RDC staff.

# NCHS RDC Procedures

## Confidentiality protection

Workstations do NOT have e-mail or internet access.

Only access to printer is through RDC staff.

Only approved materials can be taken into the RDC.

# NCHS RDC Procedures

- Confidentiality protection

All output and/or materials removed from the RDC are subject to **disclosure limitation review**.

Analytic files are created to protect confidentiality

# How does RDC operate?

- On-Site Access
- RDC staff performs Analysis
- Remote Access

# NCHS RDC

## Hardware and Software

### Hardware

Pentium computers with Windows 2000

### Software

SAS

Sudaan

Fortran

HLM

Stata

Limdep

text editors/viewers



# Number of Projects Conducted at the NCHS RDC—1998-2003

|      |    |
|------|----|
| 1998 | 7  |
| 1999 | 9  |
| 2000 | 12 |
| 2001 | 19 |
| 2002 | 36 |
| 2003 | 38 |

# Broadening of the NCHS RDC beyond NCHS Data

Other data sets resident in the NCHS RDC:

- National Evaluation of Welfare-to-Work Strategies Survey (NEWSS) – ASPE/DHHS 2001 (5 projects)
- Leavers and Diversion Studies – ASPE/DHHS 2001 (7 projects)

# Broadening of the NCHS RDC beyond NCHS Data

Data sets that are under negotiation:

- National Hospital Prevalence Study --  
NIAAA
- Health and Retirement Survey

# Implications for RDC policies and procedures

Need to incorporate requirements specific to non-NCHS data into RDC procedures

Nature of the disclosure Review does not change





# Current Practices for VSD Data Sharing

NIP used the NCHS RDC to provide access to VSD data prior to March 2004

NIP followed NCHS's general policies for accessing data in the RDC

NCHS assumed responsibility for data sharing activities for the VSD data set in March 2004 – no change in how researchers would function when visiting the RDC

# Current Practices for VSD Data Sharing

VSD specific requirements as negotiated with the MCOs are still in effect

The receipt of the VSD data and the ongoing negotiations with other data programs suggested it was time to review and clarify RDC procedures



# Revision of RDC Procedures

Current procedures were developed for use with NCHS data sets and there are references to NCHS and NCHS data sets in the text

The revised procedures will:

- apply universally to all data sets resident in the RDC
- address data set specific requirements in appendices

# Revision of RDC Procedures

The revised procedures will also:

- provide more specificity about file creation
- clarify how access to data sets is controlled so that confidentiality can be better protected

# Revision of RDC Procedures

The revision process should:

- better integrate non-NCHS data sets into RDC processes
- provide more information to users about what to expect when using the RDC