



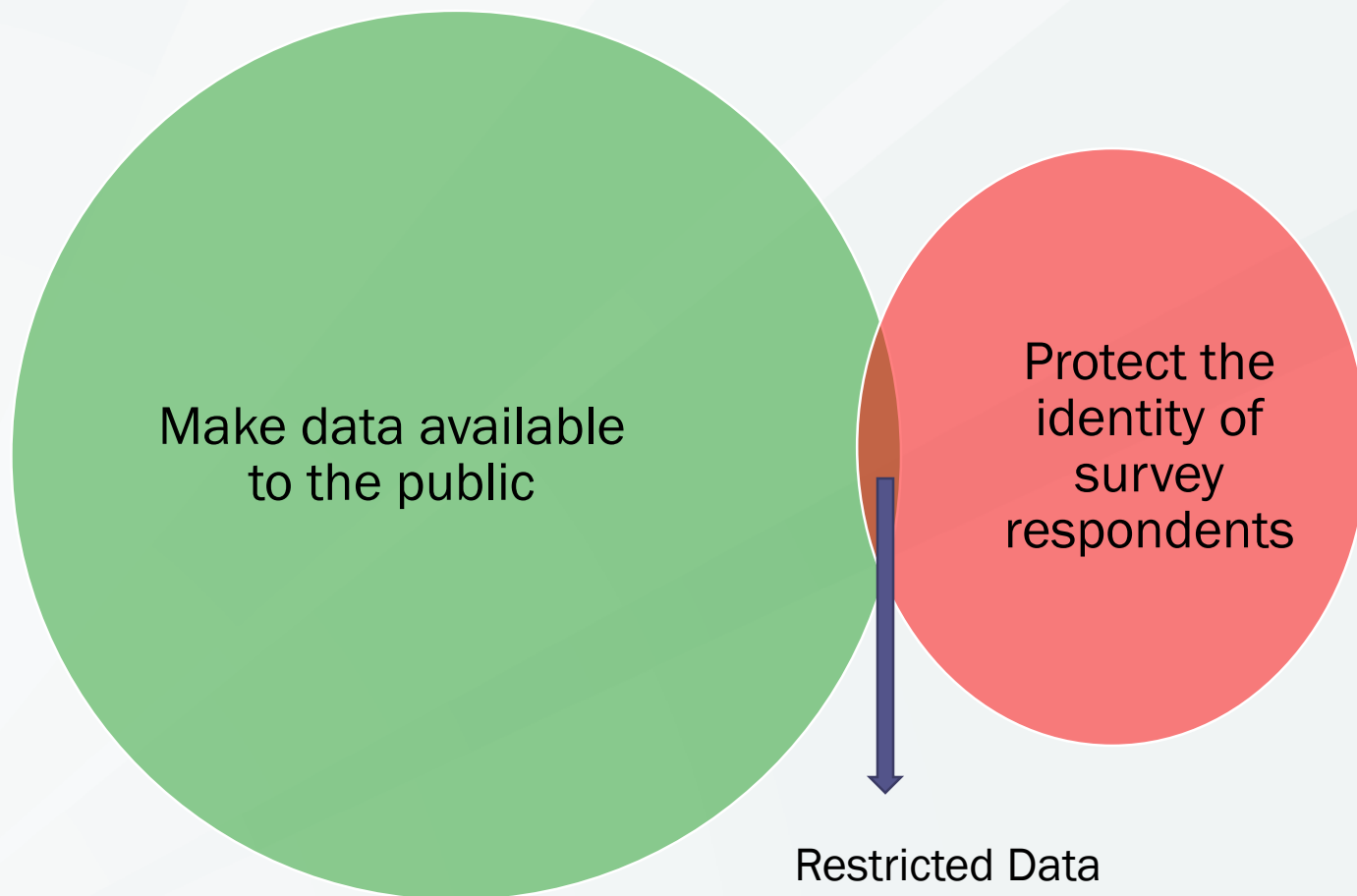
# Tips for Writing a RDC Proposal

**Pat Barnes**  
**Health Scientist**

**Tabatha McNeill**  
**Public Health Analyst**

# The NCHS RDC

# What is Restricted Data?



# Restricted Variables

## NHANES

- Geography
- Youth Variables
- Genetics

## NHIS

- Country, state, and day of birth
- Detailed race/ethnicity
- Industry and occupation codes
- Detailed income

## NSFG

- Interviewer Data
- Geography
- Detailed race/ethnicity

## NVSS

- Exact date of vital event
- Geography

## Health Care Surveys

- Number of beds and residents
- Entry wages
- Continuous ages
- Geography

## Linked Data Products

- Mortality data linked to NHIS, NHANES III, NNHS
- Retirement and Disability Data (SSA) linked to NHIS, NHANES III

# Reasons to Use Restricted Data

- To answer your main research question
- To control for confounders/examine effect modifiers
- To combine datasets/link in external data ( e.g. ARF, researcher compiled state policy information)
- To use as selection criteria for a study
- To access a more detailed version of a variable
- To make accurate estimates, you may need restricted information about the sampling frame

# RDC Proposal

# RDC Proposal Overview

- A researcher submits full research proposal
  - analytic plan
  - examples of desired output
- Provide specifications for creating analytic files
- Proposals are reviewed (6-8 weeks) by the RDC Review Committee
  - RDC Director or designee
  - RDC Analyst
  - Data System Representative (program that produces the data)
  - Confidentiality Officer

# Before Writing the Proposals

- Review Analytic Guidelines
  - Identify any analytic concerns that may affect your output
    - Can you combine survey, years, or files?
- Become familiar with the data
  - Compile a dataset that includes only the variables you need to answer your research question
  - Complete any analysis that does not need restricted variables



# Proposal Format Page 1

Proposal Format (Revised 2/04/2014)  
Project Title, Researcher

## RDC Research Proposal

General Information	
Date:	
Title of Project:	
NCHS Data System and Years:	
Non-NCHS Data Files:	
Mode of Access: (Check all that apply)	<input type="checkbox"/> NCHS RDC, Hyattsville, MD (Washington, DC-metro) <input type="checkbox"/> NCHS RDC, Washington, DC (Government Only) <input type="checkbox"/> NCHS RDC, Atlanta, GA <input type="checkbox"/> Remote Access (ANDRE) <input type="checkbox"/> Census RDC, specify: _____
Statistical Software: (Check all that apply)	<input type="checkbox"/> SAS/Sudaan <input type="checkbox"/> Stata <input type="checkbox"/> Other, specify: _____ * Remote access users can only use SAS/Sudaan
Proposed Start Date:	
Funding Source:	
Billing Address: (include contact person)	

Complete as applicable for your project. Everyone listed in this section will need to submit a C.V. and if approved, must complete the Confidentiality Orientation. There can only be one ANDRE programmer.

Research Team		
	<b>Primary Investigator</b>	<b>Co-Investigator</b>
Name		
Email		
Phone		
Institution		
Mailing Address		
US Citizen? Y or N		
	<b>Programmer</b>	<b>Programmer</b>
	<input type="checkbox"/> On-site or <input type="checkbox"/> ANDRE (account holder)	<input type="checkbox"/> On-site
Name		
Email		
Phone		
Institution		
Mailing Address		
US Citizen? Y or N		
	<b>Advisor (For Students and Post-Docs)</b>	<b>Other, specify:</b>
	<input type="checkbox"/> <a href="#">RDC-Student-Advisor Form</a>	
Name		
Email		
Phone		
Institution		
Mailing Address		
US Citizen? Y or N		

# Writing the Proposal

## General Information

- Date field should change with every iteration of the proposal
- Select only one (1) mode of access

## Research Team

- Only people who will come in direct contact with the data
- Exceptions: Student Advisors and Principal Investigators
- Non-citizens will have additional clearances

# Proposal Format Pages 3-4

**A. Abstract:** Please limit the abstract to 300 words.

**B. Research Question:** Include study purpose, hypotheses, goals, or research questions.

**C. Background:** Include a short literature review, no more than 2 pages, focusing on papers that discuss your topic or address the methodology that you plan to use. Please limit your reference list to 10 items or less.

**D. Public Health Benefit:** In one paragraph, how does your research benefit public health?

**E. Data Requirements:**

1. **Survey, Years, Files:** For example,  
NHS 2005-2007 Household, Person and Sample Adult Files  
NAMCS 2005-2006 Provider and Patient Visit Files  
NHANES 2005-2006 Examination and Demographic Files  
Medicare Denominator 2000-2007  
Linked Mortality
2. **Restricted Variables:**
  - i) List the restricted variables you are requesting.
  - ii) Explain how you will use these variables.  
If you are requesting geographic variables, please indicate how they will be used to merge files, analyze the data, and present the results.  
Are you requesting geography to merge non-NCHS data?  
Can the variables be dropped after the merge? If not, how will they be used in analysis?  
Can randomized versions of the variables be substituted?
3. **Non-NCHS Data:**  
Will you provide data from another source (such as Census or EPA)? If yes, please describe the source, list the files, and provide a general description of the data.
4. **Merge Variables:** Please describe the merge procedures needed to produce each dataset you will be analyzing. Leave blank if not applicable (NHDS, Mortality, Natality, and DHHS Hosted Data Users).
  - i) What variable(s) will be used to merge the public and restricted data files?
  - ii) What variable(s) will be used to merge the NCHS data files with any non-NCHS data?  
Note: The RDC Analyst will complete each merge for you. Your RDC Analyst may request additional information to make sure that the merge meets your needs.

**F. Methodology:**

1. **Unit or Level of Analysis and Subpopulation(s):**  
There can be many levels of analysis: be as detailed as possible. A common example for an analysis of NHANES is where the unit of analysis is the person while the subpopulation is adults ages 18-64. A common example involving geography is when you aggregate persons to the state level so you can compare states with policy A to states with policy B.
2. **Analysis Plan:** Please provide an overall analysis plan that specifies what analytic procedures or models you will use, such as prevalence estimates, logistic regression, or log-linear modeling, or list specific statistical package procedures.
3. **Complex Survey Design:** Please consult the public documentation and analytic guidelines for the survey you will use, and indicate how you will address sample weights, design variables, and other adjustments for the use of complex survey data, if applicable.

3

**G. Output:**

1. **Overview:** Please describe any output that you would like to take out of the RDC; please be detailed as this section helps the Review Committee assess disclosure risk.
2. **Examples/Table Shells:** Include detailed examples of table shells, models, and/or graphs. Please indicate the subsample or unit of analysis used in each type of table, model, or graphs.
3. **Presentation of Results:** Will you present the results in a report, publication in a peer-reviewed journal, presentation at a scientific meeting, used for internal policy analysis, etc.?

**H. Data Dictionary:** See instructions and examples for [creating the data dictionary](#).

**NHDS, Mortality, Natality, and DHHS Hosted Data Users:**

Please provide a data dictionary that includes all the variables you would like extracted for your analytic data set.

**Other NCHS Data Users (including the linked data products):**

If you are requesting data from any other NCHS data system, please include a public, restricted, and non-NCHS data dictionary.

**I. References:** Please limit the list to 10 items or less.

**J. Other Authors:** Please list the name and institution for anyone not listed in the Research Team (pg. 1) who will contribute to publications resulting from this project, but will not come into contact with the data.

**K. Resumes/C.V.:** Please include a 2-page C.V. for each member of the research team listed in the initial chart (not as attachments).

4

# The Proposal

Proposal Component	RDC	Researcher
Data Requirements	<ul style="list-style-type: none"><li>• Researcher knowledge of data</li><li>• Identify other variables that might be of interest</li><li>• Identify possible disclosure risk</li></ul>	<ul style="list-style-type: none"><li>• Helps with Data Dictionary component</li></ul>
Methods	<ul style="list-style-type: none"><li>• Identify possible disclosure risks</li></ul>	<ul style="list-style-type: none"><li>• Decreases time in the RDC (cost effective)</li></ul>
Output	<ul style="list-style-type: none"><li>• Identify possible disclosure risks</li></ul>	<ul style="list-style-type: none"><li>• Increases the chance your output will be released</li></ul>
Data Dictionary	<ul style="list-style-type: none"><li>• Used to create your final dataset</li><li>• Identify possible disclosure risks</li></ul>	<ul style="list-style-type: none"><li>• Used to create Researcher Supplied Dataset</li></ul>

# Writing the Proposal: Data Requirements

- Data Requirements Summary
  - Any “Yes” or checked responses should be addressed in later sections of the proposal
- Restricted Variables
  - Does the analytic guidelines support this type of analysis?
  - List every restricted variable
  - Explain how you will use the variable
  - Keep or drop after merge
- Non-NCHS Data
  - Must be complete for the United States
- Merge Variables
  - Instructions for your RDC analyst to create your merge
  - For complex merges consider providing your own programming code

# Writing the Proposal: Methodology

- Analytic Guidelines
  - Does the analytic guidelines support this type of analysis?
  - Be sure to address complex survey design
- Statistical Software and Methods
  - Lab computers have SAS/SUDAAN, Stata, and R
  - Extensions must be tested and approved
  - Avoid methods that will result in small cells

# Writing the proposal: Output

- Must match research questions
- Provide examples of what you are expecting your output to look like
  - Output approved for release must match these examples
- Be representative of items that will appear in your publication

# Writing the proposal: Output Recommendation

## STROBE statement - checklist of items that should be included in reports of observational studies

Main results	Give unadjusted estimates and if applicable, confounder-adjusted estimates and their precision. Make clear which confounders were adjusted for and why they were included
	Report category boundaries when continuous variables were categorized
	If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period

Reference: *STROBE statement – checklist for cohort, case-control, and cross sectional studies (combined)*. Retrieved from <http://www.strobe-statement.org/index.php?id=available-checklists>



# Writing the Proposal: Data Dictionary

- 3 Parts to the Data Dictionary
  - Public
  - Restricted
  - Non-NCHS Data
- Indicate merge variable in each dictionary
- Multiple years may require a separate dictionary for each year
- Use proc contents from an already constructed public-use file

# RDC Research Example

- Bitler, M. and L. Schmidt (2006). "Health disparities and infertility: impacts of state-level insurance mandates." *Fertil Steril* 85(4): 858-65.

Proposal Component	RDC
Data Requirements	<ul style="list-style-type: none"><li>• National Survey of Family Growth (NSFG)</li><li>• '82, '88, '95, and '02 cycles</li><li>• Reason for restricted data: Needs state to merge policy variables</li></ul>
Methods	<ul style="list-style-type: none"><li>• Multivariate logistic regression</li></ul>
Output	<ul style="list-style-type: none"><li>• Summary statistics</li><li>• Odds ratios</li></ul>
Data Dictionary	<ul style="list-style-type: none"><li>• State identifiers (restricted)</li><li>• Use of infertility treatment (public -use NSFG)</li><li>• Access to infertility treatment (non-NCHS)</li><li>• State level demographics (non-NCHS)</li></ul>

# Other Considerations

- Be flexible when choosing an access mode
- Respond to questions or comments in a timely manner
  - Highlight all revisions
- There are costs associated with using the RDC
  - Fees only applied after proposal is approved
- Estimated 6-8 weeks for proposal review
  - Allow enough to complete your analysis

# Summary

- Review all documentation
- Use the public-use data first
- Be as detailed as possible when writing the data requirements, output and methodology sections
- Submit via email [rdca@cdc.gov](mailto:rdca@cdc.gov)



NCHS : [www.cdc.gov/nchs](http://www.cdc.gov/nchs)

NCHS RDC: [www.cdc.gov/rdc](http://www.cdc.gov/rdc)

Submit a proposal or questions: [rdca@cdc.gov](mailto:rdca@cdc.gov)