Opportunities for Research Using Restricted Data

Graduate Student Information Session

Mary Campbell, PhD
Director, TXRDC
Texas A&M University

Karin Johnson, PhD
Administrator, TXRDC
US Census Bureau
Welcome everyone!
FSRDCs (or RDCs) are research facilities

Collaboration among the Census Bureau, other federal statistical agencies, and universities to provide researchers with approved projects access to restricted-use data
What do RDCs offer?

• Secure computing lab
• Joint Statistical Partnerships (JSP)
• Managed by the Center for Enterprise Dissemination (CED) at Census Headquarters
• Locally run by an on-site Census Bureau employee, the RDC Administrator
What is the Texas Consortium?

• RDCs are similar to research libraries and labs
• These universities are members, providing researcher access to valuable resources
• Talk to me (Mary) about how to access the TXRDC if you’re from another institution!
Why we need RDCs

- Federal laws and regulations protect data confidentiality
- Federal employees and contractors access restricted data only under strict guidelines

→ FSRDC network is important for research analyzing restricted statistical data to benefit society
Advantages of RDC
Access to Researchers

• Restricted microdata
• Key variables not in public versions of datasets, or detailed versions of variables
• Full counts and/or larger samples
• Original responses
• Ability to link data
RDC-Based Research is Common in Many Fields

- Business, Trade, Finance, and Management
- Crime and Crime Victimization
- Demography and Population Studies
- Economics, Labor Markets, Entrepreneurship, Employment and Industry
- Education and Education Policy
- Hazard Reduction, Environmental Impact Assessment, Pollution Abatement
- Health and Wellbeing, Health Insurance, and Health Policy
- Housing, Housing Markets, and Residential Patterns
RDCs Serve Well ...
- Discipline-based substantive research
- Research plans with long time horizons

RDCs Serve Less Well ...
- Research with short time horizons
- Research producing special tabulations and maps
- Projects producing data products for public distribution
“What kind of project could I do in the RDC?”

• Race/ethnicity, segregation
  • Decennials, Current Population Survey (CPS), Amer. Comm. Surveys (ACSs)
• Social demography (e.g., migration patterns, education)
  • Master Address File (MAF), Survey of Doctoral Recipients
• Criminology
  • National Crime Victimization Survey
• Institutions, workplaces, production
  • Long. Business Database, Long. Employer-Household Dynamics (LEHD), Census of Mining, Manufactures
• Public policy
  • SNAP, SIPP (Survey of Income & Program Participation), American Housing Survey (AHS)
• Health, mortality
  • National Longitudinal Mortality Study
### Example 1 of Data Availability

**– Census Bureau Data**

<table>
<thead>
<tr>
<th>Data Source</th>
<th>Granularity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decennial Census</td>
<td>Block</td>
</tr>
<tr>
<td>American Community Survey (ACS)</td>
<td>Block</td>
</tr>
<tr>
<td>Survey of Income and Program Participation (SIPP)</td>
<td>Tract</td>
</tr>
<tr>
<td>Current Population Survey* (CPS) – March Supplement &amp; Food Supplement</td>
<td>Tract</td>
</tr>
<tr>
<td>American Housing Survey (AHS)</td>
<td>Tract</td>
</tr>
<tr>
<td>National Longitudinal Survey (NLS) – Young/Mature Women</td>
<td>Lat/Lon Block Group</td>
</tr>
<tr>
<td>National Longitudinal Survey (NLS) – Young/Old Men</td>
<td>County</td>
</tr>
</tbody>
</table>

**Demographic Data**
- household or individual level
- more detailed geography
- less top-coding/censoring
- additional variables
- occupation codes
- possibly bigger samples
- metadata

**Combined Employer/Employee Data**
- Longitudinal Employer-Household Dynamics (LEHD)
### Example 2 of Data Availability – Census Bureau Data

#### Economic Data: Establishment, firm, transaction level

<table>
<thead>
<tr>
<th>Establishment Surveys</th>
<th>Economic Census</th>
<th>Business Register Data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data Set</strong></td>
<td><strong>Data Set</strong></td>
<td><strong>Data Set</strong></td>
</tr>
<tr>
<td>Annual Survey of Manufacturers (ASM)</td>
<td>CompuStat-SSEL  Bridge (CSB)</td>
<td>Census of Auxiliaries (AUX)</td>
</tr>
<tr>
<td>Current Industrial Reports (CIR)</td>
<td>Form 5500 Bridge File</td>
<td>Census of Construction Industries (CCN)</td>
</tr>
<tr>
<td>Manufacturing Energy Consumption Survey (MECS)</td>
<td>Integrated Longitudinal Business Database (ILBD)</td>
<td>Census of Finance, Insurance, Real Estate (CFI)</td>
</tr>
<tr>
<td>Medical Expenditure Panel Survey – Insurance Component (MEPS-IC)</td>
<td>Longitudinal Business Database (LBD)</td>
<td>Census of Manufacturers (CMF)</td>
</tr>
<tr>
<td>National Employer Survey (NES)</td>
<td>Ownership Change Database (OCD)</td>
<td>Census of Mining (CMI)</td>
</tr>
<tr>
<td>Quarterly Survey of Plant Capacity Utilization (QPC)</td>
<td>Standard Statistical Establishment List / Business Register (SSEL)</td>
<td>Census of Retail Trade (CRT)</td>
</tr>
<tr>
<td>Survey of Manufacturing Technology (SMT)</td>
<td></td>
<td>Census of Services (CSR)</td>
</tr>
<tr>
<td>Survey of Plant Capacity Utilization (PCU)</td>
<td></td>
<td>Census of Transportation, Communications, Utilities (CUT)</td>
</tr>
<tr>
<td>Survey of Pollution Abatement Costs and Expenditures (PACE)</td>
<td></td>
<td>Census of Wholesale Trade (CWH)</td>
</tr>
</tbody>
</table>

#### Transactions Data

<table>
<thead>
<tr>
<th><strong>Data Set</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Commodity Flow Survey (CFS)</td>
</tr>
<tr>
<td>Foreign Trade Data - Export (EXP)</td>
</tr>
<tr>
<td>Foreign Trade Data - Import (IMP)</td>
</tr>
<tr>
<td>Longitudinal Foreign Trade Transactions Data (LFTTD)</td>
</tr>
</tbody>
</table>
NCHS Data Sets

- National Health and Nutrition Examination Survey (NHANES)
- National Health Care Surveys
  - National Ambulatory Medical Care Survey (NAMCS) and National Hospital
    Ambulatory Medical Care Survey (NHAMCS)
  - National Hospital Discharge Survey (NHDS)
  - National Nursing Home Survey (NNHS) and National Nursing Assistant Survey
    (NNAS)
  - National Home and Hospice Care Survey (NHHCS) and National Home Health
    Aide Survey (NHHAS)
- National Health Interview Survey (NHIS)
- National Survey of Family Growth (NSFG)
- State and Local Area Integrated Telephone Survey (SLAITS)
- NCHS Data Linkage Activities
  - Linked Mortality Data Products
  - Linked Medicare Enrollment and Claims Files Data
  - Linked Social Security Benefit History Data
  - Linked Air Quality Data
- National Vital Statistics System (NVSS)

Example 3 of Data Availability – Health Data

**National Center for Health Statistics (NCHS)**

- Restricted Variables
- Geographic Variables
- Linked Data Products
- Temporal Variables
  - *e.g.* dates of birth, death, exams
- Detailed Race/Ethnicity Variables
- Sensitive Variables (*e.g.* youth sexual behavior, mental health)
- Long Term Health Care Survey Merging Variables
New Datasets Frequently Become Available

See Census FSRDC web links for compilations
- FSRDC Available Data
- Participating Agencies Data Sets
- Census CES Restricted Data Notes

National Center for Health Statistics (NCHS)
- NCHS Restricted Data Sets & Information

Agency for Healthcare Research and Quality (AHRQ)
- AHRQ Restricted Data Sets & Information

Consult with the RDC Administrator to inquire about available datasets

New Datasets are added regularly!

https://liberalarts.tamu.edu/txrdc/datasets/
What you need

**Skill Sets**

- Need to be comfortable, conscientious with using software to manage, organize, analyze data
- Must develop excellent habits of documenting workflow
- Must be comfortable diving into codebooks and being a detective about data issues
Important Practical Issues to Consider

Project Approval by Agency and Researcher Security Approval
• Agency review and approval
• Demonstrated “benefits” to federal agencies
• Special Sworn Status (3-year residency requirement for non-citizens)

The RDC Computing Environment – Software and Hardware
• Computing is conducted **in-person** in the secure lab
• Census Bureau workstation hardware/software limitations

Releasing Research Results Involves a Review Process
• Access, analysis, and disclosure of results follow strict protocols
• Sworn confidentiality for life
Other Important Practical Issues to Consider

Have a very clear vision of the project

Have support from your committee chair and members
  • Extended timeline – Assume at least 1 extra year
  • Inability to make major changes to the project once started
  • Inability to preview “preliminary” results
  • Need to obtain Special Sworn Status (SSS)

Absolutely no-nonsense!
  • Cannot game this system
  • “Full-on” professional commitment
Stage of Graduate Studies and Planning

RDC Research

**Pro** - the potential for performing cutting edge research

**Con** – the time investments associated with first projects and the extended timeline to completed research projects
Benefits to you!

Job Market Considerations
- Papers at meetings will be more complicated
- Getting results for job talks will be more complicated
- Results can be spectacular

Future Career Considerations
- Potentially transportable to other locations
- Your future research program can be more ambitious
- Your path to tenure can be enhanced
Census Mentorship Opportunity!

- Center for Economic Studies seeks candidates for its Dissertation Mentorship Program.
- **WHO**: Doctoral candidates using Census Bureau microdata at a Research Data Center.
- **WHAT**: One or more researchers are assigned as mentors and advise the participant on the use of Census Bureau microdata.
- **WHEN**: Now!
- U.S. citizenship is required.
- Students from TAMU and Univ. of Houston have participated in the program. Contact us for more info!
Is RDC Research Right for You?

If you answer YES to these questions:

• Are you organized and have the right skill set?
• Is your project path clear?
• Is the benefit of RDC data substantial to your project?
• Can you work independently? Are you self-reliant?
• Are you patient and can you respect protocols (even when they are frustrating)?
Can it Work? “You Bet’cha!”

EXAMPLE OF GRADUATE STUDENT WORK

WHY THE RDC? VIDEO
We are here to help you!

TXRDC Administrator – Karin Johnson
- Direct liaison to Census Bureau
- Assists with proposal development & project planning
- Oversees disclosure review
- Manages the secure computing lab

Director – Mary Campbell
- Oversees contracts with U.S. Census Bureau
- Promotes awareness, shares information, and encourages and assists with proposal development
- Nurtures/grows the TXRDC consortium
- Approves lab access for projects
- Coordinates with the RDC Admin to oversee the facility
I am Interested – What is Next?

- Attend TXRDC Workshops
- Contact TXRDC
- Do not delay
- Review available information
- Obtain Census Bureau approval
- Start as Soon As Possible

Start as Soon As Possible
Welcome to the Texas Research Data Center, which operates in close collaboration with the U.S. Census Bureau to enable research that expands basic scientific knowledge and provides benefits to the federal statistical system. Learn more...
THANK YOU!

Karin Johnson: karin.johnson@census.gov
Mary Campbell: m-campbell@tamu.edu