Federal Statistical Research Data Centers (FSRDC)

Samuel R. Bondurant
Dallas-Fort Worth Federal Statistical Research Data Center
Center for Enterprise Dissemination
US Bureau of the Census

Any opinions and conclusions expressed herein are those of the author and do not necessarily represent the views of the U.S. Census Bureau.
Federal Statistical Research Data Center (FSRDC)

- Federally secure research lab where authorized researchers are granted secure access to restricted-use microdata
  - Thin client access to Census linux servers
  - Census Bureau and other statistical agencies
  - Fees to maintain facility typically paid by a local consortium of research institutions
    - Additional fees for non-Census agency projects
Research Data Centers
# Census Research Data Centers

<table>
<thead>
<tr>
<th>RDC</th>
<th>State</th>
<th>Open</th>
<th>RDC</th>
<th>State</th>
<th>Open</th>
</tr>
</thead>
<tbody>
<tr>
<td>Census Bureau</td>
<td>MD</td>
<td>1982</td>
<td>Penn State</td>
<td>PA</td>
<td>2014</td>
</tr>
<tr>
<td>NBER</td>
<td>MA</td>
<td>1994</td>
<td>Yale</td>
<td>CT</td>
<td>2015</td>
</tr>
<tr>
<td>UCLA/USC/UC Irvine</td>
<td>CA</td>
<td>1998</td>
<td>U. Wisconsin</td>
<td>WI</td>
<td>2015</td>
</tr>
<tr>
<td>Chicago Fed</td>
<td>IL</td>
<td>2002</td>
<td>U. Maryland</td>
<td>MD</td>
<td>2016</td>
</tr>
<tr>
<td>Cornell</td>
<td>NY</td>
<td>2004</td>
<td>Georgetown</td>
<td>DC</td>
<td>2017</td>
</tr>
<tr>
<td>CUNY Baruch</td>
<td>NY</td>
<td>2006</td>
<td>U. Kentucky</td>
<td>KY</td>
<td>2017</td>
</tr>
<tr>
<td>Atlanta Fed</td>
<td>GA</td>
<td>2011</td>
<td>U. Colorado Boulder</td>
<td>CO</td>
<td>2017</td>
</tr>
<tr>
<td>U. Washington</td>
<td>WA</td>
<td>2012</td>
<td>U. Texas Austin</td>
<td>TX</td>
<td>2017</td>
</tr>
<tr>
<td>Texas A&amp;M</td>
<td>TX</td>
<td>2012</td>
<td>Dallas Fed</td>
<td>TX</td>
<td>2018</td>
</tr>
</tbody>
</table>
Computing Infrastructure

RDC

RDC

RDC

Census Servers (computing environment)

Disk (data)
Research Data Center (RDC)

- Main Phases of an RDC Project
  - Proposal drafting
  - Submit proposal for review
  - Security clearance (SSS)
  - RDC analysis
  - Statistical output for Disclosure Avoidance Review
  - Provide working paper and benefit to Census
Why Is Census Required to Restrict Microdata Access?

- Titles 13 (Census) /26 (IRS) U.S.C. and CIPSEA protect confidentiality
  - respondent cannot be identified
  - only Census employees and temporary staff can access microdata
  - use limited to statistical purpose
  - access must potentially provide legitimate benefits to Census Bureau programs
RDC Resources

- www.dfwrfdc.org
  - Data
  - Proposal Process
  - Presentations

- https://ideas.repec.org/s/cen/wpaper.html
  - Working paper series
  - Google “census ces working paper series”
Data Availability

- Census Bureau Data
  - Economic Data
    - establishment or firm level
  - Demographic Data
    - household or individual level
  - Combined Employer/Employee Data
    - Longitudinal Employer-Household Dynamics (LEHD)

- Health Data
  - Agency for Healthcare Research and Quality (AHRQ)
  - National Center for Health Statistics (NCHS)

- Bureau of Labor Statistics
  - NLS, NCS, SOII, CFOI

- Bureau of Economic Analysis
RDC Economic Data Advantages

• No publicly-available microdata
  ▪ Internal data at establishment and firm level
  ▪ Universal scope
  ▪ Detailed industry and geography

• Linking Data
  ▪ Consistent identifiers
  ▪ Business register
    ▪ External data
## Economic Census

<table>
<thead>
<tr>
<th>Data Set</th>
<th>RDC Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Census of Auxiliaries (AUX)</td>
<td>1977-2012</td>
</tr>
<tr>
<td>Census of Construction Industries (CCN)</td>
<td>1972-2012</td>
</tr>
<tr>
<td>Census of Finance, Insurance, Real Estate (CFI)</td>
<td>1992-2012</td>
</tr>
<tr>
<td>Census of Manufacturers (CMF)</td>
<td>1963-2012</td>
</tr>
<tr>
<td>Census of Mining (CMI)</td>
<td>1987-2012</td>
</tr>
<tr>
<td>Census of Retail Trade (CRT)</td>
<td>1977-2012</td>
</tr>
<tr>
<td>Census of Services (CSR)</td>
<td>1977-2012</td>
</tr>
<tr>
<td>Census of Transportation, Communications, Utilities (CUT)</td>
<td>1987-2012</td>
</tr>
<tr>
<td>Census of Wholesale Trade (CWH)</td>
<td>1977-2012</td>
</tr>
</tbody>
</table>
Research Example

  - RDC microdata
    - Census of Manufactures (CMF)
    - Annual Survey of Manufactures (ASM)
    - Business Register (SSEL/BR)
  - Winning and losing counties have similar trends in incumbents’ TFP prior to a large new plant opening.
  - Five years after the opening, incumbent plants’ TFP is 12 percent higher in winning counties.
### Establishment Surveys

<table>
<thead>
<tr>
<th>Data Set</th>
<th>RDC Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annual Survey of Manufacturers (ASM)</td>
<td>1973-2016</td>
</tr>
<tr>
<td>Current Industrial Reports (CIR)</td>
<td>1993-2011</td>
</tr>
<tr>
<td>Manufacturing Energy Consumption Survey (MECS)</td>
<td>1985-2010</td>
</tr>
<tr>
<td>Medical Expenditure Panel Survey – Insurance Component (MEPS-IC)</td>
<td>1996-2016</td>
</tr>
<tr>
<td>National Employer Survey (NES)</td>
<td>1994-1998</td>
</tr>
<tr>
<td>Plant Capacity Utilization (PCU)</td>
<td>1974-2017</td>
</tr>
<tr>
<td>Survey of Manufacturing Technology (SMT)</td>
<td>1988-1993</td>
</tr>
<tr>
<td>Survey of Pollution Abatement Costs and Expenditures (PACE)</td>
<td>1979-2005</td>
</tr>
</tbody>
</table>
Research Example

  - RDC microdata
    - Medical Expenditures Panel Survey-Insurance Component (MEPS-IC)
    - Longitudinal Business Database (LBD)
    - Business Register (SSEL/BR)
  - Decline in employer sponsored health insurance
    - Rising premium costs
    - Declining union coverage and collective bargaining laws
## Firm Surveys

<table>
<thead>
<tr>
<th>Data Set</th>
<th>RDC Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annual Business Survey</td>
<td>2017</td>
</tr>
<tr>
<td>Annual Capital Expenditures Survey (ACES)</td>
<td>1993-2015</td>
</tr>
<tr>
<td>Annual Retail Trade Survey (ARTS)</td>
<td>1999-2014</td>
</tr>
<tr>
<td>Business Research &amp; Development and Innovation Survey (BRDIS)</td>
<td>2008-2016</td>
</tr>
<tr>
<td>Exporter Database (EDB)</td>
<td>1992-2010</td>
</tr>
<tr>
<td>Service Annual Survey (SAS)</td>
<td>1991-2014</td>
</tr>
<tr>
<td>Survey of Business Owners (SBO)</td>
<td>1982-2012</td>
</tr>
<tr>
<td>Survey of Industrial Research and Development (SIRD)</td>
<td>1972-2007</td>
</tr>
</tbody>
</table>
# Business Register Data

<table>
<thead>
<tr>
<th>Data Set</th>
<th>RDC Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compustat-SSEL Bridge (CSB)</td>
<td>1976-2016</td>
</tr>
<tr>
<td>Form 5500 Bridge File</td>
<td>1992-2012</td>
</tr>
<tr>
<td>Integrated Longitudinal Business Database (ILBD)</td>
<td>1977-2014</td>
</tr>
<tr>
<td>Longitudinal Business Database (LBD)</td>
<td>1976-2016</td>
</tr>
<tr>
<td>Ownership Change Database (OCD)</td>
<td>1963-2002</td>
</tr>
</tbody>
</table>
Research Example

- Yichen Su (current project), “Local Multiplier Effect”
  - RDC microdata
    - Longitudinal Business Database (LBD)
    - Business Register (SSEL/BR)
      - Compustat-SSEL Bridge (CSB)
  - Previous research estimates effect at MSA
  - This project estimates the effect at the Census Tract level
Research Example

  - RDC microdata
    - Longitudinal Business Database (LBD)
    - Census of Manufacturers (CMF)
    - Annual Survey of Manufacturers (ASM)
  - Interstate banking deregulations $\rightarrow$ increased access to bank financing $\rightarrow$ increases in firms’ TFP productivity
  - Regression discontinuity around SBA funding eligibility
## Transactions Data

<table>
<thead>
<tr>
<th>Data Set</th>
<th>RDC Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commodity Flow Survey (CFS)</td>
<td>1993-2012</td>
</tr>
<tr>
<td>Foreign Trade Data - Export (EXP)</td>
<td>1992-2018</td>
</tr>
<tr>
<td>Foreign Trade Data - Import (IMP)</td>
<td>1992-2018</td>
</tr>
<tr>
<td>Longitudinal Foreign Trade Transactions Data (LFTTD)</td>
<td>1992-2016</td>
</tr>
</tbody>
</table>
RDC Demographic Data Advantages

• Individual and household data

• Restricted use RDC data
  ▪ more detailed geography
  ▪ less top-coding/censoring
  ▪ additional variables
  ▪ larger sample
## More Detailed Geography

<table>
<thead>
<tr>
<th>Data Set</th>
<th>RDC Years</th>
<th>Geography</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decennial Census</td>
<td>1950-2010</td>
<td>Block</td>
</tr>
<tr>
<td>American Community Survey (ACS)</td>
<td>1996-2016</td>
<td>Block</td>
</tr>
<tr>
<td>Survey of Income and Program Participation (SIPP)</td>
<td>1984-2011</td>
<td>Tract/County</td>
</tr>
<tr>
<td>American Housing Survey (AHS)</td>
<td>1984-2015</td>
<td>Tract</td>
</tr>
<tr>
<td>National Longitudinal Survey (NLS) – Young/Mature Men/Women</td>
<td>1966-1999</td>
<td>Tract</td>
</tr>
<tr>
<td>National Longitudinal Mortality Study (NLMS)</td>
<td>1973-2011</td>
<td>County</td>
</tr>
</tbody>
</table>
Research Example

- Kyle Walker (current project) “Demographic Inversion in the U.S. Metropolitan Areas”
  - RDC microdata
    - American Community Survey
  - Migration of Millennial populations to urban areas
  - PUMAs have a coarse geography
    - Internal ACS goes down to the block-level
Research Example

  - RDC microdata
    - Decennial Censuses
    - BLS – National Longitudinal Survey
  - Use census birth date to link to Vietnam draft lottery dates to use as an instrument for military service.
  - Long run consequence of Vietnam service were found to be schooling gains and no earnings losses.
Top-coded 2010 Public Use ACS

<table>
<thead>
<tr>
<th>State</th>
<th>Electric Payment</th>
<th>Age</th>
<th>Wage</th>
<th>Mortgage</th>
</tr>
</thead>
<tbody>
<tr>
<td>AL</td>
<td>570</td>
<td>93</td>
<td>267,000</td>
<td>3700</td>
</tr>
<tr>
<td>FL</td>
<td>570</td>
<td>94</td>
<td>330,000</td>
<td>5300</td>
</tr>
<tr>
<td>GA</td>
<td>570</td>
<td>92</td>
<td>314,000</td>
<td>4800</td>
</tr>
<tr>
<td>SC</td>
<td>560</td>
<td>93</td>
<td>281,000</td>
<td>4300</td>
</tr>
<tr>
<td>TN</td>
<td>540</td>
<td>93</td>
<td>312,000</td>
<td>4000</td>
</tr>
<tr>
<td>TX</td>
<td>580</td>
<td>92</td>
<td>322,000</td>
<td>4400</td>
</tr>
</tbody>
</table>

- RDC ACS age and mortgage expenditures are not top-coded, wages are top-coded at one million dollars.
- RDC data include tract, school and congressional district, birthday, migration place code, place of work tract code.
### 2010 ACS PUMS Top-coded Variables

<table>
<thead>
<tr>
<th>Housing Variables</th>
<th>Person Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electricity</td>
<td>Property tax</td>
</tr>
<tr>
<td>Gas</td>
<td>Insurance</td>
</tr>
<tr>
<td>Water</td>
<td>Mortgage</td>
</tr>
<tr>
<td>Heating fuel</td>
<td>2(^{nd}) mortgage</td>
</tr>
<tr>
<td>Condo fee</td>
<td></td>
</tr>
<tr>
<td>Rent</td>
<td></td>
</tr>
<tr>
<td>Mobile home costs</td>
<td></td>
</tr>
<tr>
<td>*Wages, salary income</td>
<td>Self-employment income</td>
</tr>
<tr>
<td>Interest, dividends</td>
<td>Retirement income</td>
</tr>
<tr>
<td>All other income</td>
<td>Age</td>
</tr>
<tr>
<td>Travel time to work</td>
<td></td>
</tr>
</tbody>
</table>
*Majority are matched using just the SEIN and not the SEIN+SEINUNIT pairing
Longitudinal Employer-Household Dynamics (LEHD)

- Economic Data
  - Economic Census
  - Manufacturing
  - Wholesale Trade
  - Retail Trade
  - Etc.
  - Annual Survey Manu.
  - Kauffman Firm Survey
  - Etc.

- LBD

- SSEL

- ECF/BRB

- LEHD

- Demographic Data
  - ACS
  - CPS
  - SIPP
  - Decennial
    - Short form
    - Long form
    - Etc.
Research Example


- RDC microdata
  - LEHD
  - Economic Censuses
  - Decennial Censuses

- Contribution of establishments in the upward trend in earnings dispersion

- Business Register
- Longitudinal Business Database
- Establishment Surveys
Data from Partnering Agencies Accessible in the RDC

- Agency for Healthcare Research and Quality
- National Center for Health Statistics
- Bureau of Labor Statistics

- Proposal process is with individual agencies and not Census
  - Security background check still with Census
Agency for Healthcare Research and Quality (AHRQ)

- Fully specified ICD-9 medical condition codes
- Fully specified industry and occupation codes
- State and county FIPS codes
- Census tract and block-group codes
- Non-public use data elements
  - Asset information
  - Imputed NDC codes
- Federal and state marginal tax rates
AHRQ Data Sets

- Medical Expenditure Panel Survey – Household Component (MEPS-HC)
- Nursing Home Component (1996)
- Medical Provider Component (except directly identifiable data)
- Two-Year, Two-Panel Files
- Area Resource File (county-level data that can be linked to MEPS-HC)
National Center for Health Statistics (NCHS)

Types of Restricted Variables

- Geographic Variables
- Linked Data Products
- Genetic Variables (NHANES phenotype data)
- Temporal Variables
  - e.g. dates of birth, death, exams
- Detailed Race/Ethnicity Variables
- Sensitive Variables
  - e.g. youth sexual behavior and mental health
- Long Term Health Care Survey Merging Variables
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)
Census Proposal Process

• Pre-Preliminary proposal
  ▪ Send Wenhua and I an email

• Preliminary proposal submission portal
  ▪ https://www.icpsr.umich.edu/web/pages/appfed/index.html

• Proposal development
  ▪ Involve RDC staff

• Census Review
  ▪ Other Agency Review

• Special Sworn Status application
Proposal

• Proposal
  • Overview
  • Requested Datasets and years
    • Why you need internal data
  • Methodology
    • Linking process
    • Estimating equations
    • Measures constructed
  • Expected Output
  • Duration and Funding
• Benefits to the Census Bureau
  • Minimum of two (one being the working paper)
• Abstract
13 Criteria for Benefits

1. Evaluating practices, continued relevance
2. Demog., econ. processes that affect Census programs
3. Increasing utility of data for analyzing public programs, public policy, and/or demographic, economic, or social conditions
4. Conducting data collection, processing or dissemination
5. Understanding/improving the quality of data
6. Improved methodology to collect, measure, or tabulate
7. Enhancing the data collected (e.g. imputation, links)
8. Limitations/improving the Business Register
9. New data collection needs
10. Constructing, verifying, improving sampling frames
11. Estimates(characteristics of population)
12. Methodology for estimating non-response
13. Developing statistical weights for a survey
Distribution of Benefit Criterion

- Data Quality
- Data Enhancement
- Population Estimates

Benefit Criterion

1 2 3 4 5 6 7 8 9 10 11 12 13

Percent

0 5 10 15 20 25 30 35 40 45
Proposal Review

• Census Review
  ▪ Feasibility
  ▪ Requirement of benefits to Census
  ▪ Scientific merit
  ▪ Statistical purpose
  ▪ Need for non-public data
  ▪ Risk of disclosure
  ▪ Availability of resources

• Other Agency Review

• Special Sworn Status application
Timeline

- Census review, (1-3 months)
  - Assign to a reviewer
  - Discussion committee
  - (if needed) IRS review (3-7 additional months)
- After approval, SSS Packet
- Background check (2-3 months)
  - Additional security process for non-citizens
Timeline

• Pre-Preliminary proposal
  ▪ Send Wenhua and I an email
• Preliminary proposal
• Proposal development
  ▪ Involve RDC staff
• Census Review (1-3 months)
  ▪ Other Agency Review (additional 3-6 months, if needed)
• Special Sworn Status application (2-3 months)
  ▪ Additional security process for non-citizens

• Project work (3-5 years)
## Publications by RDC Researchers and CES Staff: 2018 and Forthcoming

<table>
<thead>
<tr>
<th>Econ Journals (by rank)</th>
<th>RDC Researchers</th>
<th>CES staff</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAA (1-5)</td>
<td>12</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>AA (6-20)</td>
<td>14</td>
<td>7</td>
<td>21</td>
</tr>
<tr>
<td>A (21-102)</td>
<td>22</td>
<td>15</td>
<td>37</td>
</tr>
<tr>
<td>B (103-258)</td>
<td>12</td>
<td>11</td>
<td>23</td>
</tr>
<tr>
<td>C (259-562)</td>
<td>5</td>
<td>6</td>
<td>11</td>
</tr>
<tr>
<td>D (562-1,202)</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Journals outside of Econ</td>
<td>14</td>
<td>11</td>
<td>25</td>
</tr>
<tr>
<td>Book Chapters</td>
<td>3</td>
<td>11</td>
<td>14</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>82</strong></td>
<td><strong>63</strong></td>
<td><strong>145</strong></td>
</tr>
</tbody>
</table>
Call to Action

samuel.r.bondurant@census.gov
wenhua.di@dal.frb.gov

- Send us a quick proposal pitch
  - Send me simple data questions
- Request a meeting in person
  - Include an itinerary if possible
- www.census.gov/fsrdc/about