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.
Goals of this Presentation

- Overview of RDC program
  - Description of working environment
- Phases of an RDC Project
- Available data products and advantages
- Realistic timeline
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>Open</th>
<th>RDC</th>
<th>Open</th>
<th>RDC</th>
<th>Open</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chicago Fed</td>
<td>2002</td>
<td>Penn State</td>
<td>2014</td>
<td>U. Texas Austin</td>
<td>2017</td>
</tr>
<tr>
<td>Cornell</td>
<td>2004</td>
<td>U. Missouri</td>
<td>2015</td>
<td>Federal Reserve</td>
<td>2019</td>
</tr>
<tr>
<td>CUNY Baruch</td>
<td>2006</td>
<td>U. Wisconsin</td>
<td>2015</td>
<td>U. of Utah</td>
<td>2020</td>
</tr>
</tbody>
</table>
DFW FSRDC Consortium
What is an RDC?

*Not the actual lab
Census Servers
- Project space
- Statistical packages
- Data
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.dfwrdc.org
  - Data
  - Proposal Process
  - Presentations

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

- Note: on the dfwrfdc.org website:
  - Census proposal guidelines (pdf)
  - Links to NCHS, AHRQ, BLS, BEA proposal processes
*Census* Proposal

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

Percent

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

• Census Review
  ▪ Feasibility
  ▪ Benefits to Census strength
  ▪ Scientific merit
  ▪ Need for non-public data
  ▪ Risk of disclosure

• Additional Agency Review, if needed
  • Economic data (IRS), Numident (SSA), LEHD (states), CPS (BLS)
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
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
Special Sworn Status (SSS)

- All persons working in the RDC require SSS
  - Census, NCHS, AHRQ, BLS, BEA, etc.
  - Census absorbs SSS cost, other agencies might not
- Must reside in the U.S. for 3 of last 5 years to qualify
- Couple months to process (longer for non-citizens)
  - Application process can start after Census project approval
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
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
RDC Analysis

- All work involving restricted-use data will occur in the RDC
- SAS, Stata, R, python, GeoDA, etc. available on server
- Data Staff at HQ uploads user provided data/code to project space
- Large data files, takes time to organize
- Projects lasts from 3-5 years

*Not the actual lab*
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
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
Disclosure Avoidance Review

- All output from an RDC project must go through disclosure avoidance review
- Involved process with RDC Admin
- Disclosure requirements different across data
- HQ Disclosure Office will email you approved output
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
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
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

**Working paper**
Why Submit an RDC Proposal?

- Data availability
- Data advantages
- High level research
Why Submit an RDC Proposal?

- Data availability
- Data advantages
- High level research
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)
- UMERICs Data
- 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
## Economic Data

### Economic Censuses Years
- Auxiliaries (AUX) 1977-2012
- Construction Industries (CCN) 1972-2012
- Fin., Insurance, Real Estate (CFI) 1992-2012
- Manufacturers (CMF) 1963-2012
- Mining (CMI) 1987-2012
- Retail Trade (CRT) 1977-2012
- Services (CSR) 1977-2012
- Trans., Comms., Utilities (CUT) 1987-2012
- Wholesale Trade (CWH) 1977-2012

### Establishment Surveys
- Annual Survey of Manufacturers (ASM) 1973-2016
- Current Industrial Reports (CIR) 1993-2011
- Manufacturing Energy Consumption Survey (MECS) 1985-2010
- Plant Capacity Utilization (PCU) 1974-2017
- Survey of Manufacturing Technology (SMT) 1988-1993
- Survey of Pollution Abatement Costs and Expenditures (PACE) 1979-2005

### Business Register Data Years
- Compustat-SSEL Bridge (CSB) 1976-2016
- Form 5500 Bridge File 1992-2012
- Integrated Longitudinal Business Database (ILBD) 1977-2014
- Longitudinal Business Database (LBD) 1976-2016
- Ownership Change Database (OCD) 1963-2002
- Standard Statistical Establishment List/Business Register (SSEL/BR) 1974-2016

### Transactions Data Years
- Commodity Flow Survey (CFS) 1993-2012
- Foreign Trade Data - Export (EXP) 1992-2018
- Foreign Trade Data - Import (IMP) 1992-2018
- Longitudinal Foreign Trade Transactions Data (LFTTD) 1992-2016

### Firm Surveys Years
- Annual Business Survey 2017
- Annual Capital Expenditures Survey (ACES) 1993-2015
- Annual Retail Trade Survey (ARTS) 1999-2014
- Business R&D and Innovation Survey (BRDIS) 2008-2016
- Exporter Database (EDB) 1992-2010
- Service Annual Survey (SAS) 1991-2014
- Survey of Business Owners (SBO) 1982-2012
- Survey of Industrial R&D (SIRD) 1972-2007
## Demographic Data

<table>
<thead>
<tr>
<th>Data Set</th>
<th>RDC Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>American Community Survey (ACS)</td>
<td>1996-2016</td>
</tr>
<tr>
<td>American Housing Survey (AHS)</td>
<td>1984-2017</td>
</tr>
<tr>
<td>CPS Supplements</td>
<td>(1997-2017)</td>
</tr>
<tr>
<td>Decennial Census</td>
<td>1950-2010</td>
</tr>
<tr>
<td>National Crime Victimization Survey (NCVS)</td>
<td>2006-2017</td>
</tr>
<tr>
<td>National Longitudinal Mortality Study (NLMS)</td>
<td>1973-2011</td>
</tr>
<tr>
<td>Survey of Income and Program Participation (SIPP)</td>
<td>1984-2014</td>
</tr>
</tbody>
</table>
*Majority are matched using just the SEIN and not the SEIN+SEINUNIT pairing
Data from Partnering Agencies Accessible in the RDC

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

- Proposal process is with individual agencies and not Census
  - Security background check still with Census
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)
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)
Why Submit an RDC Proposal?

- Data availability
- Data advantages
- High level research
Why Submit an RDC Proposal?

- Data availability
- Data advantages
- High level research
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
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.
RDC Demographic Data Advantages

- Individual and household data
- Restricted use RDC data
  - more detailed geography
  - less top-coding/censoring
  - additional variables
  - larger sample
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.
## More Detailed Geography

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Geography</th>
</tr>
</thead>
<tbody>
<tr>
<td>American Housing Survey (AHS)</td>
<td>Tract</td>
</tr>
<tr>
<td>American Community Survey (ACS)</td>
<td>Block</td>
</tr>
<tr>
<td>Current Population Survey (CPS) – March Supplement</td>
<td>Tract</td>
</tr>
<tr>
<td>Decennial Census</td>
<td>Block</td>
</tr>
<tr>
<td>National Crime Victimization Survey (NCVS)</td>
<td>Block</td>
</tr>
<tr>
<td>National Longitudinal Survey (NLS) – Young/Mature Men/Women</td>
<td>Tract</td>
</tr>
<tr>
<td>National Longitudinal Mortality Study (NLMS)</td>
<td>County</td>
</tr>
<tr>
<td>Survey of Income and Program Participation (SIPP)</td>
<td>Tract/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
Longitudinal Employer-Household Dynamics (LEHD)

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

- LEHD
  - LBD
  - SSEL
  - ECF/BRB

- 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
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
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
Why Submit an RDC Proposal?

- Data availability
- Data advantages
- High level research
Why Submit an RDC Proposal?

- Data availability
- Data advantages
- High level research
Publications by RDC Researchers and CES Staff: 2018 and Forthcoming

Econ Journals (by rank) | RDC Researchers | CES staff | Total
--- | --- | --- | ---
AAA (1-5) | 12 | 2 | 14
AA (6-20) | 14 | 7 | 21
A (21-102) | 22 | 15 | 37
B (103-258) | 12 | 11 | 23
C (259-562) | 5 | 6 | 11
D (562-1,202) | 0 | 0 | 0

Journals outside of Econ | 14 | 11 | 25

Book Chapters | 3 | 11 | 14

**TOTAL** | 82 | 63 | 145
Timeline

- Census review, (1-3 months)
  - Assign to reviewers
  - Reviewing committee
  - (if needed) IRS review (3-7 additional months)
- After approval, SSS Packet
- Security clearance (2-3 months)
  - Additional time for non-citizens
- Project work from 3-5 years
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