Data available from the MU Research Data Center

A Branch of the Kansas City RDC and a unit of the Federal Statistical Research Data Centers (FSRDC)

MU Population, Education and Health Center (PEHC)

Colleen Heflin, Co-Director (presenter)

Peter Mueser, Co-Director

October 19, 2015
Presentation Outline

- Background: PEHC and the MU RDC
- Overview of data available through RDC
  - 3 other presentations this fall
- Focus on demographic data
  - Review of particular datasets and variables available
- Brief review of procedures for obtain access to MU RDC
Background

- January 2013: Population, Education and Health (PEHC) Seminar Series starts
- Summer 2013: Submission of NSF proposal to establish **Federal Statistical Research Data Center (FSRDC)** in Kansas City with branch site at MU
  - Facilitates research access to Census confidential data
  - MU is in consortium with KC Fed, Kauffman Foundation, KU and UMKC
Background

- June 2013: Census Bureau Site Visit
  - Over 40 MU researchers meet with Census Bureau staff

- August 2014: NSF Accepts Proposal
  - MU commits $1 million over 5 years to support MU RDC
  - RDC in KC and branch at MU to be in place by November 2015
PEHC Support of RDC

- Doctoral Fellowships
- Grant Writer to help obtain outside funding
- Small Grants Program
  - Support work preparing proposal for RDC data access and analysis
  - Most grants $10,000-$15,000
  - Deadline Nov. 15, 2015
  - For more information:
    - muprovostpehc@missouri.edu
    - https://truman.missouri.edu/research-publications/population-education-and-health-center
What is a Census RDC?

- A secure environment where researchers can undertake analysis of non-public data from federal agencies.
- MU will be one of 24 RDC sites nationwide by the end of 2015.
- Use of RDC allows for research that would be difficult otherwise.
Advantages of Access to Non-Public Census Data

- Data not available elsewhere
  - Establishment level business data
  - Linked worker-firm data (LEHD)
  - Data on where people live and work
- Detailed geo-spatial variables
  - Census block group
  - Longitude and latitude
- Virtually no top or bottom coding of variables
  - Earnings, income
Advantages of Non-Public Data

- Can link many Census and/or national health (AHRQ/NCHS) datasets with each other
  - Current Population Survey (CPS) with LEHD
  - Economic Census with Survey of Business Owners
- Can link to other non-Census data
  - National Center for Health Statistics data with American Community Survey
  - Other contextual data
Other Presentations

1. RDC Demographic Data Presentation
   ○ Monday, October 19, 2015  1:00-2:00 PM
   ○ Student Center room 1209

2. RDC Health Data Presentation
   ○ Friday, October 30, 2015  12:00-1:00 PM
   ○ CE705 in the Clinical Support and Education Building

2a. Using and Accessing Medical Expenditure and Panel Survey (MEPS) data
   ○ Thursday, October 22, 2015
   ○ 12:30–1:30 pm
   ○ Memorial Union N222/223
Demographic Data Sets

- Census of Population and Housing (Decennial Census)
- American Community Survey (ACS)
- Current Population Survey (CPS)
  - March & December Supplements
- Survey of Income and Program Participation (SIPP)
  - Longitudinal and Panels
  - Survey of Program Dynamics
- American Housing Survey (AHS)
- National Crime Victimization Survey (NCVS)*
- National Longitudinal Mortality Study (NLMS)*
- National Longitudinal Survey (NLS)*
## 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>1970-2000</td>
<td>Block</td>
</tr>
<tr>
<td>American Community Survey (ACS)</td>
<td>1996-2009</td>
<td>Block</td>
</tr>
<tr>
<td>Survey of Income and Program Participation (SIPP)</td>
<td>1984-2008</td>
<td>Tract</td>
</tr>
<tr>
<td>American Housing Survey (AHS)</td>
<td>1984-2009</td>
<td>Tract</td>
</tr>
<tr>
<td>National Longitudinal Survey (NLS) – Young/Mature Women</td>
<td>1966-2003</td>
<td>Lat/Lon Block Group</td>
</tr>
<tr>
<td>National Longitudinal Survey (NLS) – Young/Old Men</td>
<td>1966-1990</td>
<td>County</td>
</tr>
<tr>
<td>National Longitudinal Mortality Study (NLMS)</td>
<td>1973-2002</td>
<td>County</td>
</tr>
</tbody>
</table>
Decennial Census

- Provide annual data on demographic, social, economic and housing characteristics.
- Meet federal legislative and program needs
- Other stakeholders in state and local government and private sectors
- Improve the federal statistical system
- Person: age, sex, race, Hispanic origin, hhid relationship
- Household: occupancy, tenure, household type
Decennial Census and ACS Missing Data

Imputation Process

- **Substitution**
  - Household exists but no “data-defined” person records
  - Characteristics of a nearby household are assigned using “hot-deck” procedure

- **Allocation**
  - Assignment used other data on respondent
  - Allocation—hot deck procedure
  - Flags variables
Census: Public vs. Restricted Data

- 1%, 5% Sample
- PUMA
- Top Coding
- Collapsed Categories
- Data Swapping
- Perturbation
- No Administrative
- Example of Collapsed Categories:
  - Real Estate Taxes Paid $0 to $9,150 in 67 categories.
- 16 % Sample
- Block/Tract/Place
- Full Distribution
- Full Distribution
- Data Swapping (usually)
- Actual Ages
- Details of Collection
- Example of Full distribution:
  - Real Estate Taxes Paid $0 to $22,500+ in single dollars.
Census: Research Examples (Long Form)


American Community Survey

- Provide annual data on demographic, social, economic and housing characteristics.
- Continuous measurement rather than snapshot
- Replaced the Decennial Census “Long-Form”
- Content designed to meet the needs of federal government agencies
- 21 housing and 48 population questions
- Household Respondent provides responses for all other residents of the household
- Householder or Reference Person is “Person 1”
ACS: Public vs. Restricted Data

RDC data include tract, school and congressional district, birthday, migration place code, place of work tract code.

- 1% Sample
- PUMA
- Top Coding
- Collapsed Categories
- Perturbation
- No Administrative
- Swapped

- 2.5% Sample
- Block/Tract/Place
- Wages top-coded at $1million
- Full Distribution
- Actual Ages
- Details of Collection
- Swapped (usually)
<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\textsuperscript{nd} mortgage</td>
</tr>
<tr>
<td>Condo fee</td>
<td>All other income</td>
</tr>
<tr>
<td>Rent</td>
<td>Age</td>
</tr>
<tr>
<td>Mobile home costs</td>
<td>Travel time to work</td>
</tr>
<tr>
<td>Wages, salary income</td>
<td>Self-employment income</td>
</tr>
<tr>
<td>Interest, dividends</td>
<td>Retirement income</td>
</tr>
</tbody>
</table>
ACS Data Collection Information

- Identify the form of the survey responses, i.e. mail, telephone interview, face-to-face interview.
- Identify whether a respondent/unit called for Telephone Questionnaire Assistance.
- Identify whether a unit received a “failed edit follow up.”
ACS: Research Examples


Current Population Survey

- Supplemental studies on other topics such as health insurance coverage, food security, computer usage, and fertility/childbearing
  - Annual Social and Economic Supplement, March
  - Food Security Supplement, December
- Quick turnaround (timeliness)
- Long history
- Expert interviewers
- Topical Modules
CPS: Public vs. Restricted Data

- All Months and Topics
- All Records
- State/MSA
- Top Coding
- Collapsed Categories
- No Administrative

- ASEC (March) & Food Security (December) Only
- All Records
- Tract
- Full Distribution
- Full Distribution
- Details of Collection
CPS: Research Examples


SIPP aims to provide accurate and comprehensive information about income and program participation and their determinants.

The survey offers detailed information on cash and non-cash income on a sub-annual basis.

Includes information on taxes, assets, liabilities, and participation in government transfer programs.

SIPP has topical modules that include detailed information on special topics, including retirement and pension plan coverage, child care, and medical expenses.
Public vs. Restricted Data

- RDC provides access to detailed geographical information
- Can be linked to SSA earnings records

Research Examples

Survey of Program Dynamics

- Created after 1996 welfare reform law (PRWORA)
- Household from the 1992-3 SIPP panel were followed from 1997-2002.
- Core data were collected on employment, income, program participation, health insurance and utilization, child well-being, marital relationships, and parents’ depressions.
- Like SIPP, the SPD has topical modules that vary by year.
American Housing Survey (AHS)

- Collects data on housing, such as characteristics of homes, apartments, mobile homes, vacant units, etc. and households (income, mobility, housing size, quality and costs)
- National data is collected in odd years, and data is collected from 47 metropolitan areas each every 4 years (12/year).

Restricted Data Example:
RDC Proposal Process
Overview

1. Researchers interested in using restricted-use Census datasets will need to go through a proposal process and obtain Special Sworn Status (SSS)

2. Jake Cronin is the RDC Administrator that works full-time on campus to help researchers through the proposal process.
Proposal Process

- The goal is to draft proposals for feasible projects that provide academic merit as well as benefits to Census, with minimal disclosure risk.
  - Feasible—Is it possible to answer this research question with the data that is available?
  - Academic merit—staff at Census with expertise in topic will review. Generally pretty lenient.
- You work with the RDC Administrator to write the proposal. The RDC Administrator then forwards the final proposal to Census (or other agencies).
- Health data has a different proposal process.
Benefits to Census

- By law, researchers cannot access restricted-use data unless they document how this access benefits the mission of the Census Bureau.
- Proposals will have to explain how their proposed research benefits the Census Bureau.
- There are thirteen different benefit criterion to choose from, and with the help of the RDC Administrator, most research meets multiple criterion.
Special Sworn Status

- After the proposal has been approved by Census Bureau, researchers will need to obtain Special Sworn Status (SSS) in order to physically access the RDC.
  - SSS entails a background check and an oath to project the confidentiality of the data forever
  - Requires completing paperwork, getting fingerprints, online trainings, and identification verifications.
Working in an RDC

- Only researchers with approved projects and Special Sworn Status are allowed in the RDC.
- Researchers will work on “thin client” computers that tunnel into secure Census servers.
  - No data is stored on local machines
- All analysis must be done in the RDC.
- Output cannot be removed from the RDC prior to a formal disclosure avoidance review
Disclosure Avoidance Review

- There is a formal process to remove aggregated output out of the RDC. That process *typically* takes about three weeks.
- This is the only approved process for moving results from inside the confidential RDC to the outside public domain.
- The general idea is to ensure that no identifiable information on an establishment, firm, person, or household is disclosed.
  - In the proposal phase, it is important to think ahead about the output of the research and if it poses any disclosure risk.
More Information

- For information on restricted-use data available through the RDC:
  - [http://www.census.gov/ces/dataproducts/restricted_data.html](http://www.census.gov/ces/dataproducts/restricted_data.html)

- MU RDC Administrator: jacob.cronin@census.gov
  - 573-884-9122 (when RDC opens)

- [https://truman.missouri.edu/research-publications/population-education-and-health-center/](https://truman.missouri.edu/research-publications/population-education-and-health-center/)
Further Information

Population, Education, and Health Center

muprovostpehc@missouri.edu
Colleen Heflin, co-director, heflincm@missouri.edu
Peter Mueser, co-director, mueserp@missouri.edu

Resources

Center for Economic Studies (CES)
https://www.census.gov/ces/rdcresearch/

AHRQ
http://www.meps.ahrq.gov/mepsweb/data_stats/onsite_datacenter.jsp

NCHS
http://www.cdc.gov/nchs/r&d/rdc.htm