RDC Information Session: Demographic Data

Demographic Data available from the MU Research Data Center

Jake Cronin, MU RDC Administrator (Presenter)
Eileen Avery, Director
May 21, 2020
Overview

• Background and purpose of Federal Statistical Research Data Centers
• Data available through the RDC
  • Examples of RDC research
• Process for accessing RDC data
• Contact information and sources for additional information
• Questions
What is a Research Data Center?

• A secure environment where researchers can undertake analysis of non-public data from federal agencies

• MU joined approximately 30 RDC sites nationwide.

• The RDC allows for research that would be difficult (if not impossible) otherwise
Types of Data Available

- **Demographic Data** (today’s focus)
  - Decennial Census, CPS, ACS, SIPP, etc.

- **Health Data** (yesterday’s presentation)
  - National Center for Health Statistics
  - Agency for Healthcare Research & Quality

- **Economic Data** (Next Thursday at 3:00pm)
  - Business register data
  - Transactions data (trade flows)
  - LEHD – Longitudinal Employer/Household Data
Demographic Data

• Data on people and households
• Relative to public use, restricted use data has:
  • Lower levels of geography
  • No top-coding
  • Full samples
  • Greater detail in variables
    • Race
    • Place of birth
    • Write-in responses
Other Advantages of Access to Non-Public Data

• Can link Census datasets with each other (if PIK)
  • Decennial Census with the American Community Survey
  • Economic Census with Survey of Business Owners

• Can link to other non-Census data (external data)
  • American Community Survey (ACS) with Missouri Department of Education data
    • Data must have an identifier to merge on (e.g. social security number)

• Data not available elsewhere
  • Establishment level business data
  • Linked worker-firm data (LEHD)
<table>
<thead>
<tr>
<th>Data Set</th>
<th>Years</th>
<th>Personal Identity Key (PIK)</th>
<th>Geography Identified</th>
</tr>
</thead>
<tbody>
<tr>
<td>American Community Survey</td>
<td>1996-2014</td>
<td>Yes</td>
<td>X</td>
</tr>
<tr>
<td>Data Set</td>
<td>Years</td>
<td>Personal Identity Key (PIK)</td>
<td>Geography Identified</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>----------</td>
<td>----------------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>National Survey of College Graduates</td>
<td>2010</td>
<td>No</td>
<td></td>
</tr>
</tbody>
</table>
Decennial Census


• Short form collects basic information from every person and housing unit

• Long-form was a 1-6 sample that asked additional information on marital status, place of birth, labor force participation, occupation, income, etc.
  • ACS replaced long-form in 2010

• The restricted use data set gives a much larger sample size, narrower geographic area, greater detail in variables, and ability to merge with other datasets
American Community Survey

• Collects annual information from U.S. households similar to what was collected on the Decennial Census long-form (1996-2017)
  • E.g. marital status, place of birth, labor force participation, occupation, income, commute time to work, home value, veteran status, and other important data.

• Collected annually from about 3.5 million respondents

• Similar to the decennial census, the restricted use dataset gives a much larger sample size, narrower geographic area (block level), greater detail in variables, and ability to merge with other datasets
Current Population Survey

• Collect monthly from about 60,000 households
• Primary source of labor force statistics for the population of the United States.
  • CPS is the source of numerous high-profile economic statistics, including the national unemployment rate and the national poverty rate
  • Purpose is to produce national estimates. Estimates at lower levels may not be as reliable.
• The RDC includes supplements:
  • Annual Social and Economic Supplement in March
  • Food security in December
  • Voting and Registration Supplement
  • Fertility supplement
  • School enrollment
American Housing Survey

• Collected from approximately 50,000 households every other year
• Has data on housing conditions and costs, neighborhood characteristics, home improvements, home values, as well as household composition, demographic characteristics, and income.
  • HUD uses this data to monitor supply and demand of housing and improve housing programs appropriate for different groups
National Survey of College Graduates

• Repeated cross-sectional biennial survey that provides data on the nation's college graduates, with a focus on those in the science and engineering workforce.

• Unique source for examining various characteristics of college-educated individuals, including occupation, work activities, salary, the relationship of degree field and occupation, and demographic information.
Survey of Income and Program Participation

• Nationally representative panel survey where individuals are interviewed multiple times over approximately a four year period

• Collects extensive information on sources of income and participation in public programs.

• Topical Modules:
  • Adult well-being, assets, childcare, education, family and household, nativity and citizenship, demographics, health care utilization, school meals, food security.
National Longitudinal Mortality Survey

• Created to study the effects of demographic and socio-economic characteristics on differentials in U.S. mortality rates.
• Links CPS data and a subset of the 1980 Census with death certificate information to identify mortality status and cause of death.
  • Death records through 2011
• Has 3.8 million records with over 550,000 identified mortality cases.
National Crime Victimization Survey

• Nationally representative survey collected annually from about 90,000 households (160,000 individuals)

• Contains data on demographic characteristics and reports of crime victimization during the prior six months:
  • Demographic characteristics of individual
  • Some information on offender (age, race and Hispanic origin, sex, and victim/offender relationship)
  • Characteristics of the crime (time, place, injuries, economic consequences)
  • Victims experience with justice system

• Various supplements also available at various times: identity theft, school crime, police public contact
• BLS data just recently became available in the RDC’s
  • National Longitudinal Survey of Youth (1979 and 1997)
    • Nationally representative sample of young men and women who were 14-22 years old when they were first surveyed in 1979 or 1997, respectively. Interviewed annually (through 1994 for the 1979 cohort).
    • Restricted data has geography down to the census tract level at each interview, contextual data about places they live (education, crime, etc.), and school surveys (1997)
Moving to Opportunity (MTO)

• MTO is a unique random assignment research effort sponsored by the U.S. Department of Housing and Urban Development (HUD).

• This demonstration was designed to help very low-income families with children living in public housing or Section 8 project-based housing in extremely poor neighborhoods relocate to “opportunity neighborhoods” for greater self-sufficiency and improved individual and family well-being.

• The MTO demonstration ran in five large cities -- Baltimore, Boston, Chicago, Los Angeles, and New York -- between September 1994 and August 1998.

• A total of 4,608 families enrolled in the MTO demonstration and were randomly assigned. The demonstration combined Section 8 rental assistance with intensive housing search and counseling services to ease families' relocation to low-poverty communities and help them become self-sufficient.
UMETRICS

- Dataset constructed to understand the productivity of funding for scientific research
- Has administrative data from about 30 universities
  - Awards and award transactions at the researcher/transaction level
- Linked with Census data to get wages and establishment level data
- Linked with data from ProQuest to get publications and dissertations
Combining Workers and Firms

• LEHD: Longitudinal Employer-Household Dynamics Database:
  • Employment History Files (EHF): Quarterly UI Wage Records
  • Employer Characteristic Files (ECF): Employer data
  • Individual Characteristics Files (ICF): Person level demographic data

• Firm and worker information combined to create job level quarterly earnings history data, data on where workers live and work, data on firm characteristics, such as industry
Recent publications using restricted-use LEHD


SSA-Linked Data

• Social Security Administration (SSA)
  • SSA provides earnings records dating back to 1951. Earnings are taken from the IRS W2 records (yearly wages).
    • Also has information on births within families (from SSA registration)

• Can be linked with Census datasets (e.g. CPS and SIPP)
• Genadek (2018) linked CPS-ASEC to the SSA earnings records to estimate the motherhood wage penalty
SNAP/TANF/WIC

• Administrative state-level records documenting program participation
• Can be linked with Census data (e.g. ACS, CPS, SIPP, etc.)
• Data availability varies by state
• Must provide benefits to the Food and Nutrition Service
  • For example, your research must provide estimates to improve understanding of poverty and food insecurity
Potential Linkages

Decennial Census and ACS datasets
- 1940 Census
- 2000 Census
- 2010 Census
- 2001-2015 ACS

Household survey datasets
- CPS
- SIPP
- AHS
- NCVS

Administrative Records
- External Data Sources
MU RDC Demographic Data Projects

• A researcher was recently approved to access and link together the restricted Decennial and ACS data. She is studying the effects of an educational program which had a staggered statewide rollout. The restricted data allowed her to identify where the respondent lived at two different points to know 1) if the respondent received the treatment and 2) the long-term effects of the program on labor market outcomes.

• Another researcher was recently approved to access the restricted use SIPP data. She needed the restricted use data to identify and study a particular foreign-born sub-population.
  • Immigrant groups often aggregated to very-high levels in public-use.

• Another project was recently approved to study in and out migration in the rural Ozark region in the years around the Great Recession. Needed the restricted use data to identify current and prior county of residence.
COVID-19 Research

• Federal agencies are working to expedite the proposal process for researchers interested in studying topics related to Covid-19

• Census currently conducting the Household Pulse Survey
  • Collecting data on the socioeconomic impact of Covid-19 on American households. Survey was just conducted recently.

• Most data is dated but still possible to link current Covid data to past aggregated data to determine patterns and trends
  • Lisa Sattenspiel and Carolyn Orbann project
RDC Proposal Process
Overview

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

• Jake Cronin is the RDC Administrator who works full-time at MU 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 RDC data?
  • Academic merit—staff at Census with expertise in topic will review. Generally very lenient.

• You work with the RDC Administrator to write the proposal. The RDC Administrator then forwards the final proposal to Census or other agency for approval.

• Still accepting proposals during shutdown!
Benefits to Census

• By law, researchers must document how their research benefits the mission of the Census Bureau
  • MU’s RDC administrator will work with you to identify the appropriate benefit
    • There are thirteen different benefit criterion. Most research meets multiple criterion.
Special Sworn Status

• After the proposal has been approved, researchers need to obtain Special Sworn Status (SSS) to physically access the RDC
  • SSS entails background check and oath to project the confidentiality of the data
  • Requires completing paperwork, getting fingerprints, online trainings, identification verifications
  • Can take about 3 months to issue SSS
  • Process is tedious, but denial is uncommon
Working in an RDC

• Only researchers with approved projects and Special Sworn Status are allowed in the RDC.

• NCHS/AHRQ researchers can only work in the RDC when Administrator is present

• 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.

• 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 establishments, people, or households 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

• Demographic data:
  • https://www.census.gov/programs-surveys/ces/data/restricted-use-data/demographic-data.html

• Bureau of Labor Statistics Data
  • http://www.bls.gov/rda/restricted-data.htm

• MU RDC website (google ‘PEHC Missouri’):
  • https://coas.missouri.edu/center/population-education-and-health-center
  • RDC Annual Report
MU RDC Contact Information

Jacob Cronin
MU RDC Administrator
Jacob.Cronin@census.gov
573-884-9122

Eileen Avery
Director
AveryE@Missouri.edu