RDC Information Session: Demographic Data

Demographic Data available from the MU Research Data Center

Jake Cronin, MU RDC Administrator (Presenter)
Colleen Heflin, Co-Director
Peter Mueser, Co-Director
October 14, 2016
Overview

• Background and purpose of Federal Statistical Research Data Centers
• Information on PEHC Small Grant and Doctoral Fellowship
• 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 25 RDC sites nationwide.

• The RDC allows for research that would be difficult (if not impossible) otherwise
PEHC Small Grants

• University of Missouri research faculty and staff are eligible
• Support work preparing proposal for RDC data access and analysis
• Most grants are around $10,000-$15,000
• Applications due November 15, 2016 and May 1, 2017

More information and application can be found here:
• https://truman.missouri.edu/population-education-and-health-center/small-grant-program/
PEHC Doctoral Fellowship

• Fellowship for up to two doctoral students to conduct research using Census Research Data Center (RDC) data products.

• Each Fellowship includes a $25,000 stipend for the academic year (August-May) and full tuition remission provided by the Office of Graduate Studies.

• Eligible projects include a faculty-initiated project or a student dissertation and must require the use of restricted data only available through the Census RDC.

• Applications accepted in Spring Semester.
Types of Data Available

• **Demographic Data (today’s focus)**
  • Decennial Census, CPS, ACS, SIPP, etc.
  • Am. Housing Survey, Natl. Crime Victimization Survey

• **Health Data**
  • National Center for Health Statistics
  • Agency for Healthcare Research & Quality

• **Economic Data**
  • 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)
  • Current Population Survey (CPS) with LEHD
  • 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></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>American Community Survey</td>
<td>1996-2014</td>
<td>Yes</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td></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>Survey of Income and Program Participation (SIPP)</td>
<td>1984-2011</td>
<td>Yes</td>
<td>X X X X X</td>
</tr>
<tr>
<td>National Survey of College Graduates</td>
<td>2010</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>American Housing Survey (AHS)</td>
<td>1984-2013</td>
<td>Yes (some years)</td>
<td>X X X X X</td>
</tr>
<tr>
<td>National Longitudinal Survey</td>
<td>1966-1999</td>
<td>No</td>
<td>X X X X X</td>
</tr>
<tr>
<td>Young/Mature Men/Women</td>
<td>1966-1999</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>National Longitudinal Mortality Study</td>
<td>1973-2011</td>
<td>No</td>
<td>X X</td>
</tr>
<tr>
<td>National Crime Victimization Survey</td>
<td>2006-2012</td>
<td>Yes</td>
<td>X X X X X X X</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
  • 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

- Longitudinal biennial survey conducted since the 1970s to provide data on the nation’s college graduates.
- 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.
• 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
Research Using Restricted Use Data

Demographic Data


Bureau of Labor Statistics

• 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)
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.
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

• For information on restricted-use data available through the RDC:
• Demographic data:
  • https://www.census.gov/ces/dataproducts/demographicdata.html
• Bureau of Labor Statistics Data
  • http://www.bls.gov/rda/restricted-data.htm
• MU RDC website (google ‘PEHC Missouri’):
  • https://truman.missouri.edu/population-education-and-health-center/mu-research-data-center/
MU RDC Contact Information

**Jacob Cronin**
MU RDC Administrator
[jacob.cronin@census.gov](mailto:jacob.cronin@census.gov)
573-884-9122

Colleen Heflin
MU RDC Co-Director
[heflincm@missouri.edu](mailto:heflincm@missouri.edu)

Peter Mueser
MU RDC Co-Director
[muesermp@missouri.edu](mailto:muesermp@missouri.edu)