The Rocky Mountain Research Data Center (RMRDC)

Jani Little, Executive Director
Phil Pendergast, Administrator (Census)
Kas McLean, GRA (Econ)

ADVANCING THE FRONTIERS OF SOCIAL SCIENCE

ECONOMICS - POPULATION - HEALTH
A National Network of Scholars
What is a Federal Statistical Research Data Center (FSRDC)?

--A secure computing lab where restricted data, collected by federal agencies, can be accessed FOR STATISTICAL PURPOSES ONLY
--Confidentiality is protected by Federal Laws and Regs
--The Census Bureau directs all FSRDCs and the FSRDC Program
--FSRDCs are managed by an on-site Census employee—the administrator—who guides researchers on proposal development, enforces security guidelines, and serves as liaison with the research community.
Restricted Data to Investigate Society’s Most Pressing Challenges

- Local Labor Markets
- Health Disparities
- Immigration impacts on People and Business
- Natural Hazards impacts on People and Business
- International Trade Policies
- Rural vs Urban Disparities
- Business Sector Growth and Decline
- Crime Victims and Consequences
A Regional Network of Scholars
Today’s Topics

Jani Little—Emerging Restricted-Data Offerings—Census, BLS, NCHS
Kas McLean—Longitudinal Business Database (LBD) and Longitudinal Employer-Household Dynamics (LEHD)
Phil Pendergast—International Trade Data and National Crime Victimization Survey (NCVS)
Public-Use Microdata: created so that original records cannot be reproduced or linked to other databases

1. No obvious identifiers or PII (personally identifiable information)
2. Includes only a portion of the surveyed sample
3. Includes only a selection of variables. Sensitive variables removed.
4. Limit geographic detail
5. Limit the number and detailed breakdown of categories within variables on the file, e.g., race and ethnicity, exact dates of events
6. Truncate extreme codes for certain variables (top or bottom coding), e.g., income
7. Recode into intervals or round continuous variables
8. No unique or sensitive populations: disabled children, teenagers with STDs, LGBT, mixed race persons
9. No highly unique records
10. Perturbed values or noise added

Restricted Access Microdata

1. No obvious identifiers or PII (personally identifiable information) replaced by PIK (Protected Identification Key) to allow linkages
2. Includes all microdata records
3. Includes unlimited number of variables including sensitive variables and LINKED data records
4. Geographic detail as precise as original data
5. Unlimited categories within variables on the file, e.g., race and ethnicity, exact dates of events
6. No truncated extreme codes for certain variables (top or bottom coding), e.g., income
7. No recoding into intervals or rounded continuous variables
8. Unique and sensitive populations included
9. Highly unique records
10. Original records unperturbed and without added noise
Census Longitudinal Infrastructure Project (CLIP)

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
Current projects making use of linked Census Data

• Linking 2010, 2000 full count census to assess race response changes and how is racial identity influenced by neighborhood characteristics

• Interracial couples and children’s race – how are they reported, how does that change over time
Master Address File (MAF)

- MAF Extract
  - Complete list of addresses with lat/lon and geo information
  - Yearly snapshot of MAF can be used for research

- MAF-Auxiliary Reference File (ARF)
  - Links addresses to individuals
Current projects making use of MAF Linked Data

• Linking MAF with ACS data to get more precise estimates and characteristics of small areas and how they change over time.
Bureau of Labor Statistics (BLS) Offerings:

--Survey of Occupational Injuries and Illnesses (SOII)

--Census of Fatal Occupational Injuries (CFOI)

--American Time Use Survey (ATUS)
Current projects making use of ATUS

• How child care costs affect time spent away from home with spouse?

• How does smoking affect a single mother’s time with children?
Current projects making use of MAF Linked Data

• Linking MAF with ACS data to get more precise estimates and characteristics of small areas and how they change over time.
Federal Statistical Data
for Health Research and Policy
Spring 2020
CU Denver | Main and Anschutz Medical Campus
Instructors: Jani Little, Phil Pendergast, Kas McLean
Fridays, 9am-12noon
Federal Statistical Data for Health Research and Policy
Spring 2020

Students develop the knowledge and skills required to effectively use a variety of federal statistical data sets for health research and policy analysis.

Students get “hands on” exercises using SAS to analyze public versions of the federal data sets

The semester project is a restricted data proposal.
Anyone at University of Denver is invited to attend the lectures (9-10am) by zoom
https://ucdenver.zoom.us/j/981548769
Federal Statistical Data for Health Research and Policy
Topics Covered and Schedule
• National Vital Statistics System (NVSS)
• National Health and Nutrition Examination Surveys (NHANES)
• National Health Interview Survey (NHIS) linked with National Death Index (NDI)
• NHIS linked with Medical Expenditure Panel Survey (MEPS)
• American Community Survey (ACS) and Current Population Survey (CPS)
• National Crime Victimization Survey (NCVS)
• Survey of Occupational Injuries and Illnesses (SOII), Census of Fatal Occupational Injuries (CFOI), American Time Use Survey (ATUS)

• Jan 31
• Feb 7
• Feb 14
• Feb 28
• Mar 6
• April 3
• April 10
Longitudinal Business Database (1976-2017)

• Complete U.S. business establishment histories (with paid employees)
  • Establishment birth and death dates
  • All employer establishments in the Economic Censuses
  • Detailed linkage documentation over time and very accurate histories

• Basic establishment demographic data linkable to other business datasets by EIN, SEIN
  • Establishment size
  • Industry
  • Location
Longitudinal Employer-Household Dynamics (LEHD) [https://www.census.gov/ces/dataproducts/lehddata.html](https://www.census.gov/ces/dataproducts/lehddata.html)

Combines administrative data from states’ Unemployment Insurance systems with Census Bureau business data and the individual demographic data using the Social Security Administration number database of encrypted identifiers.

**Workers**--Employer history and wage history, Individual characteristics (sex, age, race, ethnicity, education, place of birth), Point in time residence

**Employers**--Industry, # employees, firmage, total payroll, location

Linkages between workers and employers
Longitudinal Employer-Household Dynamics (LEHD) [https://www.census.gov/ces/dataproducts/lehddata.html](https://www.census.gov/ces/dataproducts/lehddata.html)

For external projects-only available from cooperating states

- Option A: AZ AR DE DC IN IA KS ME MD NV OK TN
- Option B: AL CA CO CT FL GA HI ID IL KY LA MA MI MN MO MT NE NH NJ NM NY NC ND OR PA RI SC TX UT VT VA WA WV WI WY
Trade Data – Bureau of Economic Analysis (BEA)

- Somewhat different rules – you become a BEA “Special Sworn Employee”
- Additional requirements:
  - Must be a US citizen
  - Must not be employed by an agency with investigative, regulatory, or policy functions
  - Provide a benefit to BEA
- Go to BEA home page (www.bea.gov):
  - ‘Research’
    - “Special Sworn Employee Program”
- Expensive! Fees per year:

<table>
<thead>
<tr>
<th></th>
<th>FY20</th>
<th>FY21</th>
<th>FY22</th>
<th>FY23</th>
</tr>
</thead>
<tbody>
<tr>
<td>First year of project</td>
<td>$7,174</td>
<td>$7,265</td>
<td>$7,358</td>
<td>$7,452</td>
</tr>
<tr>
<td>Subsequent years</td>
<td>$1,830</td>
<td>$1,867</td>
<td>$1,904</td>
<td>$1,942</td>
</tr>
</tbody>
</table>
Trade Data – BEA Company-level Microdata

- Foreign Direct Investment (FDI)
  - Operations of multinational enterprises (MNEs)
    - U.S.-based MNEs’ operations worldwide (BE-10 and BE-11)
    - Foreign-based MNEs operations in the U.S. (BE-12/13/15)
  - Quarterly FDI transactions for
    - Every directly owned foreign affiliate with total assets, sales, or revenues of more than $60 million USD in the year.
    - Indirectly owned foreign affiliates meeting this threshold that have intercompany debt balance with US reporter >$1 million

- Trade In Services
Trade Data – BEA Company-level Microdata

• Foreign Direct Investment (FDI)
  • BE-10 US MNEs, BE-12 Foreign MNEs
    • Covers entire universe every 5 years (“Benchmark Years”)
    • Financials and operations, costs and expenses, employee headcount and compensation, infrastructure (property, plant, equip. spending), ex/imports
  • BE-11 US MNEs, BE-15 Foreign MNEs
    • Larger companies sampled/surveyed annually
    • Reports not required for smaller affiliates (carried forward from last BE-10)
• BE-13 Foreign MNEs
  • For new foreign direct investments in the US
  • Acquisition, value, ownership information
Trade Data – BEA Company-level Microdata

• Trade In Services data
  • BE-120/125
    • Selected Services and Intellectual property sales between MNEs
      • Patents, reported performances or events, computer software, books, etc
  • BE-45/140
    • Insurance transactions between US companies and foreign persons
  • BE-180/185
    • Financial services – brokerage, underwriting, credit services, etc.
  • BE-9/29/30/37
    • Transport services – Airline transactions, revenues, etc.
National Crime Victimization Survey (NCVS) 2006-2018

• Approximately 135,000 households each year
• AND 240,000 individuals in the HH aged 12 or older
• Dwelling units remain in sample for 3.5 years, surveyed every 6 months
• Yields annual and change estimates of criminal victimization for persons age 12 or older
• Includes persons living in group quarters (e.g. dormitories) but excludes institutional settings (correctional or hospital facilities) and homeless
National Crime Victimization Survey (NCVS) 2006-2018

• Person Data
  • Demographics (age, sex, race, ethnicity, marital status, education, income, employment)
  • Nonfatal personal crimes (i.e., rape or sexual assault, robbery, aggravated and simple assault, and personal larceny)
  • Yields potentially longitudinal personal data on marital status, education, income, employment

• Victimization incident:
  • offender (e.g., age, race and Hispanic origin, sex, and victim-offender relationship),
  • characteristics of the crime (e.g., time and place of occurrence, use of weapons, nature of injury, and economic consequences)
  • crime was reported to police, reasons the crime was or was not reported, and victim experiences with the criminal justice system
  • sought medical treatment
Student experiences with and perceptions of . . .
crime and safety at school
preventive measures employed by schools
participation in after school activities
school rules and enforcement
presence of weapons, drugs, alcohol, and gangs in school
bullying
hate-related incidents
fear of victimization at school
The Rocky Mountain Research Data Center

Jani Little, Executive Director

Jani.little@Colorado.edu

Phil Pendergast, Administrator (Census)

Kas McLean, GRA (Econ)

ADVANCING THE FRONTIERS OF SOCIAL SCIENCE

ECONOMICS - POPULATION - HEALTH