CONDUCTING RESEARCH AT THE NEW YORK FEDERAL STATISTICAL RESEARCH DATA CENTER, BARUCH

Diane Gibson, Ph.D.
OVERVIEW

• Purpose of the Federal Statistical Research Data Centers (FSRDCs)

• What types of questions can be addressed using FSRDC data?

• Demographic, economic, and health data available through the FSRDC

• Applying to use the FSRDC

• Contact information and sources for additional information
The 22 FSRDCs provide researchers with approved projects with access to Census Bureau data that are not publicly available.

The Census Bureau has agreements with NCHS and AHRQ to host restricted-access data at the FSRDCs. BLS data is coming soon.

NYRDC currently operates three FSRDCs, one at Baruch College, one at Cornell University, and one at Yale University.
The NYRDC is a consortium. Researchers with an affiliation with any of the consortium members do not incur fees for using the CED.

- Baruch College
- City University of New York
- Columbia University
- Cornell University
- Federal Reserve Bank of New York
- National Bureau of Economic Research
- New York University
- Princeton University
- The Russell Sage Foundation
- Syracuse University
- University at Albany - SUNY
- Yale University
What types of questions can be addressed using FSRDC data?
What is the effect of social interactions among neighbors on labor market outcomes? Is a person more likely to work with someone who lives in the same block than with someone who lives in a nearby block?

- **Survey:**
  - 1990 Decennial Census (long form)

- **Restricted data:**
  - Census block of each individual’s residential and employment locations

- **Need for restricted data:**
  - Block information is used to define “neighbors.”

Did the economic gains experienced by low-income neighborhoods in the 1990s follow patterns of classic gentrification?

- Survey:
  - American Housing Survey

- Restricted data:
  - Census tract of each housing unit
  - Need for restricted data:
    - Tract identifiers used to merge in Decennial Census data on neighborhoods.

Measuring income inequality using the ratio of the 90th to the 10th percentile of income deal with time-inconsistency issues in income top-coding in the Current Population Survey.

- Survey:
  - Current Population Survey
- Restricted data:
  - Wage, salary and household income data with a censoring point greater than or equal to the public use data censoring point.

What types of firms are responsible for job creation? How much does firm size and age matter?

- Data:
  - Longitudinal Business Database
  - Restricted data:
    - All - no public-use microdata on establishments

What are the product and country determinants of intrafirm trade?

- Data:
  - Foreign trade – export
  - Foreign trade – import
  - Longitudinal Business Database

- Restricted data:
  - All - no public-use microdata

Does health insurance coverage and health care and use for adults change after the implementation of health reforms?

- Survey:
  - National Health Interview Survey (NHIS)
- Restricted data:
  - State of residence
- Need for restricted data:
  - State of residence is used to identify respondents in New York and Massachusetts.

Is the county-level availability of ophthalmologists and optometrists associated with measures of realized access to eye care for individuals with diabetes, diabetic retinopathy, or age-related macular degeneration?

- Survey:
  - National Health and Nutrition Examination Survey (NHANES)
- Restricted data:
  - State, county and tract of residence
- Need for restricted data:
  - Geographic identifiers used to merge in eye care provider data from the Area Health Resources Files
What are the relationships between the frequency and quantity of alcohol consumption and mortality?

- **Survey:**
  - NHIS linked to the National Death Index
  - Restricted data:
    - Mortality information

DATA AVAILABLE

Census Bureau
- Household/individual data
- Business data
- Mixed – individual/household linked to business data

National Center for Health Statistics (NCHS)
- Household/individual data
- Provider data (e.g., nursing homes; hospitals)

Agency for Healthcare Research and Quality (AHRQ)
- Medical Expenditure Panel Survey (MEPS)
There are public versions of these data, but RDC data...
- provide more detailed geographic identification (census tract and in some cases census block)
- may include variables not available in public versions
- are not top-coded/censored

### Household and/or individual data

<table>
<thead>
<tr>
<th>Data</th>
<th>Frequency</th>
<th>Sampling Unit</th>
<th>Availability</th>
</tr>
</thead>
<tbody>
<tr>
<td>American Community Survey</td>
<td>Annually</td>
<td>Housing Unit</td>
<td>1996–2014</td>
</tr>
<tr>
<td>National Longitudinal Surveys</td>
<td>Annually</td>
<td>Individual</td>
<td>1966–1999</td>
</tr>
<tr>
<td>National Survey of College Graduates (NSCG)</td>
<td>Annually</td>
<td>Individual</td>
<td>2010 (started in 1970s)</td>
</tr>
</tbody>
</table>
ECONOMIC DATA

- Business data at the plant/establishment or firm level
  - No publicly-available micro-data

- Low levels of geography
  - Address (and name of business)

- Linking Data
  - Across censuses/surveys over time
  - Across entities (establishment to firm level)
  - External data (e.g., Compustat)

See the complete list with descriptions at http://www.census.gov/ces/
<table>
<thead>
<tr>
<th>Data</th>
<th>Frequency</th>
<th>Sampling Unit</th>
<th>Availability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Longitudinal Business Database (LBD)</td>
<td>Annually</td>
<td>Establishment</td>
<td>1976–2013</td>
</tr>
<tr>
<td>Annual Capital Expenditures Survey (ACES)</td>
<td>Annually</td>
<td>Firm</td>
<td>1993–2012</td>
</tr>
</tbody>
</table>
LONGITUDINAL EMPLOYER–HOUSEHOLD DYNAMICS (LEHD)

Combines administrative data from states’ Unemployment Insurance systems with Census data

- Linkages between workers and employers
- **Workers** (Source: State unemployment wage records/UI)
  - Employer history and quarterly wages
  - Individual characteristics (sex, age, race, DOB, county of birth)
  - Point in time residence and place of birth
- **Employers** (Source: Quarterly census of employment and wages)
  - Industry, employment, total payroll, location
- **Links to other data**
  - Business Register Bridge (link to Economic Censuses)
  - SIPP; CPS March supplement, ACS
  - Geocoded Address List (link to External Data)
Examples of restricted-access variables:
- Detailed geography
- Continuous/non top-coded variables

Some NCHS data can be linked to:
- Mortality files
- Social Security files
- Medicare/Medicaid files
- Air quality files (indirect match by detailed geography)
### Popular NCHS Data

<table>
<thead>
<tr>
<th>Data</th>
<th>Frequency</th>
<th>Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>National Health and Nutrition Examination Survey (NHANES)</td>
<td>Annual</td>
<td>5,000 persons per year, all ages</td>
</tr>
<tr>
<td>National Health Interview Survey (NHIS)</td>
<td>Annual</td>
<td>42,000 households per year</td>
</tr>
<tr>
<td>National Survey of Family Growth (NSFG)</td>
<td>Annual</td>
<td>5,000 men and women, 15-44 years of age</td>
</tr>
<tr>
<td>National Vital Statistics System</td>
<td>Annual</td>
<td>All Births (about 4 million records annually)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>All Deaths (about 2.4 million records annually)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reported fetal deaths of 20+weeks gestation (about 26,000 annually)</td>
</tr>
<tr>
<td>Longitudinal Studies of Aging (LSOAs)</td>
<td>Biennial</td>
<td>7,527 persons 70 years of age and over at the time of their 1984 SOA interview.</td>
</tr>
</tbody>
</table>
Medical Expenditure Panel Survey (MEPS) -

The MEPS Household Component collects nationally representative data on how households consume and pay for healthcare.

- Two-year panels, beginning in 1996
- Nationally representative subsample of households that participated in the National Health Interview Survey (conducted by the National Center for Health Statistics) in the prior year
- Demographic characteristics, health conditions, health status, use of medical care services, charges and payments, access to care, satisfaction with care, health insurance coverage, income, and
FSRDCS, THE PROPOSAL PROCESS, AND THE RESEARCH ENVIRONMENT

• Unfortunately, you cannot show up next week at the NYRDC to start looking at the data.
• If you are interested in using the FSRDC data, you will need to go through the proposal process.
• If your proposal is approved, you will need to go through a background investigation (SSS).
• Usually takes about one year from the time you start proposal development until you can access the data.
• NCHS and AHRQ proposals take less time on average.
• 8 cubicles at NYRDC - Baruch, each with its own computer terminal.
• Linux-based Cluster (9 servers)
• Project space for research team
• Research Assistant available

Statistical Packages
• SAS, GRASS, R, STATA
• Matlab, HLM, Gauss, SUDAAN
MORE INFORMATION

- Links
  - Center for Economic Studies
    http://www.census.gov/ces/
  - National Center for Health Statistics
    http://www.cdc.gov/rdc/
  - Agency for Healthcare Research and Quality
    http://www.meps.ahrq.gov/mepsweb/data_stats/onsite_datacenter.jsp
- Email
  - Diane Gibson (diane.gibson@baruch.cuny.edu)
  - Shirley Liu (shirley.h.liu@census.gov)
  - http://www.ciser.cornell.edu/NYCRDC/home.shtml