

Conducting Research at the New York Federal Statistical Research Data Center

Diane Gibson, Ph.D.
NYRDC – Baruch
Executive Director
Baruch College-CUNY

Shirley H. Liu, Ph.D.
NYRDC – Baruch
Administrator
US Bureau of the Census

Any opinions and conclusions expressed herein are those of the presenters and do not necessarily represent the views of the U.S. Census Bureau.

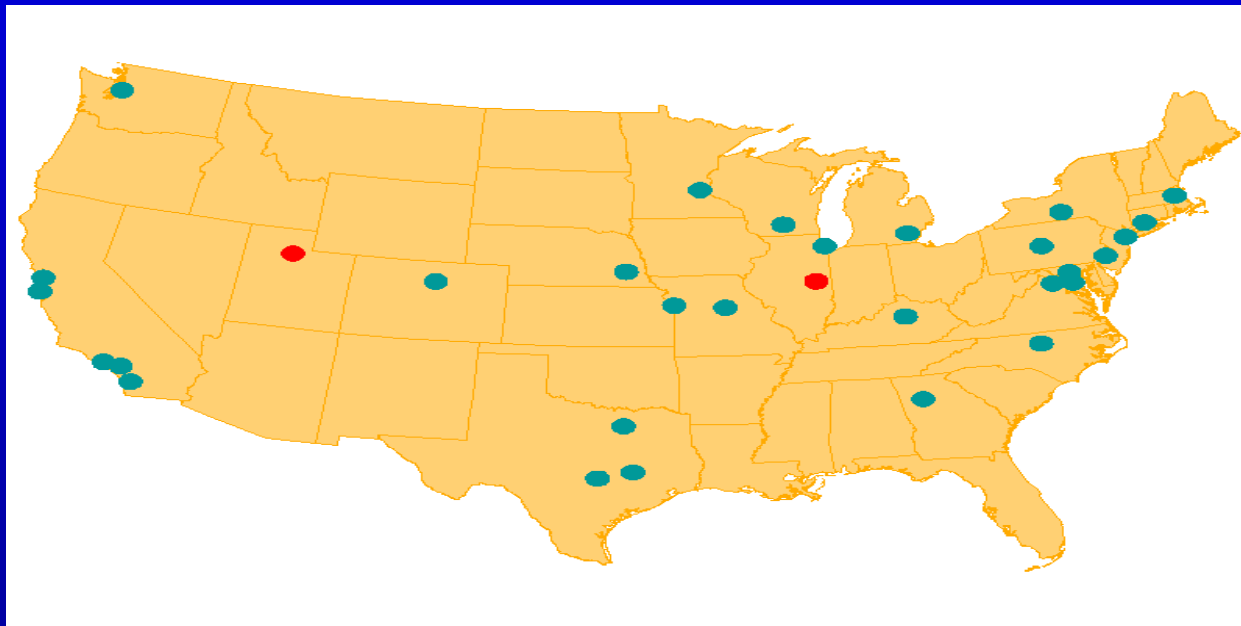
Outline of Presentation

- Overview of the Federal Statistical Research Data Center Network
- Types of questions that can be addressed using FSRDC data
- Demographic, economic, and health data available through the FSRDCs
- Application process and timeline
- Contact information and sources for additional information

Federal Statistical Research Data Center Network

The FSRDCs are secure facilities that provide researchers with approved projects with access to non-public statistical system data.

The NYRDC consortium operates RDCs at Baruch College, Yale and Cornell



Blue dots: RDCs already open; Red dots: locations soon to open

Researchers with an affiliation with one of the NYRDC consortium members do not incur fees for using the NYRDC labs

- Baruch College
- City University of New York
- Columbia University
- Cornell University
- Federal Reserve Bank of New York
- MDRC
- 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?

Are reductions in central city violent crime associated with an increase in the probability that high-income and college-educated households move into central city neighborhoods instead of the suburbs?

Data:

- 2000 Decennial Census
- 2010-2012 American Community Survey

Restricted-access variables:

- Geographic identifiers

Need for restricted data:

- Geographic identifiers used to merge in time-varying demographic, housing, and economic characteristics and central city and suburban violent crime rates

Ingrid Ellen, Katherine O'Reagan and Davin Reed. 2019. Has Falling Crime Invited Gentrification? *Journal of Housing Economics* 24:109-121.

What is the influence of the local demand for labor on SNAP participation dynamics in New York State?

Main Data Sources:

- 2010 Decennial Census
- New York State Supplemental Nutrition Assistance Program (SNAP) Administrative Records

Restricted-access variables:

- Personal Identification Key (PIK)

Need for restricted data:

- PIK used to link 2010 Census data and SNAP Administrative Records.

Erik Scherpf and Benjamin Cerf. 2019. Local Labor Demand and Program Participation Dynamics: Evidence from New York SNAP Administrative Records. *Journal of Policy Analysis and Management* 38(2): 394-425.

What happened to job mobility during the Great Recession?

Data:

- Longitudinal Employer-Household Dynamics (LEHD) infrastructure files

Restricted-access variables:

- All - no public-use microdata on employer-household dynamics

Henry Hyatt and Erika McEntarfer. 2012. "Job-to-Job Flows in the Great Recession." *American Economic Review*, 102(3): 580-83.

How did US manufacturing employment after 2000 respond to a change in US trade policy that eliminated potential tariff increases on Chinese imports?

Data:

- Longitudinal Business Database

Restricted-access variables:

- All - no public-use microdata on establishments

Justin R. Pierce and Peter K. Schott. 2016. “The Surprisingly Swift Decline of US Manufacturing Employment,” *American Economic Review* 106(7): 1632–1662.

What is the relationship between racial/ethnic residential segregation and access to health care in rural areas?

Data:

- 2005-2010 Medical Expenditure Panel Survey Household Component (MEPS-HC)

Restricted-access variables:

- Geographic identifiers

Need for restricted data:

- Geographic identifiers used to link to contextual variables from the American Community Survey and the Area Health Resources File

Molly Dondero and Jennifer Van Hook. 2017. "Racial and ethnic residential segregation and access to health care in rural areas." *Health and Place*, 43: 104-112.

What is the mother-child resemblance in dietary quality in Mexican-origin families?

Data:

- 1999/2000–2009/2010 Continuous National Health and Nutrition Examination Survey (NHANES)

Restricted-access variables:

- Household identifiers and geographic identifiers

Need for restricted data:

- Household identifiers used to link children and mothers
- Geographic identifiers used to link to contextual variables from the Decennial Census and the American Community Survey

Julia Caldwell et al. 2016. “Generational Status, Neighborhood Context, and Mother-Child Resemblance in Dietary Quality in Mexican-origin Families. *Social Science and Medicine*, 150: 212-220.

Data Available at the FSRDCs

Census Bureau

- Demographic data (household/individual)
- Economic data (firms/establishments)
- Mixed (individual 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)

Bureau of Economic Analysis (BEA)

- Company data on foreign direct investment, the activities of multinational enterprises, and international trade in services

Bureau of Labor Statistics (BLS)

- National Longitudinal Surveys of Youth 1979 and 1997

Census Demographic Data

There are public versions of the demographic 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

More Detailed Geography at the FSRDCs

Dataset	Available Years	Sampling Unit	Geography
Decennial Census	1940-2010	Household	Block
American Community Survey (ACS)	1996-2016	Housing Unit	Block
National Crime Victimization Survey (NCVS)	2006-2017	Household	Block
Survey of Income and Program Participation (SIPP)	1984-2014	Household	Tract
Current Population Survey (CPS) – March Supplement	1967-2017	Household	Tract
American Housing Survey (AHS)	1984-2017	Housing Unit	Tract
National Longitudinal Survey (NLS) – Young/Mature Men/Women	1966-1999	Individual	Tract
National Longitudinal Mortality Study (NLMS)	1973-2011	Individual	County

American Community Survey (ACS) (1996-2018)

Replaced the Decennial long form after 2010

Sample size \approx 2.9 million housing units

The questions asked include age, race, sex, educational attainment, income, place of work, occupation, household relationships, housing unit characteristics, etc.

RDC data include tract, school and congressional district, birthday, migration place code, place of work tract code. Age and mortgage expenditures are not top-coded, wages are top-coded at \$1 million dollars.

Questionnaires

<http://www.census.gov/programs-surveys/acs/methodology/questionnaire-archive.html>

Census 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)

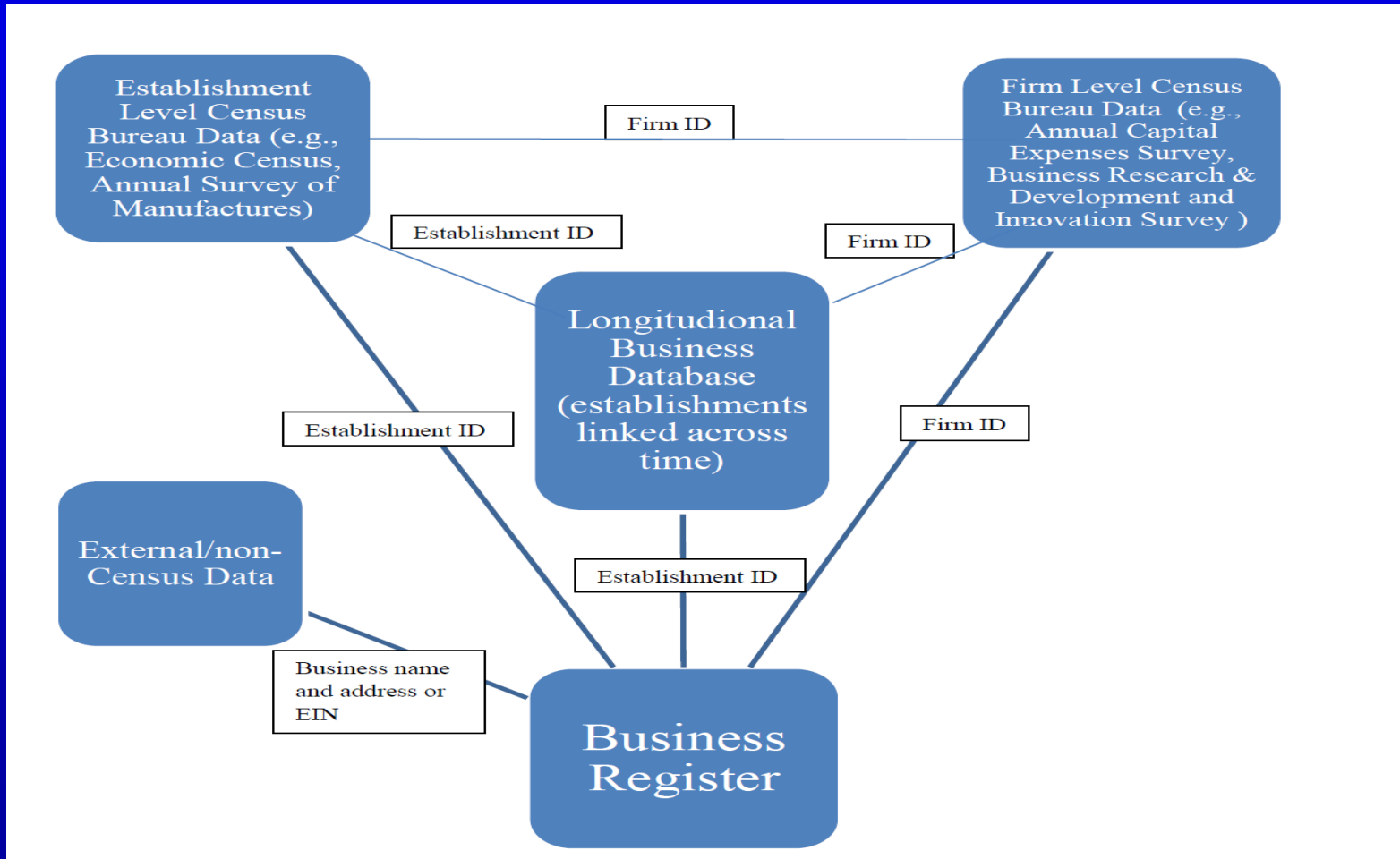
Linking Data

- Across censuses/surveys over time
- Across entities (establishment to firm level)
- External data (e.g., Kaufman Firm Survey and Compustat)

Examples of Economic Data

Data Set	RDC Years	Sampling Unit	Availability
Business Register (aka SSL)	1974-2016	Establishment	Annually
Longitudinal Business Database (LBD)	1976-2016	Establishment	Annually
Economic Censuses (Manufactures, Retail Trade, Services, Construction, Wholesale Trade, Finance & Insurance, etc)	Various (most recent 2012)	Establishment	Every 5 years
Annual Survey of Manufactures (ASM)	1973-2016	Establishment	Annually
Commodity Flow Survey (CFS)	1993 - 2012	Establishment	Every 4/5 years
Quarterly Financial Report (QFR)	1977-2015	Firm	Annually
Longitudinal Firm Trade Transactions Database (LFTTD)	1992-2016	Transaction	Monthly

Linking Business Data



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)
 - Point in time residence and country 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)

National Center for Health Statistics 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)

NCHS Data Commonly Used at the FSRDCs

Data	Frequency	Sample
National Health and Nutrition Examination Survey (NHANES)	Continuous	5,000 persons per year, all ages
National Health Interview Survey (NHIS)	Annual	42,000 households per year
National Vital Statistics System	Annual	All Births (about 4 million records annually) All Deaths (about 2.4 million records annually) Reported fetal deaths of 20+weeks gestation (about 26,000 annually)
Longitudinal Studies of Aging (LSOAs)	Biennial	7,527 persons 70 years of age and over at the time of their 1984 SOA interview.

Agency for Healthcare Research and Quality Data

Medical Expenditure Panel Survey (MEPS-HC)

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 participate in the NHIS in the prior year
- Demographic characteristics, health conditions, health status, use of medical services, charges and payments, access to care, satisfaction with care, health insurance coverage, income, employment

Restricted Variables

- Geographic detail, state identifiers
- Fully-specified ICD-9 codes (International Classification of Diseases)
- Imputed NDC for prescription drugs (National Drug Code)
- Various additional components (e.g., Provider/Insurance/Nursing Home)

Bureau of Economic Analysis Data

- U.S. Multinational Enterprises worldwide (BE-10/11)
 - Currently: 2007-2016
 - Expected: 1982-2006
- Foreign Multinational Enterprises in the United States (BE-12/15)
 - Currently: 1997-2016
 - Expected: 1981-1996
- Trade in selected services and intellectual property (BE-120/125)
 - Currently: 2007-2014
 - Expected: 2006 and 2015-2016
- Trade in financial services (BE-180/BE-185)
 - Expected: 2006-2016

The Proposal Process

Census Proposal Requirements

To access restricted Census data, each project must include

(1) Proposal

- Demonstrate scientific merit
- Document need for non–public data
- Show project is feasible given the data
- Demonstrate no risk of disclosure

(2) “Benefits Statement”

- Title 13, United States Code: all access to confidential data must benefit the Census Bureau's data collection programs
- There are 13 types of benefits a project is allowed to provide
- All projects must include a “benefits statement” specifically addressing how the project will provide two or more of these approved benefits

Proposal Requirements

Apply directly to the relevant agency to gain access to data from federal agencies other than the Census Bureau:

- NCHS

<http://www.cdc.gov/rdc/>

- AHRQ

<http://www.ahrq.gov/research/data/meps/index.html>

- BEA

<https://www.bea.gov/research/special-sworn-researcher-program>

- BLS

<https://www.bls.gov/rda/home.htm>

Once the proposal has been approved, contact the NYRDC administrator

Special Sworn Status

Apply for “Special Sworn Status” (SSS) security clearance once your project has been approved

- Must be either a U.S. citizen or a legal foreign national who has resided in the U.S. for ≥ 3 consecutive years in the last 5 years

Need to have SSS clearance before starting work at the RDC

Timeline

- Census projects typically take about 9 months or more get underway

Preparing proposal and benefits statement ~ 3 months

Census review ~ 3 months

Review by SSA and IRS if necessary ~ 3 to 6 months

SSS (concurrent with review by SSA and IRS) ~ 6 to 12 months

- NCHS and AHRQ projects take less time on average
- BEA and BLS projects timeline uncertain

Computing Environment

- 8 cubicles at the Baruch RDC – each with its own computer terminal
- Linux-based Cluster (9 servers)
- Project space for research team on the server
- Statistical Packages
 - SAS, STATA, R
 - Matlab, Gauss, SUDDAN

More Information

Links

- Center for Enterprise Dissemination (How to Apply)
<https://www.icpsr.umich.edu/web/pages/appfed/index.html>
- Center for Economic Studies
<https://www.census.gov/programs-surveys/ces/data/restricted-use-data.html>
- New York Federal Statistical Research Data Center
<https://nyrdc.cornell.edu/>
- National Center for Health Statistics
<http://www.cdc.gov/rdc/>
- Agency for Healthcare Research and Quality (MEPS)
https://meps.ahrq.gov/mepsweb/about_meps/index_researcher.jsp
- Bureau of Economic Analysis
<https://www.bea.gov/research/special-sworn-researcher-program>
- Bureau of Labor Statistics
<https://www.bls.gov/rda/home.htm>