Introduction to the Wisconsin Census Research Data Center

Economic Projects

Rachelle Hill, PhD
Administrator, MnRDC
Center for Economic Studies
U.S. Census Bureau

November 26, 2014
What is the RDC?

- Census-run facility
- Approved researchers with approved projects can access data that is not publicly available
  - Census
  - NCHS
  - AHRQ
- To be located in Sewell Social Sciences
Why use RDC data?

- Only microdata samples available
  - Full access to Economic Census
- Access to Census Business Register
  - Including Longitudinal Business Database
- Firm-linked trade data
  - Longitudinal Foreign Trade Transaction Database
  - Commodity Flow Survey
Why use RDC data?

- RDC vs. PUMS data
  - Lower levels of geography
  - No top-coding
  - Larger samples
  - Greater detail in variables
    - Race
    - Place of birth
    - Write-in responses
    - Disease codes (NHIS & MEPS-HC)
Available Data

- Economic Census
  - ASM, ARTS, AWTS, SAS
- LBD and ILBD
  - Quarterly Financial Report (QFR)
  - Survey of Business Owners (SBO)
  - MEPS-IC
Available Data

- Annual Capital Expenditures Survey (ACES)
- Business Research & Development and Innovation Survey (BRDIS)
- Manufacturers' Shipments, Inventories, and Orders (M3)
- Survey of Pollution Abatement Costs and Expenditures (PACE) and Manufacturing Energy Consumption Survey (MECS)
Available Data

- Commodity Flow Survey *(CFS)*
- Longitudinal Foreign Trade Transactions Database *(LFTTD)*
- Kauffman Firm Survey *
- LEHD *
Available Data

- Demographic and Household data
  - Decennial and ACS
  - American Housing Survey (AHS)
  - Current Population Survey (CPS) – March and December Supplements
  - Survey of Income and Program Participation (SIPP)
  - National Longitudinal Survey (NLS)
  - NCVS
Available Data

- **Recovered Data**
  - Additional CPS supplements as early as 1962
  - ASM as early as 1954
Linking Possibilities

- Internal data with PIKs or firm/establishment identifiers
  - Both surveys and administrative data
- External data CAN be linked
  - Unique individual identifiers (e.g. SSN, EIN, name, address)
  - Variables contained in both datasets (e.g. geography)
Getting Access
Census Application Process

- Identify need for data
  - www.census.gov/ces/
- Contact the RDC administrator
  - Preliminary proposal
  - 2-page summary
- Proposal and PPS
- Census (and additional agency) review
Components of Census Proposals

- Abstract
- Description of the project
- PPS (aka Benefits statement)
Project Description

- Introduction
  - Research question
  - Why you need this data
  - Benefit to Census
- Background/literature
  - Very brief (1 page single spaced)
- Methods (Focal section)
  - Data used
  - Proposed linkages
  - Analytic methods
- Plans for disclosure and disclosure risk
- Timeline and RDC usage needs
Uniqueness of Census Data

- No IRB
- No informed consent
- Mandatory participation

- Laws govern use of data (Title 13, Title 26)
  - Require “predominant purpose” be to benefit Census Bureau programs
Predominant Purpose Statement (1)

1. Evaluating concepts and practices underlying Census Bureau statistical data collection and dissemination practices, including consideration of continued relevance and appropriateness of past Census Bureau procedures to changing economic and social circumstances

2. Analyzing demographic and social or economic processes that affect Census Bureau programs, especially those that evaluate or hold promise of improving the quality of products issued by the Census Bureau

3. Developing means of increasing the utility of Census Bureau data for analyzing public programs, public policy, and/or demographic, economic, or social conditions

4. Conducting or facilitating census and survey data collection, processing or dissemination, including through activities such as administrative support, information technology support, program oversight, or auditing under appropriate legal authority

*Demographic projects only*
Predominant Purpose Statement (2)

5. Understanding and/or improving the quality of data
6. Leading to new or improved methodology to collect, measure, or tabulate a Title 13, Chapter 5 survey, census, or estimate
7. Enhancing the data collected
   - Improving imputations for non-response
   - Developing links across time or entities for data
8. Identifying the limitations of, or improving, the underlying Business Register, Master Address File, and industrial and geographical classification schemes used to collect the data
9. Identifying shortcomings of current data, collection programs and/or documenting new data collection needs
10. Constructing, verifying, or improving the sampling frame
11. Preparing estimates of population and characteristics of population
12. Developing a methodology for estimating non-response
13. Developing statistical weights
Notes about Census Proposals

- Plan ahead
  - At least 6 months to a year to get access
- Work with the RDC Administrator
- Written for a data expert rather than a content expert
  - IRS may also be a reviewer and should be considered
- Description limit is 15 pages single spaced (30 pages double)
Special Sworn Status

- Off-line paperwork and documentation
- On-line trainings and certifications
- Background check
  - Submitted online and followed with interview
    - Residential history
    - Foreign travel
    - Education and employment history
    - References
- Fingerprinting
Research Environment
Working Environment

- All work must be completed in the RDC lab
- Computing environment
  - Preliminary output viewed in the RDC
  - No internet access
  - Unix environment
  - Software: SAS, Stata, R, Matlab (others may be available)
- NCHS researchers must work when administrator is on duty
Disclosure Processes

- No output can leave RDC without review
- Plan ahead
- Performed by RDCA or the agency disclosure officer
  - Review process worked out in proposal stage
    - Turn around generally 1-2 weeks but plan for 3-4 weeks
    - Descriptive data can be problematic
Next steps
Expected Timeline

- Lab expected to open summer 2015
- Begin developing projects now
  - Check out www.census.gov/ces/
  - Contact me
  - Begin working on proposal
- Following approval, submit application for SSS
How to contact me?

- Email: rachelle.hill@census.gov
- Phone: 612-626-4365
- Schedule: Monday - Thursday & every other Friday
  - Maternity leave
    - 6 weeks off (early March through mid April)
    - 6 weeks reduced schedule (10 hours per week, mid April through late May)

- Meetings:
  - Plan to visit at least once more before March
RECOVERED DATA: HIGHLIGHTS

The economic and demographic data recovered from the Unisys will be valuable additions to the data already available at the Center for Economic Studies (CES) and the Research Data Centers (RDCs).

Most files will require additional work before they can be used for research purposes, and some may require approval by sponsoring agencies. Examples of data recovered from the Unisys include:

- Earlier years of series already available at CES
  - Censuses of Mining, Retail, Wholesale, and Services
  - Annual Survey of Manufactures (ASM)
  - Survey of Industrial Research and Development (SIRD)
  - Survey of Minority-Owned Business Enterprises
  - Commodity Transport Survey (now called the Commodity Flow Survey)
  - Decennial Census data
    ~ Puerto Rico sample and complete count files
    ~ U.S. Possessions sample and complete count files
  - Selected Current Population Survey (CPS) March Supplements

- Series not currently available at CES
  - Agriculture Surveys and Censuses
  - Annual Survey of Oil and Gas
  - Heating Fuel Survey
  - Water Use Survey
  - Survey of Construction
  - Income Surveys Development Program
  - CPS Supplements for months other than March

- New variables for series already available at CES
  - Census of Manufactures (CM) special inquiries data
  - ASM Central Administrative Office data
  - ASM and CM data flags

- Historical analysis files (see Text Box 4-2)
  - Industrial Time Series data
  - Linked CM/SIRD data created for Zvi Griliches
Information about Health Projects
Available Data

- AHRQ
  - MEPS-HC
- NCHS
  - National Health Interview Survey (NHIS)
  - National Health and Nutrition Examination Survey (NHANES)
  - National Survey of Family Growth (NSFG)
  - National Vital Statistics System (NVSS)
  - National Health Care Surveys – NAMCS, NHAMCS, NHDS, NNHS, NNAS, NSRCF, NSLTCP
Jean Abraham and Anne Beeson Royalty

Recently approved

MEPS-IC

How the provision of employer health insurance and employment outcomes are changing following the implementation of ACA.
NCHS Application Process

- Determine need for data
- Determine mode for accessing data (WI RDC)
- Prepare proposal for the agency
  1. Defined research question
  2. Identified needed restricted variables
  3. Access disclosure risk
  4. Mode of access
- Submit proposal to agency and wait for comments
AHRQ Application Process

- Contact the AHRQ Data Center manager
- Prepare proposal for the agency
- Submit proposal to agency and wait for comments

Criteria
- Feasibility
- Risk of Disclosure
- Supporting resources for the researcher
- Fits the mission of AHRQ
Notes about NCHS and AHRQ projects

- Fees for applying, merging, and technical assistance
- Relatively quick timeline
- Work directly with the agency and assigned analyst