Texas Federal Statistical Research Data Center: Opportunities for Research Using Restricted Data

---

Proposal Development Workshop

Part II – Scouting Restricted Data Sets and RDC Projects

Texas A&M University

January 31, 2020

Mark Fossett

Executive Director, Texas Research Data Center
College of Liberal Arts, Texas A&M University
College Station, Texas

Topics Covered in Part II

1. Overview Restricted Data Sets and Restricted Variables

2. Review of Examples of Projects Using Restricted Data

3. Scouting RDC Projects

Overview of Restricted Data Sets and Restricted Variables

Census Demographic and Business Data – Overview

The U.S. Census collects census and survey data in many areas including population and demography, economics and business, and more.

The following web links provide descriptions of selected data holdings and overviews of many important data sets

- **FSRDC Available Data**
- **Participating Agencies Data Sets**
- **Census CES Restricted Data Notes**

Demographic Data – Decennial Census

Decennial "Short-Form" and "Long-Form" Files

- **1950 - Long-form sample is 1%; Short-form sample is 100%**
- **1960 - Long-form sample is 25%; Short-form sample is 100%**
- **1970 - Long-form sample is 20%; Short-form sample is 100%**
- **1980 - Long-form sample is 19%; Short-form sample is 100%**
- **1990 - Long-form sample is 17%; Short-form sample is 100%**
- **2000 - Long-form sample is 17%; Short-form sample is 100%**
- **2010 – Short form only (ACS replaces Decennial Long-Form)**

Demographic Data – American Community Survey

ACS Long-form files – 1% Annual Files

All available years (2005 to present)

Restricted files have larger samples (approximately 1.5-2.0%)

Demographic Data – Other

Current Population Survey (CPS) & Supplements

American Housing Survey (AHS)

Survey of Income and Program Participation Panels (SIPP)

National Longitudinal Survey of Youth

Restricted Economic Data

Economic Data Overview

- **Economic Censuses:**
  Manufactures, Retail Trade, Wholesale Trade, Finance, Insurance and Real Estate, etc., etc.
- **Longitudinal Business Database (LBD) and Business Register**
- **Annual Survey of Manufactures (ASM)**
- **Many other firm- and establishment-level micro-data files**
- **Commodity flow and trade micro-data files**

Longitudinal Employer-Household Dynamics Data (LEHD)

- **Employment History Files (EHF)**
- **Employer Characteristics File (ECF)**

Examples of Restricted Census Variables

Finer levels of geography

- **Lat/Lon, census block, census block group, census tract, etc.**

Micro data on individuals, firms, and establishments

- **Personal Identification Key (PIK)**
- **Employer Identification number (EIN)**

Less restrictive top-coding and pre-processing

- **Detailed race codes in earlier Censuses)**
- **More detailed income data**

Restricted Health (NCHS) Data – I

National Health Status Surveys

- **National Health and Nutrition Examination Survey (NHANES I, II, and III)**
- **National Health Interview Survey (NHIS)**
- **Longitudinal Study on Aging I and II (LSOA)**
- **National Survey of Family Growth**
- **National Survey of Children's Health**
- **National Survey of Early Childhood Health**
- **National Survey of Children with Special Health Care Needs**
- **National Survey of Children with Special Health Care Needs**
- **National Asthma Survey**

Restricted NCHS Data – II

National Health Care Surveys

- **National Ambulatory Medical Care Survey**
- **National Hospital Ambulatory Medical Care Survey**
- **National Survey of Ambulatory Surgery**
- **National Hospital Discharge Survey**
- **National Nursing Home Survey (NNHS)**
- **National Home and Hospice Care Survey**
- **National Employer Health Insurance Survey**
- **National Health Provider Inventory**
- **National Immunization Survey**

Restricted NCHS Data – III

Vital Statistics

- **Mortality and Multiple Mortality**
- **Birth**
- **Fetal Death**
- **National Death Index**
- **Marriage and Divorce**

Linked Data Sets

- **Linked mortality data: NHIS, NHANES LSOA II, NNHS**
- **Linked Medicare Enrollment and Claims data: NHIS, NHANES, LSOA II**
- **Linked Social Security Administration Data: NHIS, NHANES, LSOA II, NNHS**
- **Linked EPA data**

Examples of Restricted NCHS Variables

Geographic variables (block, block group)

Most dates (date of interview, date of death, date of birth)

Income and employment data (industry codes)

Specific diagnoses (ICD-9 codes are generally coarsened)

Details about facilities (accreditation, payments, number of employees)

Some information about children and adolescents, (e.g. height and weight, depression, behavior problems, and drug use)

Contextual data (nearest hospital, % of population with diploma)

Sample design variables (necessary for estimating variances)

Restricted NCHS Data

For further information on restricted health data see the link below

http://www.cdc.gov/rdc/index.htm

Review of Examples of Projects Using Restricted Data

Project Using Restricted Demographic Data

How Low Income Neighborhoods Change: Entry, Exit and Enhancement

Researchers: Ingrid Gould Ellen and Katherine O'Regan

Datasets: American Housing Survey, Decennial Census

Summary: "This paper examines whether the economic gains experienced by low - income neighborhoods in the 1990s followed patterns of classic gentrification (as frequently assumed) – that is, through the in migration of higher income white, households, and out migration (or displacement) of the original lower income, usually minority residents, spurring racial transition in the process."

Reason for Using Restricted Data: Tract IDs

Project Using Restricted Demographic Data

Migration and Dispersal of Hispanic and Asian Groups

Researchers: William Frey and Julie Park

Datasets: American Community Survey

Summary: "This report seeks to evaluate selective migration processes of Hispanic and Asian nationality groups in the US from established settlement areas, using recent migration data from the American Community Survey. The underlying goal is to detect migration tendencies leading toward an increased dispersion of these groups associated with their migration processes."

Reason for Using Restricted Data: Finer geography, larger sample, detailed subgroups

Project Using Restricted Demographic Data

The Location of New Orleans Residents in the Year After Hurricane Katrina

Researchers: Narayan Sastry and Jesse Gregory

Datasets: American Community Survey

Summary: "Using individual data from the restricted version of the American Community Survey, we examined the displacement locations of pre-Katrina adult residents of New Orleans in the year after the hurricane. Over half (53%) of adults had returned to —or remained in—the New Orleans metropolitan area, with just under one third of the total returning to the dwelling in which they resided prior to Katrina."

Reason for Using Restricted Data: Detailed geography for location of previous residence one year ago

Project Using Restricted Demographic Data

Intergenerational Transmission of Race: Permeable Boundaries between 1970 and 2010

Researchers: Carolyn Liebler and Marie DeRousse-Wu

Datasets: American Community Survey, Decennial Census

Summary: "We study the social construction of race boundaries by investigating patterns in the race, ancestry, and Mexican origin responses provided for children of 14 types of interracial marriages using dense restricted use data from 1970 to 2010. Our broader purpose is to reveal social processes that convert a newborn child of mixed heritage into an adult person of a particular race."

Reason for Using Restricted Data: Larger sample, detailed codes for ancestry and race

Project Using Restricted Economic/Business Data

Do Environmental Regulations Disproportionately Affect Small Businesses? Evidence from the Pollution Abatement Costs and Expenditures Survey

---

Researchers: R. A. Becker, C. Pasurka, & R. J. Shadbegian

Datasets: Pollution Abatement Costs and Expenditures Survey, Annual Survey of Manufacturers, Census of Manufactures

Summary: "This study investigates whether the impact of environmental regulations differs by the size of the business … We model establishments' pollution abatement operating costs (PAOC) per unit of economic activity as a function of establishment size, industry, state, and year. Our results show that PAOC intensity increases with establishment and firm size."

Reason for Using Restricted Data: Establishment data not public

Project Using Restricted Economic/Business Data

Capital and Labor Reallocation within Firms

Researchers: Xavier Giroud & Holger M. Mueller

Datasets: Annual Survey of Manufacturers, Census of Manufactures, SSEL/Business Register, Longitudinal Business Database, Census of Auxiliary Establishments, Airline Routes (public, linked using airline name), Compustat (public, linked using Compustat-SSEL bridge)

Summary: " We document how a plant- specific shock to investment opportunities at one plant of a company ('treated plant') spills over to other plants of the same company but only if the company is financially constrained."

Reason for Using Restricted Data: Establishment data not public

Project Using Restricted Economic/Business Data

Do Housing Prices Reflect Environmental Health Risks? Evidence from Toxic Plant Openings and Closings

Researchers: J. Currie, L. Davis, M. Greenstone, & R. Walker

Datasets: Longitudinal Business Database, SSEL/Business Register, EPA Toxic Release Inventory (public, linked using TRI- SSEL Crosswalk), Housing Value (public, linked using geocode), Vital Statistics (public, linked using geocode)

Summary: "This paper measures the housing market and health impacts of 1,600 openings and closings of industrial plants that emit toxic pollutants … We find that housing values within one mile decrease by 1.5 percent when plants open, and increase by 1.5 percent when plants close."

Reason for Using Restricted Data: Establishment data not public

Project Using Restricted Economic/Business Data

University Innovation, Local Economic Growth, and Entrepreneurship

Researchers: Naomi Hausman

Datasets: Longitudinal Business Database, NSF Federal Research Funding to Universities (public, linked using geocode)

Summary: "This paper identifies the extent to which U.S. universities stimulate nearby economic activity using the interaction of a national shock to the spread of innovation from universities - the Bayh- Dole Act of 1980 – with pre- determined variation both within a university in academic strengths and across universities in federal research funding… I find that long run employment and payroll per worker around universities rise particularly rapidly after Bayh-Dole in industries more closely related to local university innovative strengths."

Reason for Using Restricted Data: Establishment data not public

Project Using Restricted Health (NCHS) Data

Disparities in Diabetes:
The Nexus of Race, Poverty, and Place

Researchers: D. Gaskin, R. Thorpe, E. McGinty, K. Bower, C. Rohde, J. H. Young, T. LaVaist, & L. Dubay

Datasets: National Health and Nutrition Examination Survey, Decennial Census (public, linked using geocode)

Summary: "We sought to determine the role of neighborhood poverty and racial composition on race disparities in diabetes prevalence…We found a race - poverty- place gradient for diabetes prevalence for Blacks and poor Whites."

Reason for Using Restricted Data: Individual health data not public

Project Using Restricted Health (NCHS) Data

Why Do They Stay? Job Tenure Among Certified Nursing Assistants in Nursing Homes

Researchers: J. Wiener, M. Squillace, W. Anderson, & G. Khatutsky

Datasets: National Nursing Home Survey, National Nursing Assistant Survey, and the Area Resource File

Summary: " This study identifies factors related to job tenure among certified nursing assistants (CNAs) working in nursing homes … Among policy- relevant domains, extrinsic rewards had the largest number of significant variables."

Reason for Using Restricted Data: Individual health data not public

Project Using Restricted Health (NCHS) Data

Racial/Ethnic Disparities in Health and Health Care among U.S. Adolescents

Researchers: May Lau, Hua Lin, & Glenn Flores

Datasets: National Survey of Children's Health

Summary: Objective: To examine racial/ethnic disparities in health status, access to care, and use of health services in U.S. adolescents. … Findings: Certain disparities were especially marked for specific racial/ethnic groups. Disparities included suboptimal health status and lack of a personal doctor or nurse for Latinos; suboptimal oral health and not receiving needed prescriptions for African Americans; no physician visit or mental health care in the past year for Asian/Pacific Islanders; obesity, insurance, receipt of routine and specialty care for American Indian/Alaskan natives."

Reason for Using Restricted Data: Individual health data not public

Project Using Restricted Health (NCHS) Data

Health Disparities and Infertility: Impacts of State-Level Insurance Mandates

Researchers: Marianne Bitler, and Lucie Schmidt

Datasets: National Survey of Family Growth

Summary: "Objective: To determine whether important racial, ethnic or socioeconomic status (SES) health disparities exist in infertility, impaired fecundity, or infertility treatment. … Results: Infertility is more common for non- Hispanic black women, non - Hispanic other race women, and Hispanic women than for non - Hispanic white women, and both infertility and impaired fecundity are more common for high school dropouts and high school graduates with no college than for 4-year college graduates, and for older women compared with women 29 and younger."

Reason for Using Restricted Data: Individual health data not public

Project Using Restricted Health (NCHS) Data

Exposures to Fine Particulate Air Pollution and Respiratory Outcomes in Adults Using Two National Datasets: A Cross-Sectional Study

---

Researchers: Keeve Nachman and Jennifer Parker

Datasets: National Health Interview Survey, EPA Air Data System (public, linked using geocode)

Summary: "The objectives of this study are to evaluate the relationship between annual average ambient fine particulate matter (PM 2.5) concentrations and respiratory outcomes for adults using modeled air pollution and health outcome data and to examine PM 2.5 sensitivity across race/ethnicity… Stratified analyses revealed significant associations for non-Hispanic blacks but not for Hispanics and non-Hispanic whites"

Reason for Using Restricted Data: Individual health data not public

Scouting RDC Projects

Scouting Previous and Current Census Projects

CES Discussion Paper Series

Every Census- based project in the RDC submits at least one working paper to the CES Discussion Paper Series

The papers can be downloaded (in PDF format) at
Archive of Working Papers Using Restricted Federal Data

The "Ideas" database is searchable

CES Annual Research Reports

Scouting Previous Health Projects

The NCHS RDC website provides a list of all publications that have come out of NCHS

- **Projects conducted in the RDC by dataset**
  - **Selected Publications Reporting Research Using Restricted Health Data**