Kentucky Research Data Center

CHARLES HOKAYEM, PhD

KRDC ADMINISTRATOR

U.S. CENSUS BUREAU

AUGUST 2019

Outline of Presentation

◦ Overview of Federal Statistical Research Data Centers (FSRDC's)

◦ Highlight Data Available in FSRDC's

◦ Access Controls

◦ Proposal Development

◦ Examples of FSRDC Projects

◦ Questions?

What is KRDC?

- NSF designated partnership between UK and U.S. Census Bureau to establish a Federal Statistical Research Data Center (FSRDC)

- Network of 29 FSRDC's (blue dots) and 2 new locations coming soon

- Qualified researchers with approved projects will be provided restricted access to microdata with safeguards to protect data confidentiality

What is KRDC?

Inaugural consortium members include:

Supporting units at UK:
◦ Office of Vice President for Research
◦ Office of Vice President for Finance
◦ Colleges of Agriculture, Arts and Sciences, Business and Economics, Medicine, Pharmacy, Public Health, Social Work

http://krdc.uky.edu

What is KRDC?

Participating Federal agencies include:
- U.S. Census Bureau
- Internal Revenue Service (IRS)
- Social Security Administration (SSA)
- National Center for Health Statistics (NCHS)
- Agency for Healthcare Research and Quality (AHRQ)
- U.S. Department of Agriculture (USDA)
- Bureau of Labor Statistics (BLS)
- More are coming!

General Advantages of FSRDC Data

- More geographic detail (some cases Census tract and block level)
  - Important for Policy Evaluation

- Additional variables
  - Census Tract Information, Genetic Information

- Larger samples
  - Decennial Census, American Community Survey

- Variables with higher top codes (earnings and income)

- More detailed coding within variables
  - Race/Ethnicity, Occupation, Exact Date of Birth, Health Diagnosis Codes

- Linkages to other administrative data
  - Birth/Death Records, Individual Data from SSA, and Business Data from IRS

FSRDC's Used to Address Many Research Topics

Business, Trade, Finance, and Management

Health and Well-Being, Health Insurance, Health Policy

Economics, Labor Markets, Entrepreneurship, Employment, and Industry

Crime and Crime Victimization

Demography, Population Distributions and Trends, Migration, and Immigration

Education and Education Policy

Hazard Mitigation, Environmental Impact Assessment, Pollution Abatement

Housing, Housing Markets, and Residential Patterns

Poverty, Social Welfare Policy, and Social Mobility

Transportation Analysis and Planning

Urban and Regional Economics and Planning

Data: Popular Census Data

Demographic Data
- Decennial Census
- American Community Survey (ACS)
- Current Population Survey (CPS)
- Survey of Income and Program Participation (SIPP)

Economic and Firm Data
- Business Register (BR)
- Longitudinal Employer-Household Dynamics (LEHD)
- Longitudinal Business Database (LBD)
- Linked Foreign Trade Transactions Database (LFTTD)
- Census and Annual Survey of Manufactures (CSM, ASM)

Data: Popular Health Data

National Health and Nutrition Examination Survey (NHANES)

National Health Interview Survey (NHIS)

National Survey of Family Growth (NSFG)

Medical Expenditure Panel Survey (MEPS can be linked to NHIS)

National Survey on Drug Use and Health (NSDUH)

National Vital Statistics System (NVSS)

National Ambulatory Medical Care Survey (NAMCS)

National Hospital Ambulatory Medical Care Survey (NHAMCS)

National Study of Long-Term Care Providers (NSLTCP)

Data: Bureau of Labor Statistics Data

Survey of Occupational Injuries and Illnesses

Census of Fatal Occupational Injuries

Employer Cost for Employee Compensation

National Longitudinal Surveys Zip Code/Census Tract Files and School Surveys

International Prices Microdata

Producer Prices Microdata

Data Linkage

- **Linked data enable broader analyses of factors that influence social, economic, and health outcomes**

- **NCHS Surveys are linked with administrative data such as**
  - The National Death Index
  - Claims data from the Centers from Medicare & Medicaid Services
  - Supplemental Security Income data from the Social Security Administration

- **Census Surveys are linked with administrative data such as**
  - IRS tax data
  - Social Security Data (earnings, retirement, disability)
  - Business databases (e.g. Compustat)
  - Extant social/economic/environmental data from 3rd party sources (e.g UKCPR National Welfare Data)

Accessing KRDC

Fees:
- Only additional cost for consortium members is for NCHS, BLS, and BEA data products
- Nominal charge for non-consortium co-author access

http://krdc.uky.edu

Access Controls

Respondent/Firm confidentiality is paramount

Researchers obtain Special Sworn Status
◦ Requires moderate level background check
◦ Oath of confidentiality is for life

Access to secure lab via two-level badge system (one UK, one Census)

Research conducted only via approved equipment in secure lab
◦ 9 thin-client station with dual 24" monitors and a room for group work
◦ Major software available (e.g. SAS, Stata, R, Matlab, QGIS, Gauss)
◦ Programs and (non-NCHS) external data submitted via KRDC Administrator

Access Controls

Analysis
- Output should be model-based; tabulations discouraged
- Mandatory disclosure avoidance review of all output
- Data stripped of SSNs; replaced with a Protected Identification Key (PIK)

Access
- 24/7 access
  - RDC Administrator **MUST** be on site for NCHS projects
- Printing available only when KRDC administrator present
- Printouts do not leave the KRDC

Access Controls

Each project has "home" RDC
- Researchers may collaborate across RDCs
- Projects may move "homes" as researchers relocate

KRDC Administrator on-site
- Guides researchers on Census proposal development
- Enforces security guidelines
- Coordinates access across RDCs
- Coordinates review of output

Project Proposals

Each agency has its own proposal review and disclosure avoidance review process

- ◦ AHRQ meps.ahrq.gov/mepsweb/data_stats/onsite_datacenter.jsp
- ◦ NCHS www.cdc.gov/rdc/B2AccessMod/ACs220.htm
- ◦ Census www.census.gov/ces/rdcresearch/howtoapply.html
- ◦ BLS https://www.bls.gov/rda/apply.htm#nonfed

All projects tracked in Census project management system

NCHS Project Proposals

Three stages of review
- Development directly with NCHS (Notify KRDC Administrator of Proposal)
- Researcher **directly** Submits Proposal to NCHS
- NCHS Review Team
  - Reviews it for disclosure (not scientific merit nor benefit to NCHS)
  - RDC analyst
  - Data representative
  - NCHS Confidentiality officer
  - NCHS RDC Director
  - Review process tends to be back-and-forth where they work with applicant to determine exactly what they need to answer the question without compromising confidentiality

Census Project Proposals

Three stages of review
- KRDC Development and Review with KRDC Administrator
  - Abstract
  - Proposal (Max 30 pages)
  - Predominant Purpose Statement
- Census Bureau Review
  - 5 concurrent reviews
- Other Agency Review
  - SSA, BLS, IRS – any agency providing some of the project's data

Census Project Proposals

Census Project Review Criteria
- ◦ Scientific merit
- ◦ Requires non-public data
- ◦ Is feasible
- ◦ Poses no risk of disclosure of individual or business
- ◦ Provides benefit to Census Bureau programs

Benefits to Census Bureau

Helps Census check data it collects, edits, and tabulates
◦ Permits rigorous analysis of confidential data
◦ Tests validity of data processing rules
◦ Evaluates conceptual and processing assumptions

Prepares new economic or population estimates

Demographic Example

"Do Walmart Supercenters Improve Food Security?" Courtemanche, Carden, Ndirangu, and Zhou. *Applied Economic Perspectives and Policy* (Forthcoming)

◦ Link external data on Walmart Supercenter openings to **Current Population Survey Food Security Supplement** to examine whether Supercenters affect food security

◦ Calculated distance from household's Census tract to nearest Supercenter

◦ Results: Greater proximity to a Walmart Supercenter improves food security

Firm Example

Greenstone, Michael, Richard Hornbeck, and Enrico Moretti. 2010. "Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings." *Journal of Political Economy* 118(3): 536-598.

◦ Linked external data on new plant openings to the **Business Register, Annual Survey of Manufactures,** and **Census of Manufactures**

◦ Calculated TFP for each plant located in the "winning" counties and "losing" counties

◦ TFP of plants in winning counties is 12% higher than TFP of plants in losing counties, suggesting that there are substantial spillovers to agglomeration

Health Example

"The Effects of CHIPRA 2009 New Option for Pregnant Non-citizens on their Hospital Utilization and Infant's Health" Hwang. In Progress

◦ Uses **National Health Interview Survey (NHIS)** to estimate effect of Medicaid/SCHIP coverage on female immigrant health care utilization and infant health outcomes

◦ Takes advantage of NHIS restricted variables in a difference-in-differences analysis:

State of Residence, State of Birth, Length of Time in US, Citizenship Status

◦ Results: Stay Tuned!

Use FSRDC Census Project as a Research Agenda

Don't think of one FSRDC project corresponding to one research paper

Think of the FSRDC project as a research agenda and include several papers

My work: Use same data under one project for several papers about nonresponse
◦ Nonresponse and Poverty (*Journal of the American Statistical Association*)
◦ Nonresponse and Distributional Effects (*Journal Political Economy*)
◦ Nonresponse and Volatility (In Progress)
◦ Nonresponse and Measurement Error (In Progress)

NCHS projects more restricted in this domain

Contact Information

James P. Ziliak, PhD

KRDC Executive Director

University of Kentucky

jziliak@uky.edu

859.257.6902

Charles Hokayem, PhD

KRDC Administrator

U.S. Census Bureau

charles.hokayem@census.gov

859.323.7075

Website: http://krdc.uky.edu

Resources for NCHS Projects

◦ Helpful NCHS Websites
  - ◦ List of available NCHS surveys and data: https://www.cdc.gov/rdc/b1datatype/Dt122.htm
  - ◦ List of Geocodes by NCHS survey: https://www.cdc.gov/rdc/b1datatype/Dt123Geocod.htm
  - ◦ Proposal Template and Sample: https://www.cdc.gov/rdc/b3prosal/pp300.htm

◦ Dedicated NCHS support for RDC Researchers
  - ◦ NCHS Analyst Assigned to Each Project
  - ◦ Send any data/project related questions to rdca@cdc.gov

◦ General Questions about CDC Data
  - ◦ https://wwwn.cdc.gov/dcs/ContactUs/Form