

Triangle Census Research Data Center

A Multi-University Effort

Restricted Health Data at Census Research Data Centers

Restricted Health Data at RDCs

- o What are Census RDCs (Research Data Centers)?
- o Why do Census RDCs have access to restricted health data?
- o What type of data is it?
- o Why is data restricted?
- o What data is restricted?
- o How can I access it?
- o Why should I want to?

What are Census RDCs (Research Data Centers)?

- o Supported by the Center for Economic Studies, Census Bureau
- o Nine centers across the US
- o Census intended as “leading source for quality data.”
- o Mission includes responsibility to “honor privacy, protect confidentiality, share our expertise globally, and conduct our work openly.”
- o Dual (and somewhat conflicting) mission necessitates existence of RDCs

What are Census RDCs?

TCRDC

Triangle Census Research Data Center

- o Secure computing lab on Duke's campus
- o Provide access to both Census microdata and confidential data from other government agencies
- o Supported by funding from Duke and UNC system
- o Use is free to Duke community and members of UNC system

Why is there health data at the Census RDCs?

- o This data is collected by:
 - o National Center for Health Statistics (NCHS)
 - o Agency for Healthcare Research and Quality (AHRQ)
- o Dual mission: to provide broad access to health data and statistics, while protecting the privacy of respondents
- o NCHS and AHRQ RDCs (in MD) created to disseminate data
- o Now available at all Census RDCs

What type of data is it?

NCHS Data

National Health Status Surveys

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

National Health Care Surveys

- o National Ambulatory Medical Care Survey
- o National Hospital Ambulatory Medical Care Survey
- o National Survey of Ambulatory Surgery
- o National Hospital Discharge Survey

- o National Nursing Home Survey (NNHS)
- o National Home and Hospice Care Survey
- o National Employer Health Insurance Survey
- o National Health Provider Inventory
- o National Immunization Survey

Vital Statistics

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

Linked Data Sets

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

What type of data is it?

NCHS data

National Health and Nutrition Examination Survey (NHANES I, II, III)

- o Periodic surveys since 1972
- o Includes data from self report, physical exam, and lab tests
 - o Nutrition measures and self-reported nutritional intake
 - o Self-reported health problems
 - o Medical, dental, and physical measurements
 - o Age, race, and ethnicity
- o Applications
 - o Sets national standards for height, weight, and blood pressure
 - o Used to estimate disease presence and risk factors

What type of data is it?

NCHS data

National Health Interview Survey (NHIS)

- o Cross-sectional, representative survey, collected regularly since 1957
- o Data collected:
 - o Self-report health status and health conditions
 - o Recent health events
 - o Utilization of medical care and health insurance
 - o Demographic, employment and income information
- o Applications
 - o Used to estimate insurance coverage and healthcare utilization
 - o Health disparities by age, race, and socio-economics status

What type of data is it?

NCHS data

Vital Stats

- o Collects data on all births, deaths, and fetal deaths annually
- o Data on all marriages and divorces before 1995
- o Data includes:
 - o Causes of death
 - o Age at death
 - o Demographics of mothers at birth
 - o Birthweight, length of gestation and sex of children
- o Applications
 - o National estimates of causes of death
 - o Rates of fetal mortality

What type of data is it?

NCHS data

- o Recent findings:
- o Vital Stats: Nearly one million fetal deaths occur annually in the US (MacDorman, 2009)
- o NHIS: In 2008, 55 million Americans (18%) had no insurance for at least part of the year. (Cohen, 2009)
- o NSFG: Nearly 4 in 10 births were to unmarried women in 2007. (Ventura, 2009)
- o NHANES: Obesity rates have tripled among young adults in the past three decades, rising from 8 percent in 1971-1974 to 24 percent in 2005-2006. (NCHS, 2008)

What type of data is it?

AHRQ data

- o Medical Expenditure Panel Survey – Household Component
 - o Large, longitudinal study of civilian non-institutionalized population
 - o Collects information from all members of household
 - o Demographics
 - o General health status and health problems
 - o Health events within survey period
 - o Healthcare utilization and expenditure
 - o Insurance coverage

What type of data is it?

AHRQ data

- o Medical Expenditure Panel Survey – Insurance and Provider Components
 - o Follows up with medical providers and pharmacists to confirm household-reported diagnoses and procedures
 - o Collects data on total charges, payments, and sources
 - o Follows up with employers and collects data on insurance options
 - o Collects data on characteristics of firms

What type of data is it?

AHRQ data

- o MEPS files include:
 - o Household component
 - o Provider Component
 - o Insurance/Employer Component
 - o Nursing Home Component (1996 only)
 - o Area Resource File
 - o Two-year two panel file
 - o MEPS-NHIS linked data
- o **Only Household Component and portions of Provider Component are publicly available**

What type of data is it?

AHRQ data

- o Recent findings:
 - o Total health care expenditures for obese adults increased 80% from 2001-2006. (Stagnitti, 2009)
 - o Young adults 19-23 were almost twice as likely to be uninsured as adults ages 45-64. (Beauregard, 2009)
 - o 13.4 million (14.6%) non-elderly adults worked full-time and lived in families below 200% of the poverty line. (Carroll, 2009)
 - o In 2005-2006, 1% of the population accounted for 18.7% of total health care expenditures. The top 5% of spenders accounted for 44% of aggregate spending, while the bottom half of spenders accounted for only 4.3% of spending. (Cohen, 2009)

Why is data restricted?

- o NCHS and AHRQ's mandating legislation demands confidentiality for participants
- o All data is collected with guarantee of confidentiality
- o Much of the data collected is sensitive and could potentially identify individuals or establishments

What is restricted?

- o Every survey has at least some data that is restricted for confidentiality
- o Data can be restricted in a number of ways:
 - o Individual variables:
 - o Removed
 - o Top-coded, bottom-coded, coarsened or masked
 - o Artificial information is substituted
 - o Pieces of datasets are restricted
 - o Whole datasets are unavailable (particularly linked files)

What's restricted?

Variables

Examples of restricted variables:

- o Geographic variables (state, county, or metropolitan area)
- o Most dates (date of interview, date of death, date of birth)
- o Income and employment data (industry codes)
- o Specific diagnoses (ICD-9 codes are generally coarsened)
- o Details about facilities (accreditation, payments, number of employees)
- o Some information about children and adolescents, (e.g. height and weight, depression, behavior problems, and drug use)
- o Some information about race, ethnicity, and country of origin
- o Contextual data (nearest hospital, % of population with diploma)
- o Sample design variables (necessary for estimating variances)

What's restricted?

Pieces of datasets

Examples

- o Contextual data: data can be linked to information about area (e.g., number of hospitals, education in county, MEPS Area Resource File)
- o Medical Expenditure Panel Survey: Provider, Insurance, and Nursing Home Component
- o NHANES III: Youth Conduct Disorder Datasets, Los Angeles Demographic Dataset, Diagnostic Interview Schedule for Children
- o National Survey on Family Growth: self-report data and interviewer comments

What's restricted?

Datasets

- o Linked data sets:

- o Mortality files linked to NHANES, NHIS, LSOA

- o EPA emissions data linked to NHDS, NHIS, NHANES

- o Social Security linked to NHANES, NHIS, LSOA

- o Medicare files linked to NHANES, NHIS, LSOA

- o Other datasets unavailable:

- o National Employer Health Insurance Survey

- o National Death Index

How can I access it?

- o Submit a proposal to NCHS or AHRQ
- o NCHS/AHRQ evaluates for feasibility, availability of computing resources, and likelihood of disclosure of confidential info (NOT for scientific merit)
- o If approved, researcher sends public use data and code
- o NCHS/AHRQ staff merges public use data with restricted data to create a file for use by researcher
- o Files are *only* created by NCHS/AHRQ staff

How can I access it?

- o Proposal must include
 - o Full research proposal
 - o Explanation of why public-use files are insufficient
 - o Data dictionary, which must identify files and years, target sample, and variables
 - o Sample code, examples of desired output, and software requirements
 - o Resumes of researchers, sources of funding, and proposed dates when analysis will take place

How can I access it?

(Working through NCHS/AHRQ)

- o Working at NCHS or AHRQ RDCs (both in Hyattsville, MD)
 - o RDC analyst prepares data prior to researcher's arrival
 - o Researchers cannot merge own data sets or work with more than one data set at time
 - o All output and notes must be reviewed before removal; data files cannot be removed
 - o Support is available from RDC staff
- o Working with NCHS remotely
 - o Researchers send code via email and receive output back via email
 - o Only certain SAS/SUDAAN procedures permitted; no access to micro data
- o Working with AHRQ remotely
 - o AHRQ has no remote server
 - o Possibility of writing task order for AHRQ

How can I access it?

TCRDC

- o Submit proposal to AHRQ or NCHS
- o Obtain special sworn status with Census Bureau
- o Working at Census
 - o Requested data supplied to RDC server
 - o Same data restrictions apply (review of output, no merging of datasets)
 - o TCRDC staff not as familiar with datasets
 - o BUT – fees are waived and you do not have to travel to Maryland!

Why do I want to access it?

Reasons for using restricted data:

- o Greater accuracy in estimates (through access to design variables)
- o Access to restricted variables
- o Access to interesting data (linked data sets)
- o Ability to link data yourself

Why do I want to access it?

- o Recent findings using restricted data:
 - o NHIS: Maternal employment increases chance of child hospitalization and accidental injury/poisoning, after controlling for mother's likelihood of being employed (Morrill 2009)
 - o MEPS-IC and Economic Census: Employers who offer health insurance have 25% greater productivity and 32% higher pay, all other variables held constant (McCue and Zawacki 2006)
 - o MEPS-IC: Higher employee contributions for insurance are associated with lower enrollment (Cooper and Vistnes, 2006)

References

- Beauregard, Karen and Kelly Carper. 2009. Characteristics of Uninsured Young Adults: Estimates for the US Civilian Non-institutionalized Population, 19-23 Years of Age, 2006. Agency for Healthcare Research and Quality, Statistical Brief #246.
- Carroll WA, Miller GE. 2009/ Full-Time Poor and Low Income Workers: Demographic Characteristics and Trends in Health Insurance Coverage, 1996–97 to 2005–06. Agency for Healthcare Research and Quality. MEPS Chartbook No. 18.
- Cohen, Robin A. and Michael Martinez. 2008. Health Insurance Coverage: Early Release of Estimates from the National Health Interview Survey. National Center for Health Statistics.
- Cohen, Steven B. and Frederick Rhode. The Concentration in Health Expenditures over a Two Year Time Interval, Estimates for the U.S. Population, 2005-2006. Agency for Healthcare Research and Quality, Statistical Brief #244.

References

- Cooper, Philip F. and Jessica P. Vistnes. 2003. “Worker's Decisions to Take-Up Offered Health Insurance Coverage: Assessing the Importance of Out-of-Pocket Premium Costs.” *Medical Care*, 41(7): III-35–III-43.
- Hyson, Rosemary and Alice Zawacki. 2008. “Health-related research using confidential US Census Bureau Data.” Center for Economic Studies 08-21.
- MacDorman M, Kirmeyer S. 2009. The challenge of fetal mortality. NCHS data brief, no 16. Hyattsville, MD: National Center for Health Statistics.
- McCue, Kristin and Alice Zawacki. 2006. “Using Census Business Data to Augment the MEPS-IC.” *Journal of Economic and Social Measurement*, 31(1): 47–67.
- Morrill, Melinda. “The Effects of Maternal Employment on the Health of School-Age Children,” under review.

References

- National Center for Health Statistics. Health, United States, 2008, With Special Feature on the Health of Young Adults. Hyattsville, MD.
- Selden, Thomas M. and Bradley M. Gray. 2006. "Tax Subsidies for Employment-Related Health Insurance: Estimates for 2006." *Health Affairs*, 25(6): 1568–79.
- Stagnitti, Marie. 2009. Trends in health Care expenditure by body mass index (BMI) category for adults in the US civilian noninstitutionalized population, 2001 and 2006. Agency for Healthcare Research and Quality. Statistical Brief #247.
- Ventura SJ. 2009. Changing patterns of nonmarital childbearing in the United States. NCHS data brief, no 18. Hyattsville, MD: National Center for Health Statistics.

Contact Information

o People:

- o Executive Director: Gale Boyd
 - o gale.boyd@duke.edu
- o Administrator: Bert Grider
 - o grider@email.unc.edu
- o Research Assistant: Daria Pelech
 - o daria.pelech@duke.edu

o Resources:

- o TCRDC website: <http://www.econ.duke.edu/tcrdc/>
- o AHRQ:
http://www.meps.ahrq.gov/mepsweb/data_stats/onsite_datacenter.jsp
- o NCHS: <http://www.cdc.gov/nchs/r&d/rdc.htm>

Further Information

o People:

- o Executive Director: Gale Boyd
 - o gale.boyd@duke.edu
- o Administrator: Bert Grider
 - o grider@email.unc.edu
- o Research Assistant: Daria Pelech
 - o daria.pelech@duke.edu

o Resources:

- o TCRDC website: <http://www.econ.duke.edu/tcrdc/>
- o AHRQ:
http://www.meps.ahrq.gov/mepsweb/data_stats/onsite_datacenter.jsp
- o NCHS: <http://www.cdc.gov/nchs/r&d/rdc.htm>

Further Information

o Events

- o Executive Director: Gale Boyd
 - o gale.boyd@duke.edu
- o Administrator: Bert Grider
 - o grider@email.unc.edu
- o Research Assistant: Daria Pelech
 - o daria.pelech@duke.edu

o Resources:

- o TCRDC website: <http://www.econ.duke.edu/tcrdc/>
- o AHRQ:
http://www.meps.ahrq.gov/mepsweb/data_stats/onsite_datacenter.jsp
- o NCHS: <http://www.cdc.gov/nchs/r&d/rdc.htm>