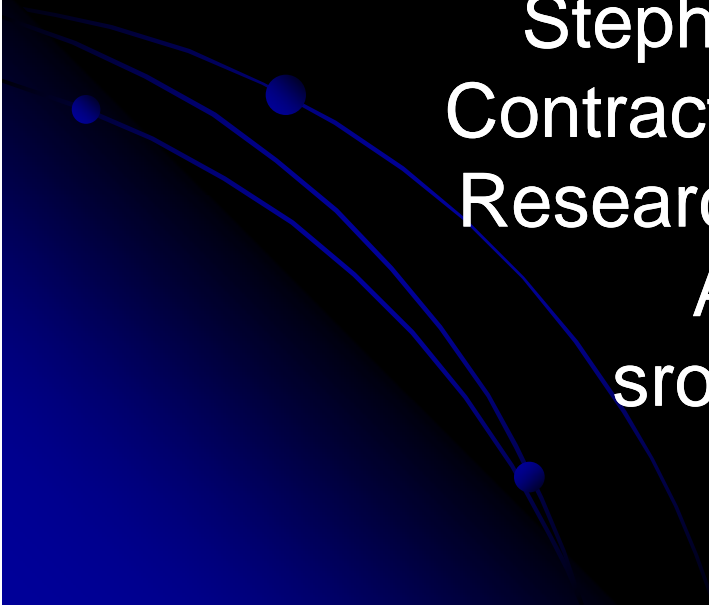
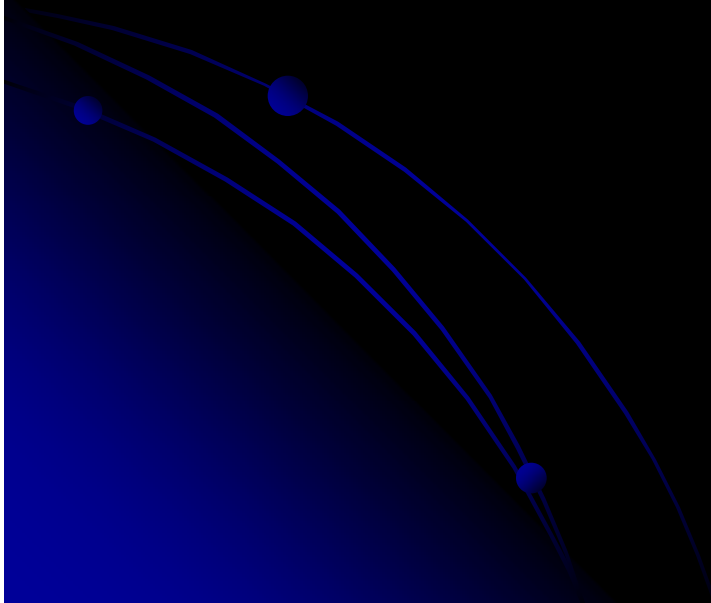


# National Center for Health Statistics Research Data Center



Stephanie Robinson, MPH  
Contractor, Northrop Grumman  
Research Data Center Analyst  
Atlanta, Georgia  
srobinson7@cdc.gov

This presentation, advertised as “New Opportunities in Health Research: Using Restricted Access Health Data at the Chicago Census Research Data Center,” was made on Jan. 29, 2010, at the Institute for Health Research and Policy at the University of Illinois at Chicago.



# Legalities

- NCHS is legally required to
  1. Collect and disseminate health information on as wide a basis as possible
  2. To do so in a manner that will not in any way harm the providers of these statistics
- Confidential Information Protection and Statistical Efficiency Act (CIPSEA) established harsh penalties
  - Up to 5 years imprisonment and up to \$250,000 in fines

# Establishment of the RDC

- Hyattsville Research Data Center
  - Established 1989
- Remote Access System
  - Established 1991
- Agreement with Census RDCs
  - Established 2007
- Atlanta Research Data Center
  - Established Spring 2009

# Confidential Information

- Direct Identifiers

- Name
- Address
- Social Security Number

- Indirect Identifiers

- Geography
- Race Ethnicity
- Date of exam, birth, or death
- Occupation

# RDC Provides Access to...

## Indirect Identifiers Necessary for Public Health Research

1. Geographic Variables
2. Content Variables
3. Genetic Variables
4. Linking Variables
5. Controlling Variables
6. Design Variables
7. Continuous/Non Top-Coded Variables


# RDC Provides Access to...

## NCHS Products Created Using Direct and Indirect Identifiers

- Linked Mortality Files
- Linked Social Security Files
- Linked Medicare/Medicaid Files
- Linked Air Quality Files (indirect)

RDC does not provide access to direct identifiers

# NCHS Surveys

- Nationally representative
  - Different collection methods
    - Laboratory Tests/Examination (NHANES)
    - Record Extraction (Health Care, Birth, Death)
    - In-person Interview (NHIS, NSFG, NHANES)
    - Random Digit Dial Interview (SLAITS)
  - Sample size changes disclosure risk
- 



# Restricted: Country of Origin

- National Health Interview Survey (NHIS)
- Sought to examine differences in overweight and diabetes prevalence based on country of origin
- Used country of origin to group into 9 regions: Europe (referent) Mexico/Central America, Caribbean, South America, Russia, Africa, Middle East, Indian Subcontinent, Central Asia, Southeast Asia
- Conclusion: Considerable heterogeneity in both prevalence of overweight and diabetes by region of birth highlights the importance of making a distinction among US immigrants to better identify subgroups at higher risks of these conditions.

Oza-Frank, R. & Narayan, V. (2009). Overweight and Diabetes Prevalence Among US Immigrants. *American Journal of Public Health*, 99(9), 1-8.

# Restricted: Census Tract

- National Health and Nutrition Examination Survey (NHANES) III
- How do neighborhood factors including segregation and the concentration of disadvantage explain ethnic disparities in body mass index?
- Used the Census tract of the NHANES respondents to add contextual information from Census to the data set.
- Discussion: The increase in BMI for Mexican-Americans associated with an increase in the proportion of Hispanics in a neighborhood is somewhat surprising given the literature on the salutatory health effects of ethnic enclaves.

Do, D.P., Dubowitz, T., Bird, C.E., Lurie, N., Escarce, J.J., & Finch, B.K. (2007). Neighborhood context and ethnicity differences in body mass index: a multilevel analysis using the NHANES III survey (1988-1994). *Economics and Human Biology* 5, 179-203.

# Restricted: Genetic Data

- National Health and Nutrition Examination (NHANES) III
- Purpose: Estimate allele frequency and genotype prevalence for 90 variants in 50 genes chosen for their potential public health significance by age, sex, and race/ethnicity in non-Hispanic whites, non-Hispanic blacks, and Mexican Americans.
- Potential Use: Provide reference for investigations into US population structure, for examinations of gene-disease associations in the NHANES data set, for calculation of attributable risk, and for design of future studies aiming to discover associations of alleles and genotypes with common diseases.

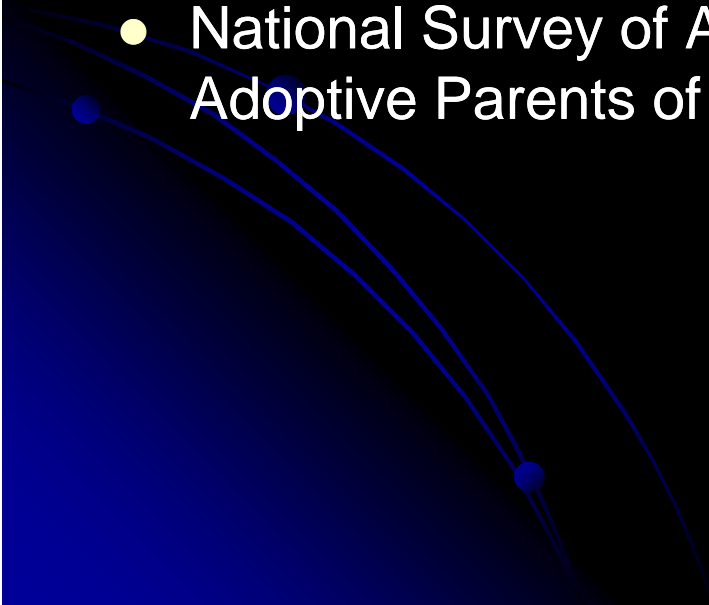
Chang, M. et al. (2008). Prevalence in the United States of Selected Candidate Gene Variants. *American Journal of Epidemiology*.

# Restricted: NNHS-NNAS Linking Variable

- National Nursing Home Survey (NNHS) and the National Nursing Assistant Survey (NNAS)
- Examined the factors influencing CNAs tenure
- Conclusions: Wages, fringe benefits, job security, and alternative choices of employment are important determinants of job tenure that should be addressed.

Anderson, W.L., Wiener, J.M., Squillance, M.R., & Khatutsky, G. (2009). Why Do They Stay? Job Tenure Among Certified Nursing Assistants in Nursing Homes. *The Gerontologist*.

# Linking Variables

- National Home and Hospice Care Survey → National Home Health Aide Survey
  - National Survey Children's Health → National Survey of Children with Special Health Care Needs
  - National Survey of Adoptive Parents → National Survey of Adoptive Parents of Children with Special Health Care Needs
- 

# Other Examples

## NHIS Study of Occupation and Morbidity/Mortality

- Industry and Occupation
- Mortality Files

## NAMCS Study of Medical Training in Emergency Departments

- Emergency medicine residence completion
- Emergency medicine board completion

## NHANES Study of STI prevalence

- Adolescent sexual behavior and STI information
- Region

# More Examples

## NSFG Study of Pregnancy in American Indian women

- Race/ethnicity

## NHANES Studies of Vitamin D

- Latitude → Sun Exposure
- Date of Exam → Seasonality

## NHIS Study of Region and Diabetes

- Duration of Residence → Acculturation
- Age at Migration → Acculturation
- Citizenship Status → Acculturation

# RDC Provides Access to...

## NCHS Products Created Using Direct and Indirect Identifiers

- Linked Mortality Files
- Linked Social Security Files
- Linked Medicare/Medicaid Files
- Linked Air Quality Files (indirect)

RDC does not provide access to direct identifiers



# Linked Mortality

## Restricted: Mortality data

- NHANES III 1988-1994
- Question: How does overall obesity and body fat distribution predict risk of mortality?
- Findings: Waist-to-hip ratio (WHR) in women associated with mortality in middle-age women. BMI and waist circumference (WC) exhibited U- or J-shaped associations. In older adults, a higher BMI in both sexes and WC in men were associated with increased survival.

Reis, J.P., Macera, C.A., Araneta, M.R., Lindsay, S.P., Marshall, S.J. & Wingard, D.L. (2009). Comparison of Overall Obesity and Body Fat Distribution in Predicting Risk of Mortality. *Obesity*.

# Linked Mortality

- National Health Interview Survey 1986-2004
- NHANES I Epidemiologic Follow-up Study 1971-1992
- NHANES II 1976-1980
- NHANES III 1988-1994
- NHANES 1999-2004
- The Second Longitudinal Study of Aging 1994-2000
- National Nursing Home Survey 1985, 1995, 1997, 2004

## Potential Study Questions:

- What is the association between health status and mortality?

# Linked Social Security

- National Health Interview Survey 1994-2005
- NHANES I Epidemiologic Follow-up Study 1971-1992
- NHANES III 1988-1994
- NHANES 1999-2004
- The Second Longitudinal Study of Aging 1994-2000
- National Nursing Home Survey 1985, 1995, 1997, 2004

## Potential Study Questions:

- What is the association between health status and characteristics of Social Security disability applicants and recipients?

# Linked Medicare

- National Health Interview Survey 1994-1998
- NHANES I Epidemiologic Follow-up Study 1971-1992
- NHANES II 1976-1980
- NHANES III 1988-1994
- The Second Longitudinal Study of Aging 1994-2000

## Potential Study Questions:

- How have health status and health care utilization/expenditures changed over time in the elderly and disabled population?

# Linked Air Quality

## EPA Air Pollution Data Linked by

- Block Group to NHIS to 1986-2005
- Zip Code and Admin Date to NHDS 1999-2005
- Block Group and Exam Date to NHANES III

## Possible Study Questions:

- How do air pollution values affect prevalence of childhood asthma?
- How do sudden increases in air pollution affect admissions for respiratory diseases?

# Summary of Restricted Variables

- Geography to add policy
- Geography to add context
- Geography
- Genetic data
- Linking within surveys
- Industry and occupation
- Sensitive sexual information
- Smaller racial/ethnic groups
- Doctor characteristics
- Acculturation variables
- Linkage products

# Proposal Review Process

- Submit a Proposal
  - Research Question
  - Public Health Benefit
  - Data Needed
  - Sample Output
- Review Committee
  - RDC Analyst, Confidentiality Officer, RDC Director, Representative from the Data System(s)
  - 6-8 Weeks
  - Assess Disclosure Risk

Peter Meyer  
Research Data Center Director  
Hyattsville, MD  
pmeyer1@cdc.gov

Stephanie Robinson  
Research Data Center Analyst  
Atlanta, Georgia  
srobinson7@cdc.gov

