



American Indians and Alaska Natives (AIAN) in National Survey Data

Pamela Jo Johnson, MPH, PhD,
State Health Access Data Assistance Center
University of Minnesota

AIAN MCH Conference Workshop
April 25, 2008



Overview

- Presenter's background
- National survey data for AIAN research
 - NCHS sources
 - CDC sources
 - Access to data
- Overview complex survey data
- Analysis of NSFG data



Federal Data Sources

- NCHS
 - Vital Statistics
 - Surveys
 - NHIS – National Health Interview Survey
 - NHDS – National Hospital Discharge Survey
 - NSCH – National Survey of Child Health
 - NSFG – National Survey of Family Growth
- CDC
 - BRFSS – Behavioral Risk Factor Surveillance System
 - YRBSS – Youth Risk Behavior Surveillance System
 - PRAMS – Pregnancy Risk Assessment Monitoring System



NCHS Data Sources



Vital Records

- Natality, 2005
 - Total number of registered births: 4,138,349
 - AIAN registered births: 44,813
- Mortality, 2005
 - Total number of records: 2,452,506
 - AIAN deaths: 13,918
- LBD, 2004
 - Race of mother is AIAN
 - Births: 43,927 (N=4,112,055)
 - Infant deaths: 371 (N=27,860)
 - Neonatal deaths: 187 (N=18,602)
 - Post-neonatal deaths: 184 (N=9,258)



Natality (Birth) Data

- Over 4 million births annually in US
 - Births to foreign residents occurring in US are included
 - Births to US citizens occurring outside US are not included
- Files released for each calendar year
 - Public use data available for 1968 – 2005
 - 100% coverage since 1985
 - Separate files for US states and for US territories
 - Geographic identifiers
 - States, MSA, metro/non-metro county
 - County and City > 100,000 population
 - NO geographic identifiers beginning with 2005



Nativity Data Elements

- Mother and father
 - Age
 - Education
 - Hispanic origin
 - Race
 - Marital status
- Labor & Delivery
 - Medical risk factors
 - Obstetric procedures
 - Method of delivery
 - Complications of L & D
 - Attendant at birth
- Mother
 - Nativity
 - Pregnancy history
 - Prenatal care
 - Tobacco use
 - Weight gain
- Infant
 - Birth weight (grams)
 - Gestational age (wks)
 - Plurality (single, twin)
 - Abnormal conditions
 - Congenital anomalies



Mortality (Death) Data

- Over 2.3 million death certificates filed annually
 - Represents **all** deaths occurring in the US
 - Deaths of US citizens and members of the Armed Forces occurring outside the US **are not** included
- Files released for each calendar year
 - Public use data available for 1968 – 2005
 - 100% coverage since 1983
 - Separate files for US states and for US territories
 - Geographic identifiers
 - States, MSA, metro/non-metro county
 - County and City > 100,000 population
 - **NO geographic identifiers beginning with 2005**



Mortality Data Elements

- **Decedent**
 - Age at death
 - Race & Hispanic origin
 - Sex
 - Education
 - Marital status
 - Industry or business
 - Occupation (usual)
 - Place of birth
- **Death event**
 - Day of week
 - Month of death
 - Place of death
 - Manner of death
 - Underlying cause of death (ICD codes)
 - Multiple condition codes
 - Place of injury



Linked Birth/Infant Death Data

- **Birth and death data for infants born in the US who die in the US within first year**
 - ~ 27,000 infant deaths each year linked with data from the birth certificate (numerator)
 - ~ 4 million birth records (denominator)
 - 2-3% of infant deaths not linked (unlinked)
- **Benefits of LBD for infant mortality analyses**
 - Additional variables available from the birth certificate useful in infant mortality analysis
 - Numerator and denominator in a single dataset
 - Race of mother used to classify infants so numerator and denominator are comparable



Linked Birth/Infant Death

- Cohort files (1983 – 1991, 1995 – 2001)
 - Infants born in a year (cohort), linked with death data for infants that die within first year
 - Gold standard for infant mortality research
 - Timeliness of data availability is a problem
- Period files (1995 – 2004)
 - Infant deaths in a year (period), linked with birth data and all other births in same year
 - Standard for national infant mortality surveillance
- No national linked files produced 1992 – 1994



National Health Interview Survey



National Health Interview Survey

- Purpose
 - Monitor health and healthcare of the U.S. population
- Target Population
 - Non-institutionalized, civilian population in U.S.
- Frequency
 - Continuous data collection throughout year
 - Annual data release, usually in June for previous year
- NHIS fielded annually since 1957
 - Longest running national health survey in the U.S.



NHIS Data

- Data collection
 - Household, in-person interview
 - Survey 100,000 people in 40,000 households each year
 - Core questionnaires – all persons
 - Sample adult/sample child for additional information
 - Supplemental questions, which may vary by year
- Data availability
 - 1968 – 2006
 - AIAN identifiable since 1978
- Public-use data files
 - http://www.cdc.gov/nchs/about/major/nhis/quest_data_related_doc.htm
 - <http://www.ihis.us>



NHIS, 2006

- File characteristics
 - Datafiles = 6
 - HH, person, sample adult, sample child, injury, injury verbatim
 - Variables = 1,423
 - Records = 75,716
 - Format = ASCII data with Stata, SAS, or SPSS syntax
- Geography
 - Census regions
- AIAN sample
 - AIAN alone = 621
 - AIAN alone or main race = 767



Content of NHIS

- Health conditions
 - Chronic diseases
 - Disability status
- Health care access
 - Health insurance
 - Health care utilization
 - Barriers to care
- Health behaviors
 - Tobacco use
 - Alcohol use
 - Physical activity



Integrated Health Interview Series

- IHIS is an interactive, web-based, data extraction system
 - Original NHIS data recoded for comparability over time
 - Free to any user
 - User specifies file format, years of data, and variables
- Currently over 1,000 variables for 1969 – 2006
 - AIAN adults and children identified from 1978 forward
 - Really good comparability for 1997 forward
- July 2008 will include over 2,000 variables
 - Linked mortality files 1984 – 2000; Cancer screening
 - 1968 data



- Available at: <http://www.ihis.us>

National Survey of Children's Health (SLAITS)



National Survey of Children's Health

- Purpose:
 - To examine aspects of children's health and well-being (including physical and mental health, health care, social well-being) as well as aspects of family & neighborhood that affect children's health
- Target population
 - Children ages 0-17 years old
- Frequency
 - Periodic
- One of the SLAITS surveys
 - The State and Local Area Integrated Telephone Survey (SLAITS) collects important health care data at State and local levels



SLAITS Data

- Data collection
 - State and Local Area Integrated Telephone Survey
 - Random digit dial (RDD) sample of households with children < 18 years
 - One child randomly selected as subject of survey
- Data availability
 - National Survey of Children's Health (NSCH)
 - Data collected 2003-04
 - Data collected 2007-08 (**not released yet**)
 - Available at
 - <http://www.cdc.gov/nchs/slaits.htm>



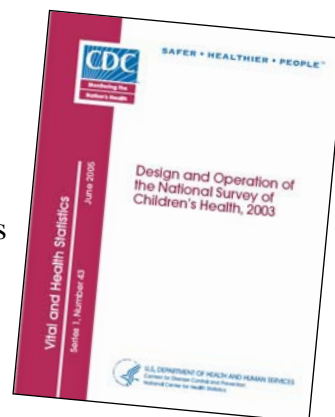
NSCH, 2003

- Data Collection: Jan 2003 - Jul 2004
- File characteristics
 - Datafiles = 1
 - Variables = 301
 - Records = 102,353
 - Format = SAS file
- Geography
 - States
- AIAN sample
 - AIAN alone or main race = 1,017
 - Only identified in AK, AZ, MT, ND, NM, OK, SD



Content of NSCH

- Child
 - Health and functional status
 - Health insurance, healthcare access and utilization
 - Medical home
 - School, after-school experiences
- Family
 - Family interactions
 - Parent health
- Neighborhood
 - Neighborhood characteristics



Other SLAITS Data

- SLAITS data availability
 - National Survey of Early Childhood Health (NSECH)
 - Data collected 2000
 - National Survey of Children with Special Health Care Needs (NSCSHCN)
 - Data collected 2000-02; 2005-06
 - National Asthma Survey (NAS)
 - Data collected 2003
 - National Survey of Adoptive Parents (NSAP)
 - Data collected 2007
 - Survey of Adult Transition and Health (SATH)
 - Follow-back sample, data collected 2007



National Hospital Discharge Survey



National Hospital Discharge Survey

- Purpose
 - The National Health Care Surveys are a family of provider-based surveys designed to meet the need for objective, reliable information about the organizations and providers that supply health care, the services rendered, and the patients they serve.
- Target Population
 - Nationally representative sample of inpatient discharges
 - Non-Federal, short-stay hospitals in the United States
 - Excludes federal, military, and VA hospitals; hospital units of institutions such as prisons; and hospitals with < 6 staffed beds
- Frequency
 - Annual data release



NHDS Data

- Data collection
 - Sample
 - Hospitals:
 - About 500 hospitals sampled per year
 - Discharges:
 - Over 300,000 sampled per year
 - Chart abstraction
 - NHDS are data about *hospitalizations*, not people
- Data availability
 - 1996 – 2005
 - <http://www.cdc.gov/nchs/about/major/hdasd/nhds.htm>



NHDS, 2005

- File characteristics
 - Data files = 1
 - Variables = 32
 - Records = 375,372 (includes 39,702 newborns)
 - Represents 34.7 million discharges (excluding newborns)
 - Format = ASCII file
- Geography
 - Census regions
- AIAN sample (31% of all records missing race)
 - AIAN = 1,034 hospitalizations
 - 150 AIAN women hospitalized for delivery
 - 155 AIAN newborns



Content of NHDS

- Patient & Hospital characteristics
 - Age, sex, race
 - Number of discharges
 - Average length of stay
- Diagnoses
 - e.g. Childbirth
- Procedures
 - e.g. C-section



Other Health Care Surveys

- **Ambulatory and hospital care**
 - National Ambulatory Medical Care Survey (NAMCS)
 - National Hospital Ambulatory Hospital Medical Survey (NHAMCS)
 - National Survey of Ambulatory Surgery (NSAS)
 - National Hospital Discharge Survey (NHDS)
- **Long-term care**
 - Nursing Home Survey (NNHS)
 - National Home and Hospice Care Survey (NHHCS)
- **Data availability**
 - <http://www.cdc.gov/nchs/nhcs.htm>



CDC Data Sources



Behavioral Risk Factor Surveillance System

- Purpose
 - a state-based system of health surveys that collects information on health risk behaviors, preventive health practices, and health care access primarily related to chronic disease and injury
- Target Population
 - Non-institutionalized, adult population in U.S.
- Frequency
 - Continuous collection; Annual data release
- Data collection
 - Telephone interview by states
- Data availability
 - 1984 – 2007 (all states represented since 1994)



BRFSS, 2006

- File characteristics
 - Datafiles = 1
 - Variables = 302
 - Records = 355,710
 - Format = ASCII or SAS transport file
- Geography
 - State, county, zip code, residence within or outside a metropolitan statistical area
- AIAN sample
 - 6,234 respondents reported AIAN as the race that best described them



Pregnancy Risk Assessment Monitoring System

- Purpose
 - PRAMS collects state-specific, population-based data on maternal attitudes and experiences before, during, and shortly after pregnancy.
- Target Population
 - Women (approximately 1,300-3,400 per state) who delivered a live-born infant in the past 2 to 4 months
- Mailed survey
- Health-related content
 - Attitudes and feelings about most recent pregnancy,
 - Content and source of prenatal care, Infant health care,
 - Alcohol and tobacco consumption, Contraceptive use, Physical abuse before and during pregnancy



PRAMS Data Availability

- PRAMS data from multiple states
 - Requests reviewed by CDC and participating PRAMS states using a standard proposal format.
 - See proposal guidelines
- PRAMS data for a single state
 - Requests should be directed to state's PRAMS coordinator
- Data availability
 - 1988 – 2004 (varies by state)
 - <http://www.cdc.gov/prams/index.htm>



PRAMS Participation

Thirty-seven states, New York City, and South Dakota (Yankton Sioux Tribe) currently participate in PRAMS. Six other states previously participated.


The map shows the following states in blue (current participation): Washington, Oregon, California, Nevada, Arizona, New Mexico, Texas, Colorado, Kansas, Oklahoma, Nebraska, Minnesota, Iowa, Missouri, Arkansas, Louisiana, Mississippi, Alabama, Georgia, Florida, South Carolina, North Carolina, Virginia, West Virginia, Maryland, Delaware, Pennsylvania, New Jersey, New York, Connecticut, Rhode Island, Massachusetts, Vermont, New Hampshire, and Maine. States in light green (previous participation): Montana, Wyoming, Idaho, Utah, New Mexico, Arizona, California, Nevada, Oregon, Washington, Alaska, and Hawaii.

**SHA
DAC**
State Health Assessment
Data Analysis Center

SOURCE: <http://www.cdc.gov/prams/states.htm>; last modified 06/22/2007

Youth Risk Behavior Surveillance System

- Purpose
 - Survey established by CDC to monitor prevalence of youth behavior that most influences health. The priority health risk behaviors that contribute markedly to the leading causes of death, disability, and social problems among youth and adults in the United States
- Target Population
 - 9th to 12th grade students
- Frequency
 - Biennial
- Content
 - Health behaviors
 - Sexual health
 - Behavior related to injury and violence



SHA
DAC
State Health Assessment
Data Analysis Center

YRBSS Data

- Data collection
 - Self-administered questionnaire in classroom
- Data availability
 - National data
 - 1991 – 2005 (odd years)
 - <http://www.cdc.gov/HealthyYouth/yrbs/data/index.htm>
 - State and local data
 - Controlled by jurisdictions that conducted the survey
 - Some have given CDC permission to distribute data upon request
 - Others manage the distribution of their data themselves
 - For information on acquiring data files contact CDC
 - email at cdc-info@cdc.gov
 - call 1-800-CDC-INFO

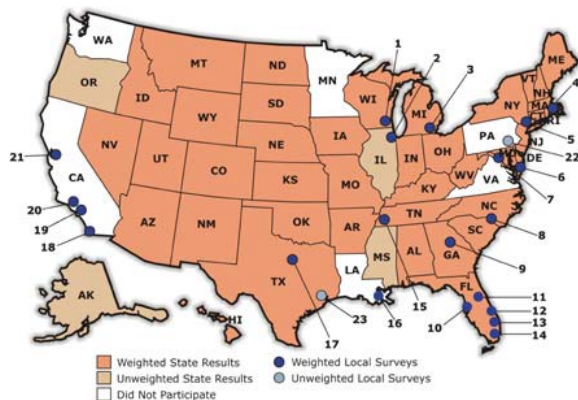


YRBSS, 2005

- File characteristics
 - Data files = 1
 - Variables = 204
 - Records = 13,917
 - Format = ASCII, SAS, SPSS files
- Geography
 - National, no geographic identifiers in public file
- AIAN sample
 - 147 respondents classified as AIAN-only, non-Hispanic
 - Original response variables allow identification of 536 respondents who identified AIAN alone or in combination with other races or with Hispanic ethnicity



YRBSS Participation, 2005



SOURCE: <http://www.cdc.gov/HealthyYouth/yrbss/map.htm>

Accessing NCHS Data



Accessing NCHS Data

- Public-use Data Files
 - Download
 - Request CD-ROM
- NCHS Research Data Center (RDC)
 - Remote submit
 - Staff-assisted programming
 - On-site access
- Census RDCs



RDC Data Access Procedures

- Research Data Center
 - At NCHS in Hyattsville
- To access NCHS restricted data a user must:
 - Submit a research proposal
 - Review committee receives, reviews, approves researcher proposals
 - Proposals are evaluated primarily on the confidentiality disclosure risk.
 - Scientific merit is not an evaluation criteria.
 - Sign an affidavit of confidentiality and promise not to use any method to attempt to identify respondents
- More information
 - <http://www.cdc.gov/nchs/r&d/rdc.htm>



NCHS RDC Access Methods

- Three methods to access restricted data
 - On-Site access
 - Use local computing resources at NCHS RDC
 - Remote access
 - Submit programs electronically to be executed in the RDC with output returned by email
 - Staff-assisted programming
 - RDC staff provide on-site programming for off-site approved researchers
- For all methods, restricted data files remain at RDC, output is inspected for disclosure violations



On-Site Access

- RDC staff constructs necessary data files, including merged user data
- Most statistical packages available with sufficient lead time
 - SAS, Stata, SUDAAN, LIMDEP, HLM, SPSS
 - Other software can be made available with sufficient lead time
- Output subject to disclosure review
- Open during normal business hours



Remote Access Method

- RDC constructs data files, including merged user data
- SAS programs only
 - Certain SAS procedures and functions not allowed
 - Additional software options expected in future
- Submitted programs and output undergo a programmed disclosure limitation review
- Features of **AN**alytic **D**ata **R**esearch by **E**-mail (ANDRE)
 - Registered subscribers only
 - Proposals already reviewed and approved
 - Unlimited access during the subscription period
 - Completely automated system
 - Operates round the clock without human intervention
 - Registered users submit data requests by email at any time
 - Results released to specified email address that has been certified as secure and approved by NCHS/RDC



RDC Staff-assisted Programming

- Subcontract with the RDC staff to perform programming tasks
- Useful for those planning to use statistical software not available for the remote system and who are not able to travel to the RDC facility
- Cost is estimated for each research project



Staff-Assisted Remote Access

- Analysis done through a particular staff member
 - RDC staff determine cost based on time
 - RDC staff create data, run e-mailed programs, check and return output to researcher
- Staff can do exploratory analysis, if needed
- Staff can help check modeling problems
- Commonly done after on-site visit



RDC On-Site Fees

- New file creation
 - Minimum setup charge of \$500 per day (\$250 for the Mortality file)
 - Additional \$500 per day is charged as needed for special handling, such as merging of additional data or creating custom file formats
 - Complex projects may require discussion with RDC staff to determine cost of file creation
- Daily programming costs
 - \$200 per day (consecutive 2-day minimum and 10-day maximum, with extensions negotiated subject to scheduling requirements)



SOURCE: <http://www.cdc.gov/nchs/r&d/rdcuserfees.htm> ; last reviewed May 23, 2007

Remote Access Fees

- **Linked Mortality files**
 - \$250 per month for a single survey wave
 - \$125 per month for each additional survey wave
- **National Survey of Family Growth Contextual Data**
 - \$500 per year
- **Polio file for National Health Interview Survey**
 - \$500 per year
- **For all other files:**
 - \$500 per month for a single survey wave file
 - \$250 per month for each additional survey wave



SOURCE: <http://www.cdc.gov/nchs/r&d/rdcuserfees.htm> ; last reviewed May 23, 2007

Census RDCs

- | | |
|---|--|
| <ul style="list-style-type: none"> • Boston Census Research Data Center
National Bureau of Economic Research
Cambridge, MA 02138
617-613-1215 • California Census Research Data Center
University of California
Berkeley, CA 97420-5100
510-643-2262 • California Census Research Data Center
University of California at Los Angeles
Los Angeles, CA 90095-1484
310-267-5588 • Center for Economic Studies
US Bureau of the Census
Washington, DC 20233
301-763-8840 • Chicago Census Research Data Center
Federal Reserve Bank of Chicago
Chicago, IL 60604-1413
312-423-4692 | <ul style="list-style-type: none"> • Michigan Census Research Data Center
Institute for Social Research
University of Michigan
Ann Arbor, MI 48106-1248
734-615-2535 • Baruch School of Public Affairs
New York, NY 10010
646-660-6788 • Cornell Census Research Data Center
CISER, Cornell University
Ithaca, NY 14850
607-255-8603 • Triangle Census Research Data Center
Duke University
Durham, NC 27708-0097
919-660-6893 |
|---|--|



SOURCE: <http://webserver02.ces.census.gov/index.php/ces/researchlocations>