

**National Center for Health Statistics
Office of Analysis and Epidemiology
Special Projects Branch
Record Linkage Program**

**Christine S. Cox, SPB Branch Chief, OAE
NCHS Board of Scientific Counselors Meeting
April 24, 2008**

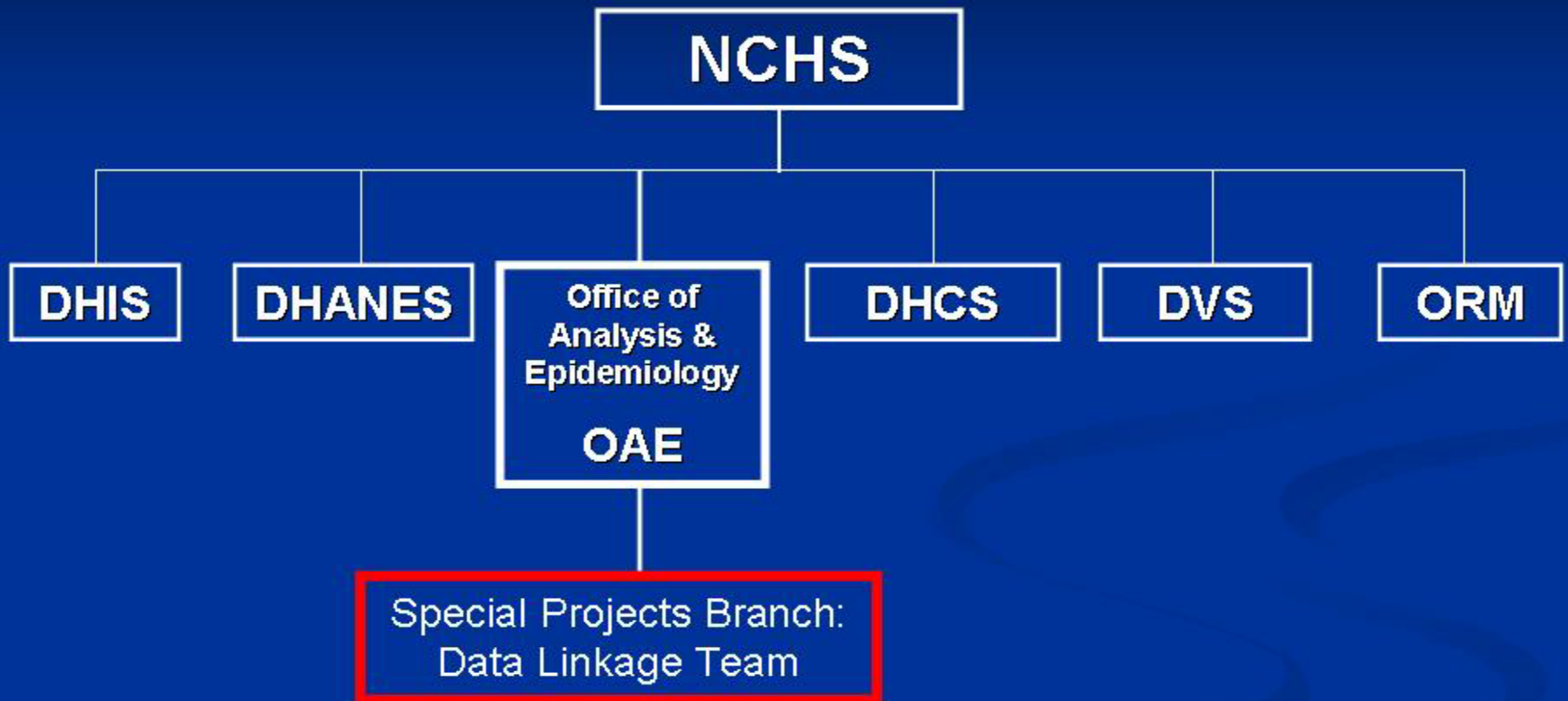


SAFER • HEALTHIER • PEOPLE™

**U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Centers for Disease Control and Prevention
National Center for Health Statistics**



NCHS Organizational Chart



Data Linkage & Tracking Teams

■ Data Linkage Team

- Stephanie Bartee (contractor)
- Jim Brittain (contractor)
- Cordell Golden
- Kimberly Lochner
- Donna Miller
- Gloria Wheatcroft

■ Data Tracking Team

- Dawn Scott
- Keith Zevallos

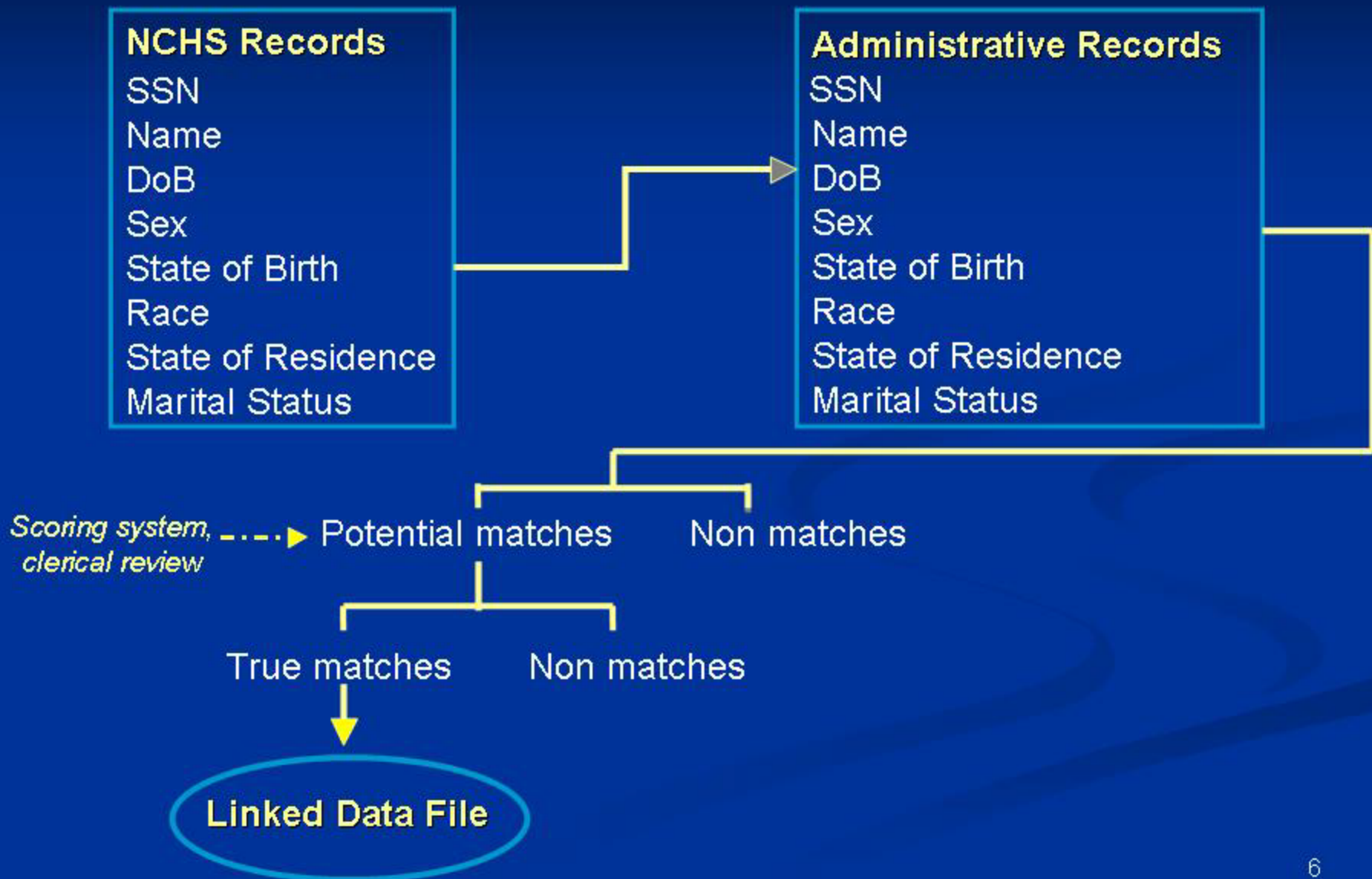
Why Do Linkage

- Augments available information for major diseases, risk factors, and health service utilization
 - Links exposures to outcomes
 - Provides longitudinal component to survey data
- Reduces cost burden
 - Re-contacting survey respondents for follow-up information can be expensive
- Increases accuracy and detail of data collected

Types of Data Linkage

- Person-level or facility level record
 - Person survey data linked with administrative data (e.g. Medicare)
 - Hospital survey data linked facility characteristics (e.g. American Hospital Associations Annual Survey of Hospitals)
- Contextual data
 - Geocoded to standard Census geo-areas
 - Census data – population & housing
 - EPA data – Environmental air quality
 - State level data – generosity of Medicaid payments

How Records are Linked



Research Potential of NCHS Linked Data

- Aging
 - Risk factors for poor health outcomes (hip fractures, stroke, etc.)
- Disability
 - Effects of chronic illness and obesity on disability and mortality
- Disparities
 - Mortality patterns by race/ethnicity or socioeconomic status
- Health Services
 - Functional impairment and health care costs
- Genetics
 - Genetic variants and health outcomes
- Methodologic Studies
 - Validation of self-reports vs. administrative records

NCHS Linkage Program

- Early 1980's
 - NMCUES link to Medicare in 1981-1986
 - NHEFS
 - NDI linkage in 1982-1992
 - Medicare linkage in 1980-1986
- 1990's – NHIS mortality linkage to NDI
- 2000 to present – NCHS expanded linkage program
 - Division specific linkages
 - OAE/SPB record linkages

Division Linkage Activities

- Division of Vital Statistics
 - Linked Birth and Infant Death Files
 - 1983-1991 (birth cohort linkage)
 - 1995-2004 (period and birth cohort linkages)
- Division of Health Care Statistics
 - 2004 National Nursing Home Survey
 - CMS Long Term Care Minimum Data Set on resident assessments and facility characteristics
- Division of Health Interview Survey
 - Medical Expenditure Panel Survey (MEPS) Linkage Files
 - NHIS survey cohorts are linked by person identification number

OAE Record Linkage Activities

- Mortality
 - National Death Index
- Retirement and Disability
 - Social Security data from the Retirement, Survivors, Disability Insurance (RSDI) and Supplemental Security Income (SSI) programs
- Medicare enrollment and payments
 - Enrollment and claims data

Summary Linked Mortality Data Files

	Current Linkage (death data through 2000/2002)		Next Linkage (death data through 2006)	
	Restricted- use	Public-use	Restricted-use	Public-use
NHIS 1986-2000	X	X	X	X
NHIS 2001-2004			X	X
LSOA II (1994-2000)	X	X	X	X
NHEFS (1971-1992)	X		X	X
NHANES II (1976-1980)	X		X	X
NHANES III (1988-1994)	X	X	X	X
NHANES 1999-2004			X	X
NNHS 1985	X		X	
NNHS 1995, 1997, 2004			X	

Research Potential of Linked Mortality Data

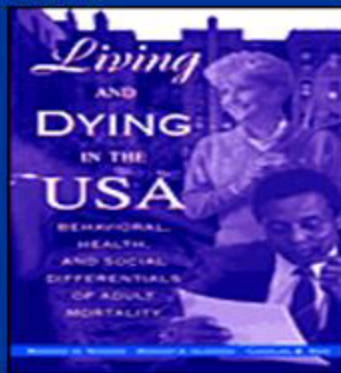


Excess Deaths Associated with Underweight, Overweight, and Obesity

KM Flegal, BI Graubard, DF Williamson, MH Gail;
JAMA, 2005;293:1861-1867.

Living and Dying in the USA:

Behavioral, Health, and Social
Differentials of Adult Mortality
RG Rogers, CB Nam, RA Hummer;
2000.



Suicide among male veterans: a prospective population-based study

MS Kaplan, N Huguet, BH McFarland, JT Newsom;
J Epidemiol Community Health, 2007; 61:619-624.

Linked Medicare Files

- Medicare enrollment and claims data for the years 1991-2000
 - Denominator file
 - MEDPAR Inpatient hospitalization
 - MEDPAR Skilled nursing facility (SNF)
 - Hospital outpatient
 - Home Health Agency (HHA)
 - Hospice
 - Carrier (physician/supplier Part B file)
 - Durable Medical Equipment (DMERC)

Research Potential of Linked Medicare Data

- Examine risk factors for health conditions
- Examine reliability of survey data
 - Compare survey reported Medicare enrollments to Medicare claims records
 - Examine survey report of disability with program participation eligibility criteria
- Examine disparities in Medicare service utilization

Research Potential of Linked Medicare Data

- Examine risk factors for health conditions
- Examine reliability of survey data
 - Compare survey reported Medicare enrollment to Medicare claims records
 - Examine survey report of disability with program participation eligibility criteria
- Examine disparities in Medicare service utilization

Publications & Current Projects

Using Linked Medicare Data

■ Publications:

- Looker AC. Mussolino ME. Serum 25-hydroxyvitamin D and hip fracture risk in older U.S. white adults. *Journal of Bone & Mineral Research*. 23(1): 143-150, 2008 Jan.

■ Current Projects:

- Assessing the Economic Burden of Chronic Kidney Disease in the United States
- The Association of Obesity and Overweight with Higher Medical Care Costs in Medicare Beneficiaries
- Comparing Self-Reported Chronic Conditions with Medicare Claims Data

Linked Social Security Files

- Retirement, Survivor, & Disability Income
 - Master Beneficiary Record (MBR), 1962-2003
 - Program eligibility, benefit amount, payment status, dual entitlement
 - Payment History Update System (PHUS), 1984-2003
 - Benefit payment amounts, including withholding information for Medicare Part B premiums
- Supplemental Security Income
 - Supplement Security Record (SSR), 1974 to 2003
 - Program eligibility, benefit information, and payment status

Social Security Linkage

	Current Linkage SSA data 1962-2003	Next Linkage SSA data 1962-2007
NHIS 1994-1998	X	X
NHIS 1999-2005		X
LSOA II (1994-2000)	X	X
NHEFS (1971-1992)	X	X
NHANES III (1988-1994)	X	X
NHANES 1999-2004		X
NNHS 1985	X	X
NNHS 1995, 1997, 2004		X

Research Potential of Linked Social Security Data

- Examine reliability of survey information for SSA program participation and benefits
- Compare the health characteristics of early retirees (age 62) to those who postpone benefits
- Policy and analysis using validated survey data
 - Predicting the number of people who will become disabled based upon survey reported health conditions
 - Determining whether current disability entitlement funding levels will be adequate as the population ages

Publications & Current Projects Using Linked Social Security Data

■ Publications:

- Riley GF. Health Insurance and Access to Care among Social Security Disability Insurance Beneficiaries during the Medicare Waiting Period. *Inquiry* 43:222-230, 2006 Fall.

■ Current Project:

- The Importance of Objective Health Measures in Predicting Exit From the Labor Force Via Early OA, DI, and SSI Programs
- Where Are They Now? The Subsequent Labor-Force Participation and Health Status of Rejected Disability Applicants
- Concordance between Self-Reports of SSDI Application & Receipt

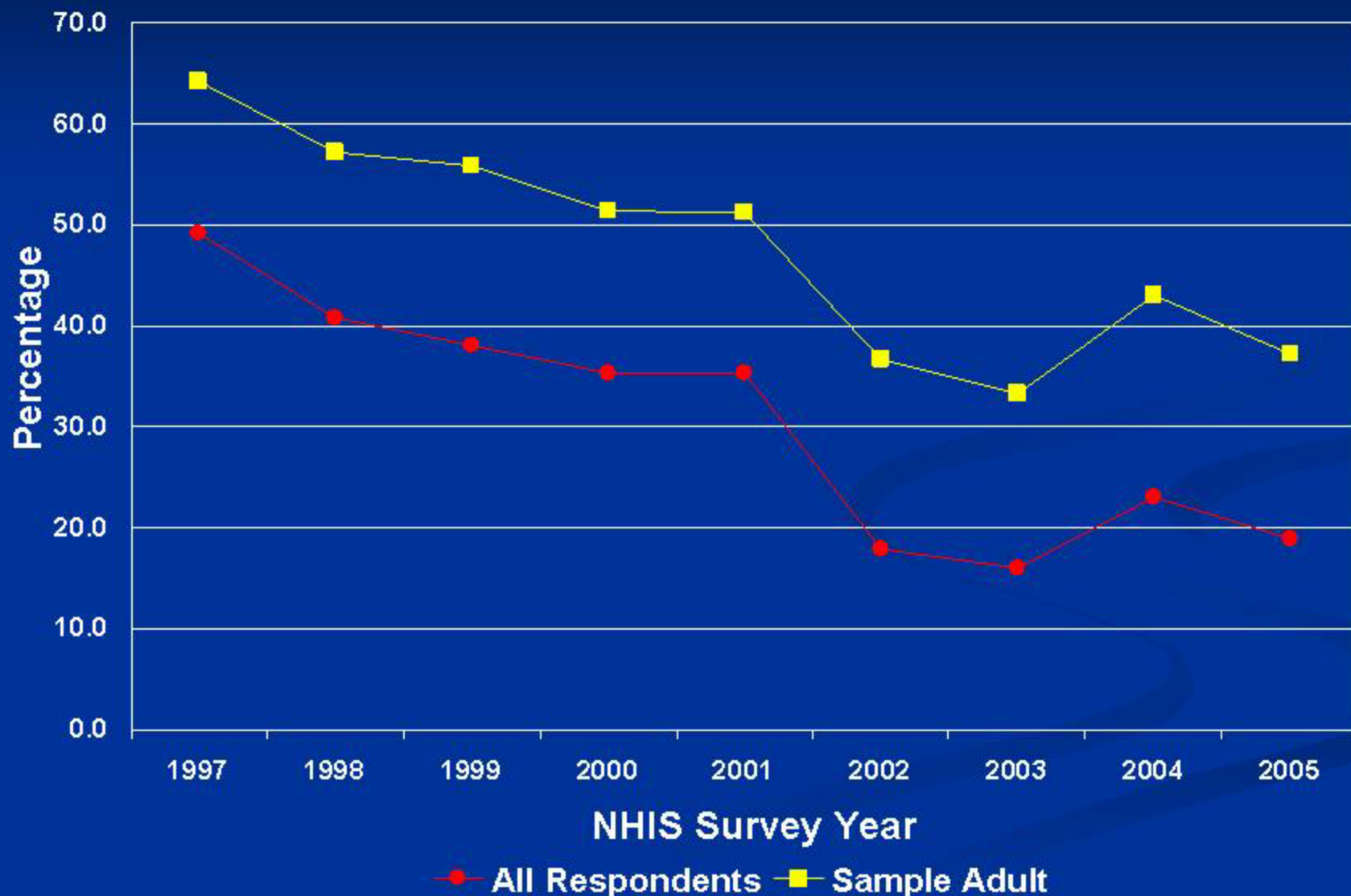
Challenges

- Obtaining informed consent
- Improving identification data
- Developing interagency agreements
- Balancing limited resources
- Improving data access

Informed Consent

- Satisfying institutional requirements for permission to link
- Communicating the importance of record linkage to survey respondents and gaining their cooperation

NHIS Participants Providing SSN



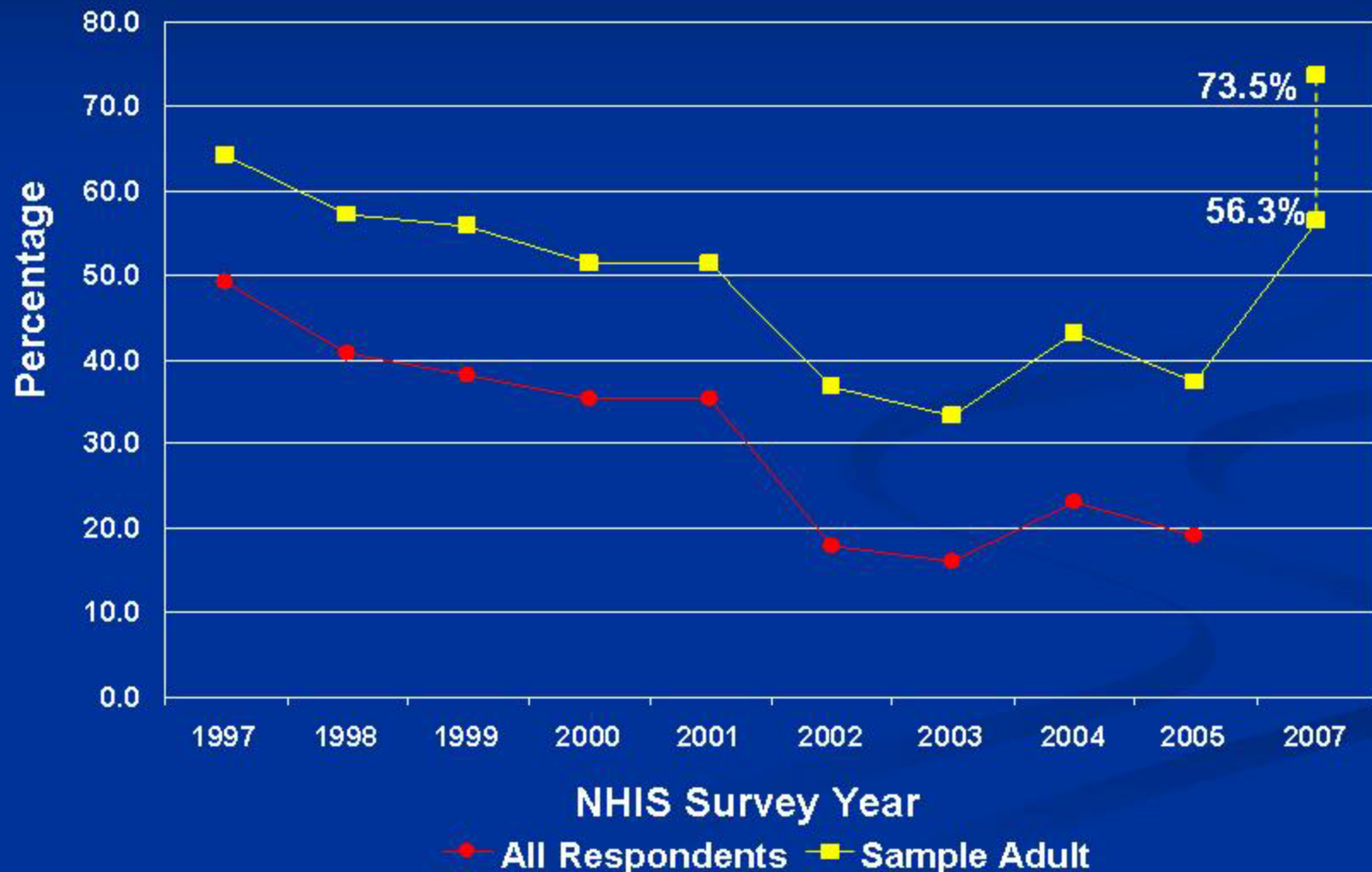
2007 NHIS split-ballot experiment

- Tested two options for obtaining consent to record linkage
 - Treatment 1: Ask permission to link survey data with health-related records of other government agencies
 - If no, end
 - If yes, ask for last four digits of SSN
 - Treatment 2: Ask last four digits of SSN; consent to link embedded in the question
 - If partial SSN reported, end
 - If partial SSN not reported, as permission to link without it

NHIS Split-Ballot Results

- Treatment 2 was associated with significantly higher odds of consent overall, and consent with or without an SSN, compared to treatment 1
- Treatment 2 substantially increased the percentage of sample adults consenting to record linkage compared to 2005-06

NHIS Participants Providing SSN



Next Steps

- Design and test matching algorithms that utilize last four digits of SSN
- Work with other agencies who currently require all 9 digits of SSN for linkage

Identification Data

- Incomplete or inaccurate identification data from the survey interview can lead to potential biases in linked data files
 - Names
 - Addresses

Names

- Issues with collection and cleaning of survey participant names
 - Created standardized procedure to identify non-names in survey data
 - Developed nickname to proper name conversion table

Addresses

- Important to keep address information current
 - Improve linkage accuracy
 - Particularly important for common surnames
- Conduct passive tracking
 - National change of address matches
 - Standardize and update addresses
 - Address updates current as of June 2007
 - Now geo-coding new address data

Interagency Agreements

- Complexity in drafting agreements
- Agency differences in legislative mandates and requirements to protect data and survey respondent confidentiality
- Resolving issues of data ownership and access to linked data files, e.g.
 - Can a public-use file be created?
 - If restricted access required, where will data reside?
Who will control access?

Interagency Agreements

- NCHS taking leadership role in working across federal agencies through Federal Committee on Statistical Methodology (FCSM)
 - 2006: NCHS & Census presentation highlighting difficulties in developing agreements
 - Included a recommendation that FCSM facilitate the development of a IAA template
 - 2008: FCSM convenes sub-committee on usage of administrative records
 - Defining best practices across federal agencies
 - Developing an interagency agreement template for record linkage projects

Balancing Limited Resources

- Limited resources to both assist data users and conduct new linkages
 - Developing user documentation, web pages, analytic guideline & other user tools
 - Transforming administrative data into analytic data
 - Providing technical assistance to data users

Data User Tools

- File layouts & detailed notes
- Sample SAS & STATA input statements for public-use linked mortality files
- Dummy data
- Matching methodology reports
- Linkage rates for SSA & CMS linked data
- Analytic guidelines
 - Weighting and variance estimation (NHIS)

Data User Tools (cont.)

- Summary Medicare and SSA files
- Feasibility data files for SSA & CMS Files –
Download from web
- Comparative analysis of the public-use and restricted-use linked mortality data
- Compare the mortality experience of the NHIS linked mortality cohorts to U.S. population

Data Access

- Expand data access for restricted files
 - Expand RDC locations (e.g. NCHS restricted data now accessible through 9 Census RDCs)
 - Create designated agents
- Create public use file, possible but difficult
 - Assess disclosure risk
 - Develop synthetic public-use micro data files that are analytically useful and valid

Public-use Linked Mortality Files

- Risk of survey participant re-identification required restricted-use designation for the most recent mortality update
- Great degree of motivation to create public-use linked mortality files to enhance utilization
 - Developed perturbation strategy to ensure protection of respondent identity, while maintaining statistical validity
 - Conducted comparative study to determine whether findings based upon the public-use files reproduce those using the restricted-use files.

Selected Variables on the NHIS Linked Mortality Files

	Restricted-use	Public-use
Final mortality status	Yes	Yes
Death date	Yes (month, day, year)	Yes (quarter, year)
Underlying cause-of-death	Yes	Yes (grouped recode)
Multiple cause-of-death	Yes	Yes (selected conditions)
Age at interview	Yes	Yes (top-coded at 85)
Age at death	Yes	No
Age last presumed alive	Yes	No
Date of birth	Yes (month, day, year)	Yes (month, year)
Interview date	Yes (month, day, year)	Yes (quarter, year)

Comparative Analyses Results

- The public-use linked mortality files (NHIS, NHANES III, and LSOA II), with a limited amount of perturbed data and reduced number of mortality variables, yield similar results as the restricted-use data
- Findings from the NHIS comparative analyses forthcoming in the *American Journal of Epidemiology*

Outreach

- Announcements
 - Listserves
 - HCFO, AMSTAT News
- Presentations
 - Conferences: APHA, SER, PAA, JSM, NCHS DUC
 - Committees: CNSTAT, NCVHS
 - Agencies: CBO, CMS
- Workshops
 - Academy Health, SER, ASHE

Future Linkage Activities

- Periodic Mortality Linkage
 - Plan to conduct mortality linkage every three years
 - Develop both restricted and public-use versions
- Continue Medicare and Social Security Linkage
 - Periodicity of these linkages is outside NCHS control

Future Linkage Activities

■ Medicaid

- Multi-agency collaborative project exploring differences in survey and program estimates of Medicaid enrollment
 - 2001-2002 NHIS linked to MSIS data files
 - Comparison of NHIS survey estimates of Medicaid enrollment with counts from CPS
 - Expansion of partnership planned to include collection of linked Medicaid data for NHIS and NHANES

■ State based linkages

- State Cancer Registries
 - National Program of Cancer Registries (NPCR)
 - SEER registries