

Data Linkage Activities at NCHS

Jennifer Madans, Ph.D.
NCHS Associate Director of Science

Christine S. Cox, M.A.
NCHS Special Assistant for Record Linkage

September 18, 2006



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

SAFER • HEALTHIER • PEOPLE™



Purpose

- Overview of NCHS Data Linkage
 - Past, Present, and Future
- Current linked data access procedures
- Challenges in conducting record linkages
- Future directions

Why do record linkage?

- Increases accuracy and detail of data collected
- 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, cost for re-contacting survey respondents for follow-up information can be expensive

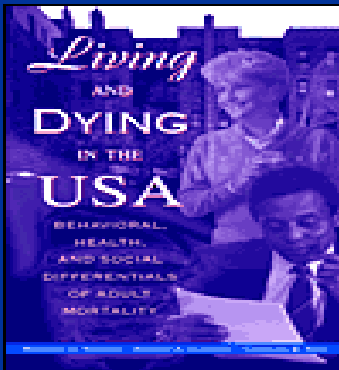
Research Potential of Linked Data



The Income-Associated Burden of Disease in the United States
P Muennig, P Franks, H Jia, E Lubetkin and MR Gold

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



A Semiparametric Analysis of the Body Mass Index's Relationship to Mortality
JT Gronniger

Types of Record Linkage

- Person-level or facility-level
 - Person survey data linked with administrative data (e.g. Medicare)
 - Hospital data linked to 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

NCHS Data Record Linkage Program

- Early (1980's)
 - NMCUES- link to Medicare - MADRS1981-86
 - NHEFS- links to NDI in 1982-1992 and Medicare in 1980-1986 MADRS
- 1990's- NHIS mortality linkage to NDI
- 2000 to present- expanded linkage program

Completed Survey Data Linkage as of 2006

	Mortality (NDI)	Medicare (CMS)	Retirement & Disability (SSA)
NHIS 1986-2000	X		
NHIS 1994-1998	X	X	X
LSOA II	X	X	X
NHANES I	X	X	X
NHANES II	X	X	
NHANES III	X	X	X
NNHS 1985	X		X

Other Completed NCHS Data Linkage

- Linked Birth and Infant Death Files
 - 1983-1991 (birth cohort linkage)
 - 1995-2003 (period and birth cohort linkage)
- National Hospital Discharge Survey
 - American Hospital Associations Annual Survey of Hospitals

Ongoing Linkage

- National Death Index (NDI)
 - NHIS 1986-current
 - NHANES I,II,III, 1999-current
 - National Nursing Home Surveys
- CMS Medicare Enrollment & Utilization
 - NHIS 1994-current
 - NHANES I,II,III, 1999-current
 - 2004+ National Nursing Home Surveys
- Linked Birth and Infant Death Files

Planned Linkage

- 2001-2002 NHIS & NHANES
 - CMS Medicaid Enrollment data
- 2004 NNHS
 - CMS Minimum Data Set
- NHANES
 - Food Assistance programs

Linkage Evaluation Studies

- Methodologic reports
- Bias analysis
 - Non-response assessments
- Evaluate and improving matching algorithms

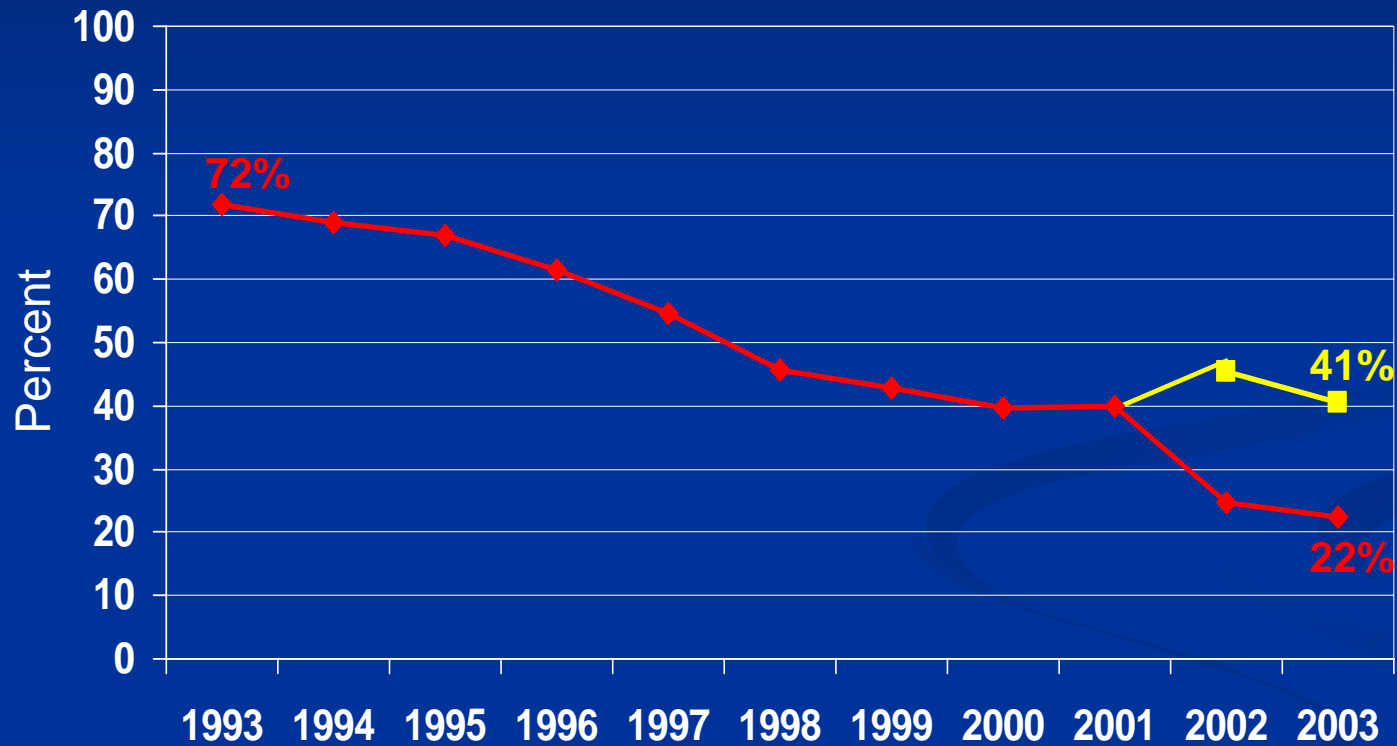
Current Linked File Access Procedures

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

Challenges: Human Subjects/Privacy Issues

- Obtaining informed consent
 - Satisfying NCHS institutional requirements for permission to link
 - Satisfying provider's institutional requirements for permission to link
 - Communicating the importance of linkage and gaining the cooperation of survey respondents

NHIS adults providing SSN



Challenges: Institutional Requirements

- Complexity in drafting agreements
- Differences in privacy requirements
 - Legislative mandates between collaborating agencies

Challenges: Balancing Resources

- Development of user documentation, web pages, analytic guidelines & workshops
- Need to transform administrative data into analytic data
- Experts in the data reside in the providing institution but are not available to assist data users
- Difficult to balance limited resources- balancing the effort to assist data users with resources to conduct new linkages

Challenges: Improving Data Access

- Creating public use files from linked data very difficult
 - Must shield identity of survey respondents from the agency that provided the administrative records
- Using sworn agent status
- Perturbing linked data for public release?
- Share knowledge and experience across agencies

Standardization across Federal Agencies

- Increases efficiency and reduces cost
- Reduce need to develop user documentation
- Seek standards for:
 - Linkage procedures
 - Data handling requirements
 - Administrative record extracts & documentation

Future Directions

- Increase collaboration & communication among agencies
- Develop new linkage projects
- Develop standards for cross-agency linkage
- Improve survey techniques for obtaining informed consent and adequate identification data
- Data access
 - Expand access at RDC's
 - Develop new methodologies to:
 - Assess disclosure risk
 - Develop public-use micro data files