Challenges and Opportunities with NCHS Linked Data Files
Council of Professional Associations on Federal Statistics (COPAFS)

• Provides government policy decision makers with information that demonstrates the value of federal statistics and the imperative of a strong federal statistical system.

• Organizes forums in which federal statisticians and agency representatives interact with the users of their statistics.

• Conducts independent, third-party evaluations of statistical tools, processes and accessibility options, and prepares white papers.

• Holds seminars, workshops, and webinars on challenges and opportunities facing groups that rely on federal statistics and those facing the federal agencies that produce them.
National Center for Health Statistics (NCHS)

• Monitor the nation’s health by collecting, analyzing and disseminating health data.

• A variety of data collection mechanisms are used to obtain accurate information from multiple sources, including individual interviews, physical examinations, medical records, and birth and death records.

• NCHS developed a data linkage program to maximize the scientific value of the Center's population-based surveys.
Challenges and Opportunities with NCHS Linked Data Files

Jennifer D. Parker, Ph.D.
National Center for Health Statistics/CDC
Webinar March 20, 2014
Outline

• NCHS Record Linkage program
• Administrative Data
• Protection of Personally Identifying Information and Linked Data Files
  • Research Data Center
  • Linkage eligibility
• Examples highlighting data uses and analytic issues
• Summary
NCHS Record Linkage Program

- **Population health surveys** linked to administrative records
  - Enhance analytic potential
  - Longitudinal
  - Reduce respondent burden

- **NCHS surveys add**: more detailed demographic information and family structure, socio-economic and educational factors, underlying health characteristics and health behaviors

- Surveys include:
  - National Health Interview Survey
  - National Health and Nutrition Examination Surveys
  - National Health Care Surveys
Administrative Data Linked to NCHS Surveys

- **Administrative data add**: cross-sectional and longitudinal program eligibility, enrollment and participation, services billed
  - Surveys and years of administrative linked vary

- **Linked Data Available**
  - Centers for Medicare and Medicaid Services (through 2009)
  - Social Security Administration
  - National Death Index (through 2011)

- **Pilot projects**
  - Florida Cancer Data System, Supplemental Food and Nutrition Program (SNAP), Housing and Urban Development
Protection of personally identifying information (PII) and linked data files

- Laws and regulations
  - Section 308(d), Public Health Service Act
  - Privacy Act
  - Confidential Information Protection and Statistical Efficiency Act (P.L. 107-347)
  - Partner agencies may have additional regulations and policies

- Record linkage projects approved by NCHS Ethics Review Board (ERB)

- Most linked data files are restricted use and only available through the **NCHS Research Data Center**
  - On-site at NCHS RDC in Hyattsville, Atlanta, or D.C. (federal agencies)
  - On-site at Census RDC

- Feasibility files for CMS and SSA linked data available online

- Public-use Linked Mortality Files available online
Research Data Center

NCHS Research Data Center (RDC)

The National Center for Health Statistics (NCHS) developed the Research Data Centers (RDC) to allow researchers access to restricted data. Today, in addition to providing access to NCHS data, the RDC also hosts restricted data from a variety of groups within the Department of Health and Human Services (DHHS).

The RDC is responsible for protecting the confidentiality of survey respondents, study subjects, or institutions from which data were collected. In order to do this, we request all researchers submit a research proposal outlining the need for this more sensitive data. The proposal provides a framework for us to identify potential disclosure risk. Once approved, we work with you to create a data file specific to your research question. We cannot send you the dataset, but we offer several modes of access.

People interested in using the RDC are typically, university trained researchers who are comfortable using SAS or a similar statistical package to analyze standard datasets. Past researchers have been epidemiologists, economists, demographers and health scientists.

If this is your first time using the RDC we have provided step-by-step instructions to guide you along the process.

STEP 1 Restricted Data

STEP 2 Access Modes
Linkage eligibility

• For linkages other than to the NDI, survey respondents must
  • provide **sufficient PII** (SSN, name, date of birth),
  • not explicitly refuse linkage
  • or, explicitly allow linkage w/o SSN

• Percent linkage eligible **varies among surveys**, over time, by respondent characteristic

• Criteria and ascertainment have evolved over time

• Currently in the NHIS, **only primary respondents are asked for SSN**. Other household members included in the survey are not linkage eligible

• Child survey respondents are not linkage eligible for administrative data obtained after their 18th birthday per NCHS ERB guidelines

• Most survey respondents eligible for linkage to the NDI
  • Eligibility based on data availability, not consent to link
Percent of NHIS sample adults refusing to provide SSN, by age at interview and survey year

Analytic considerations

• Linkage eligibility
  • Current recommendation is to re-weight sample weights but other approaches are being considered

• Temporal alignment of survey and administrative data
  • Child survey respondents
  • Program eligibility
  • Discontinuous coverage
  • Residential mobility

• Program and data characteristics
  • Medicaid and Medicare managed care
  • State-based program (e.g. Medicaid, Florida Cancer Data System)
Example 1: Medicaid ascertainment by interview and administrative data linkage. Children, 1999-2004 NHANES-MAX linked data

Example 1: Medicaid ascertainment by interview and administrative data linkage. 1999-2004 NHANES-Medicaid linked data

Table: Child characteristics by concordance

<table>
<thead>
<tr>
<th></th>
<th>Agreement: Yes in NHANES</th>
<th>Discordant: Yes in NHANES</th>
<th>Agreement: No in NHANES</th>
<th>Discordant: No in NHANES</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes Linked to MAX</td>
<td>Not Linked to MAX</td>
<td>Not Linked to MAX</td>
<td>Yes Linked to MAX</td>
</tr>
<tr>
<td>Gender: Male, %</td>
<td>51.6</td>
<td>52.1</td>
<td>50.4</td>
<td>55.8</td>
</tr>
<tr>
<td>Age less than 6 years%</td>
<td>42.3</td>
<td>39.2</td>
<td>28.9</td>
<td>55.8</td>
</tr>
<tr>
<td>Race/ Ethnicity, %</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>non-Hispanic white</td>
<td>58.5</td>
<td>56.6</td>
<td>83.3</td>
<td>56.2</td>
</tr>
<tr>
<td>non-Hispanic black</td>
<td>26.4</td>
<td>22.2</td>
<td>8.0</td>
<td>32.5</td>
</tr>
<tr>
<td>Mexican American</td>
<td>15.1</td>
<td>21.2</td>
<td>8.7</td>
<td>11.3</td>
</tr>
<tr>
<td>Poverty Index, mean/SE</td>
<td>1.0 (0.03)</td>
<td>1.2 (0.06)</td>
<td>3.2 (0.05)</td>
<td>1.2 (0.05)</td>
</tr>
</tbody>
</table>
Example 2: How many children are enrolled in Medicaid any time during a 5 year period?

Figure. Percent of children enrolled in Medicaid at any time during 2004–2008, by number of years of enrollment: children ages 0-13 at time of interview, 2004 NHIS linked to 2004-2008 MAX

SOURCE: Simon et al, Pediatrics 2013
Example 3: What are the characteristics of children with longer vs. shorter gaps in Medicaid?

Figure: Number of enrollment periods in Medicaid during 10 year observation period: children ages 5-13 at time of interview, 2004 NHIS linked to 1999-2008 MAX

- Once, 44.3% (42.1—46.5%)
- Twice, 28.2% (26.2—30.3%)
- Three times, 15.3% (13.8—17.0%)
- 4 or more times, 12.2% (10.9—13.7%)

SOURCE: Simon et al, in press AJPH
Example 4: Obesity and Medicaid costs

Using the MAX-NHANES Merged Data to Evaluate the Association of Obesity and Medicaid Costs

Allison Hedley Dodd and Philip M. Gleason

The Medicaid Analytic eXtract (MAX) data set is derived from the state reporting of Medicaid eligibility and claims data and is designed to enable research on Medicaid enrollment, service utilization, and expenditures per calendar year at the enrollee level. The National Health and Nutrition Examination Survey (NHANES), a nationally representative survey of the U.S. noninstitutionalized population, collects measured heights and weights. The NHANES data set is merged with state-level Medicaid claims data to create the MAX-NHANES merged data set.

About This Series

The MAX Medicaid policy issue brief series highlights the essential role MAX data can play in analyzing the Medicaid program. MAX is a set of annual, person-level data files on Medicaid eligibility, service utilization, and payments that are derived from state reporting of Medicaid.
Example 4: Obesity and Medicaid costs

<table>
<thead>
<tr>
<th>Analysis Population</th>
<th>Number of Records</th>
</tr>
</thead>
<tbody>
<tr>
<td>1999–2004 NHANES participants</td>
<td>31,126</td>
</tr>
<tr>
<td>1999–2004 linkage-eligible NHANES participants</td>
<td>25,750</td>
</tr>
<tr>
<td>Total records in which NHANES and MAX data matched in the same year</td>
<td>7,915</td>
</tr>
</tbody>
</table>

Number of records after imposing each NHANES restriction

<table>
<thead>
<tr>
<th></th>
<th>Number of Records</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age 20 or older</td>
<td>2,180</td>
</tr>
<tr>
<td>Not pregnant</td>
<td>1,909</td>
</tr>
<tr>
<td>Has BMI</td>
<td>1,666</td>
</tr>
</tbody>
</table>

Number of records after imposing each MAX restriction

<table>
<thead>
<tr>
<th></th>
<th>Number of Records</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not SCHIP only</td>
<td>1,663</td>
</tr>
<tr>
<td>Full-benefit</td>
<td>1,473</td>
</tr>
<tr>
<td>Not dual</td>
<td>894</td>
</tr>
<tr>
<td>Adult or disabled</td>
<td>856</td>
</tr>
<tr>
<td>Never enrolled in comprehensive managed care</td>
<td>455</td>
</tr>
<tr>
<td>One MAX record per NHANES participant</td>
<td>446</td>
</tr>
<tr>
<td>Not underweight</td>
<td>433</td>
</tr>
<tr>
<td>Not in “Other” race/ethnic group</td>
<td>419</td>
</tr>
<tr>
<td>Not in state with 5 or fewer records</td>
<td>375</td>
</tr>
</tbody>
</table>


Note: Data restrictions accumulate from the top of the table to the bottom. The number of records shown in a row includes all data restrictions shown on the row and above.
Example 4: Obesity and Medicaid costs

Table 3. Estimated Annual FFS Costs for Adults by Obesity Status in 2004 dollars

<table>
<thead>
<tr>
<th></th>
<th>Not Obese</th>
<th>Obese</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Person in Data Set¹</td>
<td>$95.02</td>
<td>$225.72</td>
</tr>
<tr>
<td>30 year-old white female non-smoker with no high school degree who resided in:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>State A²</td>
<td>32.80</td>
<td>78.81</td>
</tr>
<tr>
<td>State B²</td>
<td>247.45</td>
<td>585.65</td>
</tr>
<tr>
<td>State C²</td>
<td>568.22</td>
<td>1,343.06</td>
</tr>
<tr>
<td>40 year-old black male smoker with a high school degree who resided in:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>State A²</td>
<td>3.97</td>
<td>10.73</td>
</tr>
<tr>
<td>State B²</td>
<td>35.52</td>
<td>85.24</td>
</tr>
<tr>
<td>State C²</td>
<td>82.67</td>
<td>196.58</td>
</tr>
</tbody>
</table>

Example 5: Enrollment in Medicare Advantage at age 65

**Figure.** Adjusted percent enrollment in Medicare Advantage in 2006-2009 and 2000-2005 by number of self-reported health conditions. 1997-2005 NHIS participants aged 60-64 years with 6 months of Medicare enrollment data (2006-2009) upon turning 65.

Example 5: Florida Cancer Linkage


  - Linking prior years of survey data to FCDS for examining risk factors, SES, screening history, access to care, and comorbidity information as predictors of cancer, cancer stage, cancer treatment

  - Linking subsequent years of survey data to FCDS for examining quality of life, demographic information, health, insurance, access to care after diagnosis and treatment
Example 5: Florida Cancer Linkage

- NHIS participants linked to FCDS=8,110
- Florida NHIS participants linked to FCDS=6,281
- non-FL NHIS participants who moved to FL after survey interview and linked to FCDS=1,829
- FL NHIS participants who left FL after survey interview and were not linked
- non-FL NHIS participants who moved to FL after interview and were not linked
Example 5: Florida Cancer Linkage

Figure 1. Percent of cancers diagnosed at a late stage (regional plus distant) by education level, overall and by sex, among Florida National Health Interview Survey participants aged 18 years and over linked with the Florida Cancer Data System.

*Relative standard error >30%.

SOURCE: CDC/NCHS, 1986-2009 National Health Interview Survey linked to Florida Cancer Data System data.
Example 6. Excess deaths associated with underweight, overweight and obesity [Reference Range BMI 18.5 – 24.9]

Flegal et al *JAMA* 293:1861, 2005
Example 7. Mortality associated with health insurance.

Table. Proportional Hazards Estimates of the Lack of Health Insurance on Survival Time controlling for baseline covariates. 2006 NHIS-Linked Mortality Files

<table>
<thead>
<tr>
<th>Covariates</th>
<th>Hazard ratio</th>
<th>Confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, gender</td>
<td>1.71</td>
<td>(1.65, 1.76)</td>
</tr>
<tr>
<td>Race, ethnicity, immigration status, veteran status</td>
<td>1.65</td>
<td>(1.59, 1.70)</td>
</tr>
<tr>
<td>Education</td>
<td>1.46</td>
<td>(1.41, 1.51)</td>
</tr>
<tr>
<td>Marital status</td>
<td>1.37</td>
<td>(1.33, 1.42)</td>
</tr>
<tr>
<td>Family income, telephone in home, mobile home</td>
<td>1.25</td>
<td>(1.20, 1.29)</td>
</tr>
<tr>
<td>Labor force participation</td>
<td>1.20</td>
<td>(1.15, 1.24)</td>
</tr>
<tr>
<td>Smoking and BMI</td>
<td>1.10</td>
<td>(1.03, 1.19)</td>
</tr>
<tr>
<td>Health status and activities</td>
<td>1.03</td>
<td>(0.95, 1.12)</td>
</tr>
</tbody>
</table>

Kronick, HSR 2009
Summary

• **Disclosure protection** of survey participants’ information critical

• Access linked data files through the NCHS Research Data Center

• Analysis of linked survey-admin data requires assessment and **consideration** of many issues, including:
  • re-weighting for linkage eligibility
  • temporal correspondence between survey and admin data
    • study design, time-varying covariates
  • geographic correspondence between survey and admin data
  • program eligibility and program characteristics (e.g. Medicare MA)

• Linkage between NCHS surveys and administrative data provide **opportunities** to study characteristics of program beneficiaries, patterns of use and care
  • Study areas include: smoking, obesity, cancer, managed care, Medicaid and many others