Record and Geographic Linkages to Inform Health Disparities

Jennifer Parker and Lauren Rossen
Office of Analysis and Epidemiology
Outline

• Why combine data?
• Linked birth and infant death files
• NCHS record linkage program
  – Linked mortality files
  – Linked Medicaid and Medicare data
• Geographically linked data
• Access to NCHS linked data
• Summary
Why link data for disparities?

- Possibly better ascertainment of race and ethnicity information
- Additional information on
  - socioeconomic related variables
  - contextual exposures
  - urban-rural status
  - outcomes or intermediate endpoints related to outcomes
Infant Mortality Rates by Race
United States, 1915-2005

Black : White Infant Mortality Ratio
United States: 1915-2005

Year


Infant deaths per 1,000 live births

Ratio of infant mortality rates

0 0.5 1 1.5 2 2.5 3


Year
Why link infant death records to infant birth records?

Birth records add:

– Race of mother versus race of infant
  • Better information for smaller race and ethnicity groups
– Mother’s demographic information (e.g. education, marital status)
– Infant and maternal health information (e.g. birth weight, gestational diabetes)
2005 Infant Mortality Rates by Mother’s Race: Unlinked and Linked Births

<table>
<thead>
<tr>
<th>Mother’s Race</th>
<th>Unlinked</th>
<th>Linked</th>
</tr>
</thead>
<tbody>
<tr>
<td>American Indian or Alaska Native</td>
<td>7.97</td>
<td>8.06</td>
</tr>
<tr>
<td>Asian or Pacific Islander</td>
<td>3.78</td>
<td>4.89</td>
</tr>
<tr>
<td>Black or African American</td>
<td>13.73</td>
<td>13.26</td>
</tr>
<tr>
<td>White</td>
<td>5.73</td>
<td>5.73</td>
</tr>
</tbody>
</table>

SOURCE: CDC Wonder
Infant mortality by maternal education and race
United States, 2005

Deaths per 1000 live births

- White
- Black

Mother’s Education, years

- <12
- 12
- 13-15
- 16+
Pre-term (<37 weeks) Infant Mortality rates, 1989-2006

Deaths per 1000 live births

- American Indian
- Asian and Pacific Islander
- Central/South American
- Cuban
- Mexican
- non-Hispanic Black
- non-Hispanic White
- Other/Unknown Hispanic
- Puerto Rican

1989-90: 50.1, 29.7, 32.3, 25.3, 43.3, 36.0, 60.6, 29.1, 37.5
2005-06: 32.3, 27.3, 26.5, 28.1, 51.3, 47.3, 38.5, 27.1, 37.5
NCHS Record Linkages

• NCHS record linkage program links survey data to administrative records using confidential personal identifying information (e.g. names, Social Security Numbers, dates)

• Administrative records
  – Mortality
  – Centers for Medicare and Medicaid Services (CMS)
  – Social Security Administration
  – Pilot projects (e.g. Florida Cancer Data System, Texas Supplemental Food and Nutrition Program )
NCHS Record Linkages

Why link survey data to administrative records?

- **Survey data**
  - better detail on race, ethnicity, socioeconomic indices (education, income), baseline health status, self-reported program participation

- **Administrative records**
  - program participation (e.g. Medicaid, SNAP)
  - costs of medical care, benefits
  - longitudinal data adds ability examining health prior to or after the survey
  - mortality and cause of death
Life expectancy at age 25

<table>
<thead>
<tr>
<th>Year</th>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td>1996</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No high school diploma</td>
<td>47</td>
<td>53</td>
</tr>
<tr>
<td>High school graduate or GED</td>
<td>50</td>
<td>57</td>
</tr>
<tr>
<td>Some college</td>
<td>51</td>
<td>58</td>
</tr>
<tr>
<td>Bachelor’s degree or higher</td>
<td>54</td>
<td>59</td>
</tr>
<tr>
<td>2006</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No high school diploma</td>
<td>47</td>
<td>52</td>
</tr>
<tr>
<td>High school graduate or GED</td>
<td>51</td>
<td>57</td>
</tr>
<tr>
<td>Some college</td>
<td>52</td>
<td>58</td>
</tr>
<tr>
<td>Bachelor’s degree or higher</td>
<td>56</td>
<td>60</td>
</tr>
</tbody>
</table>

NOTE: GED is General Educational Development high school equivalency diploma.
SOURCE: CDC/NCHS, Health, United States, 2011, Figure 32. Data from the National Health Interview Survey Linked Mortality File.
Percent distribution of parental education by children’s enrollment in Medicaid during a 5 year period

2004 NHIS linked to 2004-2008 Medicaid

- less than high school
- high school diploma
- some college/associates degree
- college degree or greater

Never enrolled in Medicaid: 4.6% (less than high school), 14.8% (high school diploma), 31.3% (some college/associates degree), 49.3% (college degree or greater)

Enrolled in Medicaid during 1 year only: 12.4% (less than high school), 31.4% (high school diploma), 31.4% (some college/associates degree), 16.9% (college degree or greater)

Enrolled in Medicaid during all 5 years: 32.5% (less than high school), 33.6% (high school diploma), 28.2% (some college/associates degree), 5.7% (college degree or greater)

SOURCE: Simon et al, preliminary results
Percent hypertension by Medicare enrollment (fee for service or Medicare Advantage) by race/ethnicity

1999-2004 NHANES linked to 2007 Medicare

SOURCE: Mirel et al 2012
Association between ethnicity and mortality using 3 linkage criteria

Foreign born Hispanic versus US born white: 1.24
US born Hispanic versus US born white: 1.14
Foreign born white versus US born white: 0.97

Relaxed NCHS standard
Tightened

SOURCE: Lariscy 2012
Geographic linkages

Why link survey data to geographic data?

• Measures of
  – contextual SES indices (median income, % poverty)
  – exposures (e.g. pollution, liquor stores)
  – access (e.g. Community Health Centers)

• Very limited geographic detail on public use files
  – External data can be merged by administrative units or using GIS methods (Research Data Center)
Percent of Children (2-18 yr) in Poverty (Family and Tract level) and “Double Jeopardy”, NHANES 2001-2010 linked to 2000 Census

SOURCE: Rossen et al, preliminary results

- Reference group is non-Hispanic white
Distance to nearest road (log) by poverty status & race/ethnicity
NHANES 1999-2008 linked to 2005 National Highway Planning Network

Beta (95% CI)

Non-Hispanic white

Non-Hispanic black

Mexican American

<100% vs 100-200% vs 200-400% vs >400%
<100% vs 100-200% vs 400% vs >400%
<100% vs 200% vs 400% vs >400%

SOURCE: Berko et al unpublished results
Percent fair/poor health by distance to nearest road
NHANES 1999-2008 linked to 2005 National Highway Planning Network

SOURCE: Parker et al 2012
Considerations

• Temporal relationship between data sources
  – Prospective and retrospective analyses
  – Need to keep track of data collection years for each source

• Linkage bias
  – All survey records may not be linked
  – Linkage quality depends on accurate identifiers
  – All geographic areas in survey may not have contextual data

• Record linked and geographic linked data pose risks of disclosure
  – Access through the NCHS Research Data Center
Summary

- So, why link data for disparities?
  - Survey data
    - better detail on race, ethnicity, socioeconomic indices (education, income), baseline health status, self-reported program participation
  - Administrative records for individuals
    - program participation (Medicaid, SNAP)
    - costs of medical care, benefits
    - longitudinal data adds ability examining health prior to or after the survey
    - mortality and cause of death
  - Geographic data
    - contextual socioeconomic status
    - measures of exposure (pollution, locations of exposures)
    - measures of access (locations of services)

- However,
  - Most analyses of NCHS linked and geographic data must be done in the RDC
  - Care must taken to understand temporal and geographic relationships
Acknowledgements

• Nataliya Kravets
• Jeff Berko
• Ken Schoendorf
• Alan Simon
• Lisa Mirel
• Kim Lochner
More information

Jennifer D. Parker
(301) 458-4419, jdparker@cdc.gov
Disparities in Exposure to Air Pollution during Pregnancy

Tracey J. Woodruff, Jennifer D. Parker, Amy D. Kyle, and Kenneth C. Schoendorf

Relationship between high air pollution index and maternal race. US Natality 98-99
Infant Mortality Rates Among Term Infants (≥ 37 weeks), 1989-2006