Analytic Challenges with National Data Linked to State-Level Data

The National Health Interview Survey – Florida Cancer Data System Linkage

Eric A. Miller
National Center for Health Statistics (NCHS)
Pilot Project

• Linkage of the Florida Cancer Data System (FCDS) Data to National Health Interview Survey (NHIS)

• Objectives
  • Feasibility
  • Value / Utility
Data Linkage

  ~2 Million Records

  ~2.5 Million Records

8,217 linked survey participants
How Does This Apply to Other Linkages?

• Some of the issues we have encountered with this linkage could be relevant for other National/State linkages

• For example
  – State-level analysis of national survey data (e.g. NHIS) linked to Medicaid or Supplemental Nutrition Assistance Program (SNAP)
Description of the Data

• Cancer Registries
  – Collect, manage, and analyze data about cancer cases and cancer deaths
  – Are essential for monitoring progress in cancer prevention and control
Data Collected by Cancer Registries

- Cancer-related
  - Incident cancers
  - Type, extent (i.e. stage) and location of tumor
  - Date of diagnosis
  - Type of initial treatment

- Demographics

- Vital status
Description of the Data

• NHIS
  – In-person household survey
  – Conducted continuously by the CDC’s NCHS since 1957
  – Large sample sizes
    • ~35,000 households in the U.S. per year
    • Complex sampling with some populations oversampled
Data Collected by NHIS

- Risk factors (*e.g.* smoking, alcohol use, obesity)
- Health conditions, diseases, and disabilities
- Cancer screening history (*selected years*)
- Occupation/Industry
- Socioeconomic information (*e.g.* income, education, health insurance/access to care)
Data Linkage

- Linking the information from these two sources could potentially provide a valuable resource for cancer research.

- Linkage adds:
  - Longitudinal component to survey
  - Quality of life/health after diagnosis
  - Risk factor, SES, screening history, access to care, and comorbidity information to registry data
Also Adds Complexity

• NHIS is a nationally representative sample of the civilian, non-institutionalized (CNI) population of the United States – i.e. not just Florida

• FCDS is intended to capture (almost) all cancers diagnosed among Florida residents
Examples of Challenges

- Creation of survey weights
- Survey participant mobility
Challenge #1 – Survey Weights

• NHIS weights were available to represent the US CNI population

• Weights needed to be created to represent the Florida population
  – NCHS (Dean Judson) created weights to be representative of the Florida CNI population for each year of the survey
Creation of Florida Weights

• Used NHIS sample weights
  – Limited to Florida survey participants
  – Adjusted for linkage ineligibility using PROC WTADJUST in SUDAAN
    • Based on race, sex and age

  – Linkage ineligibility
    • Did not refuse
    • Did not provide sufficient personally identifiable information
Creation of Florida Weights

- Post-stratified to the Florida CNI population
  
  Method 1: Using Florida CNI estimates directly from NHIS

  Method 2: Using estimates of the CNI population based on average CNI percent of total Florida population

- Methods highly correlated \((r=0.99)\) and had little effect on estimates
Comparison

Percent of survey participants with a cancer record in the FCDS who ever smoked by race/ethnicity, and post-stratification method

<table>
<thead>
<tr>
<th></th>
<th>Post-stratification Method 1</th>
<th></th>
<th>Post-stratification Method 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ever Smoked % (SE)</td>
<td></td>
<td>Ever Smoked % (SE)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>48.2 (3.52)</td>
<td></td>
<td>48.0 (3.44)</td>
</tr>
<tr>
<td>White Non-Hispanic</td>
<td>63.9 (1.63)</td>
<td></td>
<td>64.1 (1.52)</td>
</tr>
<tr>
<td>Black Non-Hispanic</td>
<td>45.7 (3.33)</td>
<td></td>
<td>46.2 (3.23)</td>
</tr>
</tbody>
</table>
Consequence of Current Weighting Strategy

- Participants are weighted to the CNI Florida population in the year of their survey.

- This means if you were interviewed in Minnesota but diagnosed with cancer in Florida, you get a weight of 0.
  - Data for these respondents are not included in the analysis.
  - Not a trivial number.
Challenge #2 - Movers

• People moved to Florida after the survey
  – Some were diagnosed with cancer
    • In the FCDS
  – Some were not

• Analytic implications
  – With current weighting strategy loss of sample size limits the ability to look at individual cancer types or at demographic differences
Movers to Florida

- Number of Survey Participants Linked to FCDS=8,217
Movers to Florida

- Number of Survey Participants Linked to FCDS=8,217
- Number of FL survey participants linked to FCDS=6,366
Movers to Florida

- Number of Survey Participants Linked to FCDS=8,217
- Number of FL survey participants linked to FCDS=6,366
- Number who moved to FL after survey and were dx’d with cancer=1,851 (23%)
Reason for Current Strategy

Don - Florida Native

Walter from MN

Jack - MN Transplant

If Walter moves to Florida and is dx’d with cancer, he is more comparable to Jack who moved to Florida and did not get cancer.
Walter from MN

Don - Florida Native

Jack - MN Transplant

But we do not have a way to know about Jack in the data.
Comparison of Demographic Characteristics Among Participants Linked with FCDS by Residency at Survey

<table>
<thead>
<tr>
<th></th>
<th>Florida Residents %</th>
<th>Movers %</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Male</strong>*</td>
<td>49.6</td>
<td>52.7</td>
</tr>
<tr>
<td><strong>Race/Ethnicity</strong>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>83.6</td>
<td>91.8</td>
</tr>
<tr>
<td>Black</td>
<td>14.3</td>
<td>5.8</td>
</tr>
<tr>
<td>Hispanic</td>
<td>21.0</td>
<td>5.5</td>
</tr>
<tr>
<td><strong>Mean Age (as of 2009)</strong>*</td>
<td>61.3</td>
<td>56.7</td>
</tr>
<tr>
<td><strong>Education Level</strong>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; High School (HS)</td>
<td>7.0</td>
<td>14.0</td>
</tr>
<tr>
<td>HS Graduate</td>
<td>45.3</td>
<td>46.3</td>
</tr>
<tr>
<td>&gt; HS</td>
<td>47.7</td>
<td>39.7</td>
</tr>
</tbody>
</table>

*P<0.05 for difference between groups
Comparison of Health Characteristics Among Participants Linked with FCDS by Residency at Survey

<table>
<thead>
<tr>
<th>Smoking Status*</th>
<th>Florida Residents %</th>
<th>Movers %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Never</td>
<td>40.0</td>
<td>33.0</td>
</tr>
<tr>
<td>Current</td>
<td>23.5</td>
<td>27.7</td>
</tr>
<tr>
<td>Former</td>
<td>36.3</td>
<td>39.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Self-rated Health*</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Excellent / Very Good / Good</td>
<td>76.3</td>
<td>85.0</td>
</tr>
<tr>
<td>Fair / Poor</td>
<td>23.7</td>
<td>15.0</td>
</tr>
</tbody>
</table>

*P<0.05 for difference between groups
Comparison of Cancer Types Among Participants Linked with FCDS by Residency at Survey

<table>
<thead>
<tr>
<th>Cancer</th>
<th>Florida Residents</th>
<th>Movers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>%**</td>
<td>%**</td>
</tr>
<tr>
<td>Bladder</td>
<td>6.2</td>
<td>6.2</td>
</tr>
<tr>
<td>Breast (Female)*</td>
<td>19.0</td>
<td>15.4</td>
</tr>
<tr>
<td>Colorectal</td>
<td>12.8</td>
<td>12.4</td>
</tr>
<tr>
<td>Lung</td>
<td>13.8</td>
<td>13.4</td>
</tr>
<tr>
<td>Prostate</td>
<td>18.6</td>
<td>20.0</td>
</tr>
<tr>
<td>Thyroid</td>
<td>1.5</td>
<td>1.2</td>
</tr>
<tr>
<td>Uterus</td>
<td>3.6</td>
<td>2.7</td>
</tr>
</tbody>
</table>

*P<0.05 for difference between groups
**Percent of all cancer diagnoses. Cancer types are not mutually exclusive and table does not include all categories. Not intended to add up to 100%.
Alternate Weighting Strategy

• Statistical matching
  – For movers (in-migration) find a similar survey participant from Florida and split weight between Florida and non-Florida resident
    – Could limit to those diagnosed with cancer within a certain number of years (e.g. with 5 years of survey)
    – Could base magnitude of split on number of years between survey and diagnosis in Florida
      • e.g. 5 years: 90% FL /10% not FL, 1 year: 50%/50%
Challenge #2B: Movers Out of Florida

• People moved out of Florida after the survey
  – Some were diagnosed with cancer
  – Some were not
Movers Out of Florida

• Change of address data are available to see who moved from Florida
  – But we do not have a way to identify survey participants who were diagnosed with cancer in another state
    • This would require linkages with cancer registries nationally
Movers Out of Florida

• Analytic implications
  • Can affect the representativeness of the estimates if a sizeable number of participants moved out of state
  • Currently do not have a way to address movers out of state
Conclusions

• Linking national survey data to state-level data produces additional analytic considerations
  – And opportunities for further research

• When linking national and state-level data, it is important to consider the potential impact of “movers”
Co-Authors/Collaborators

• NCHS
  – Donna Miller
  – Dean Judson
  – Hannah Day
  – Yulei He
  – Jennifer Parker

• FCDS/University of Miami
  – Cristina Fernandez
  – Monique Hernandez
  – Jill MacKinnon
  – Laura McClure
  – Brad Wohler
  – Recinda Sherman
  – Bill LeBlanc
  – Sharon Christ
  – David Lee
For More Information

Eric Miller: bwe6@cdc.gov

Data Linkage at NCHS:

datalinkage@cdc.gov

www.cdc.gov/nchs/data_access/data_linkage_activities.htm

Thanks!