Race/Ethnic Data Collection: Population Surveys and Administrative Records

Jennifer Madans, PhD
National Center for Health Statistics
Centers for Disease Control and Prevention
NCHS Data Systems

- NCHS data systems are designed to collect data on racial and ethnic populations
- These data systems also provide for comparisons across time, providers, and geographic areas
- Ongoing data collections allow for combining data across years
Challenge for Population Surveys

• Obtain stable estimates for population subgroups defined by race/ethnicity AND geography AND other demographic characteristics
  • Conceptually straightforward
  • Technically feasible
  • Major cost and logistical issues
## Minimum Sample Size for Proportions

Design Effect = 1.5

<table>
<thead>
<tr>
<th>Proportion</th>
<th>RSE 30</th>
<th>RSE 20</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>17</td>
<td>38</td>
</tr>
<tr>
<td>40 or 60</td>
<td>25</td>
<td>56</td>
</tr>
<tr>
<td>30 or 70</td>
<td>39</td>
<td>88</td>
</tr>
<tr>
<td>20 or 80</td>
<td>67</td>
<td>150</td>
</tr>
<tr>
<td>10 or 90</td>
<td>150*</td>
<td>338</td>
</tr>
<tr>
<td>5 or 95</td>
<td>317</td>
<td>713</td>
</tr>
<tr>
<td>2 or 98</td>
<td>817</td>
<td>1838</td>
</tr>
</tbody>
</table>

* For a 10% statistic the CI is 4-16%
Calculation of Number of Completed Interviews Needed

• To estimate a 10% statistic; RSE 30%; Design Effect=1.5
• When no other stratification is needed e.g., SES or health condition
• With intensive screening to obtain required sample sizes
• With appropriate mechanisms in place to handle impact on confidentiality of data
## The Math

<table>
<thead>
<tr>
<th>Age Groups</th>
<th>3</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 Race/Ethnic Groups</td>
<td></td>
<td></td>
</tr>
<tr>
<td>60 places</td>
<td>162,000</td>
<td>342,000</td>
</tr>
<tr>
<td>1000 places</td>
<td>2.7 mil</td>
<td>5.4 mil</td>
</tr>
<tr>
<td>12 Race/Ethnic Groups</td>
<td></td>
<td></td>
</tr>
<tr>
<td>60 places</td>
<td>324,000</td>
<td>648,000</td>
</tr>
<tr>
<td>1000 places</td>
<td>5.4 mil</td>
<td>10.8 mil</td>
</tr>
</tbody>
</table>

For gender specific estimates – multiply by 2
National Health Interview Survey

- Nation’s largest household interview survey
- Provides data on health status and disability, health insurance, immunizations, health behaviors, use of health services
- Obtains data on expanded racial and ethnic groups
  - Hispanics – includes eight subgroups
  - Asian and Pacific Islanders – includes 11 subgroups
- Oversamples blacks, Asians and Hispanics
- Source of bridging data
### National Health Interview Survey (NHIS) sample sizes

<table>
<thead>
<tr>
<th></th>
<th>2005</th>
<th>2007</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hispanic *^</td>
<td>23,642</td>
<td>18,788</td>
</tr>
<tr>
<td>White*</td>
<td>78,650</td>
<td>56,827</td>
</tr>
<tr>
<td>Black*</td>
<td>13,723</td>
<td>11,813</td>
</tr>
<tr>
<td>AIAN*</td>
<td>808</td>
<td>893</td>
</tr>
<tr>
<td>Asian*</td>
<td>3748</td>
<td>4730</td>
</tr>
<tr>
<td>NHOPI</td>
<td>161</td>
<td>124</td>
</tr>
<tr>
<td>Multiple race*</td>
<td>1559</td>
<td>1377</td>
</tr>
</tbody>
</table>

*released on puf  
^ Hispanic persons can be of any race
## NHIS samples sizes – Hispanic population group detail

<table>
<thead>
<tr>
<th>Group</th>
<th>2005</th>
<th>2007</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hispanic*</td>
<td>23,642</td>
<td>18,788</td>
</tr>
<tr>
<td>Puerto Rican*</td>
<td>1,923</td>
<td>1,644</td>
</tr>
<tr>
<td>Mexican*</td>
<td>16,006</td>
<td>12,439</td>
</tr>
<tr>
<td>Cuban*</td>
<td>1,017</td>
<td>655</td>
</tr>
<tr>
<td>Dominican*</td>
<td>547</td>
<td>500</td>
</tr>
<tr>
<td>Central/South Amer*</td>
<td>3,168</td>
<td>2,762</td>
</tr>
<tr>
<td>Other Hispanic groups*</td>
<td>981</td>
<td>788</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>2005</td>
<td>2007</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>AIAN*</td>
<td>808</td>
<td>893</td>
</tr>
<tr>
<td>American Indian</td>
<td>795</td>
<td>808</td>
</tr>
<tr>
<td>Alaska Native</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>NHOPI</td>
<td>161</td>
<td>124</td>
</tr>
<tr>
<td>Native Hawaiian</td>
<td>74</td>
<td>44</td>
</tr>
<tr>
<td>Guamanian</td>
<td>38</td>
<td>23</td>
</tr>
<tr>
<td>Samoan</td>
<td>24</td>
<td>8</td>
</tr>
<tr>
<td>Other Pacific Islander</td>
<td>25</td>
<td>49</td>
</tr>
</tbody>
</table>
## NHIS samples sizes – Asian population group detail

<table>
<thead>
<tr>
<th>Asian*</th>
<th>2005</th>
<th>2007</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asian Indian*</td>
<td>704</td>
<td>804</td>
</tr>
<tr>
<td>Chinese*</td>
<td>704</td>
<td>983</td>
</tr>
<tr>
<td>Filipino*</td>
<td>872</td>
<td>1153</td>
</tr>
<tr>
<td>Japanese</td>
<td>261</td>
<td>296</td>
</tr>
<tr>
<td>Korean</td>
<td>287</td>
<td>401</td>
</tr>
<tr>
<td>Vietnamese</td>
<td>394</td>
<td>475</td>
</tr>
<tr>
<td>Other Asian groups</td>
<td>526</td>
<td>618</td>
</tr>
</tbody>
</table>
Combining Years of Data

- Access to Dental Care Among Hispanic or Latino Subgroups: United States, 2000-03.
- To obtain estimates about the NHOPI population – 8 years of data
What To Do?

- Larger sample sizes are needed, but how large can we go?
- Does all content need to be collected for all geographic units?
- In-person designs are expensive but cheaper methods present quality concerns - is there a way to combine methods
  - Use of a state-specific telephone survey and a large national in-person interview survey to improve racial/ethnic data
Challenges for Vital Statistics
(birth and death data obtained from states)

- Even with a complete universe – problem of small numbers
- Method of collection race/ethnic information affects data quality - Death rates for some race and Hispanic origin groups appear questionable
  - Numerator from Death Certificate - race and Hispanic origin identification by funeral director per family member report or based on observation of decedent.
  - Denominator from Census - race and Hispanic origin identification is self-reported or reported by family/household member
- Comparability evaluated using data from the National Longitudinal Mortality Study
Effect of Correction on Age Adjusted Death Rates

- Observed and Corrected Age Adjusted Death Rates and Percentage Relative to White
  - **AIAN**
    - Observed: 85% of White AADR
    - Adjusted: 110% of White AADR
  - **API**
    - Observed: 60% of White AADR
    - Adjusted: 64% of White AADR
Effect of Correction on Age Adjusted Death Rates

- **Observed and Corrected Age Adjusted Death Rates and Percentage Relative to Non Hispanic White**
  - **Hispanic**
    - Observed: 79%  Adjusted: 83%
  - **Mexican**
    - Observed: 88%  Adjusted: 88%
  - **Puerto Rican**
    - Observed: 93%  Adjusted: 96%
  - **Cuban**
    - Observed: 81%  Adjusted: 82%
  - **Central/South American**
    - Observed: 87%  Adjusted: 69%
  - **Other Hispanic**
    - Observed: 45%  Adjusted: 76%
Challenges for Health care surveys

- Surveys of hospitals, physicians offices, nursing homes and hospices, outpatient and emergency departments
  - Obtain data on facilities, services provided, and characteristics of patients served
- Data are obtained from providers
  - Little control over data collection
  - No requirement for providers to follow any specific standards in data collection can affect data quality
National Hospital Discharge Survey

Number and Percent* of Records (n=200,368)

- Multiple race: 108  .09%
- Not stated: 104,887  24%

*weighted estimates
Ambulatory Care Surveys

Percent* of records for multiple race persons

- **NAMCS:** 0.3
- **NHAMCS-ED:** 0.3
- **NHAMCS-OPD:** 1.4

*weighted estimates; missing responses imputed (27% for NAMCS to 9.8% for NHAMCS-ED)