New Measures to Access the Quality of Race/Ethnicity Reporting in State Databases

David Zingmond, MD, PhD
AHRQ monthly grantees call
January 10, 2012
Aim

- Develop validated audit measures for race/ethnicity reporting that can be used for any state’s statewide databases
Background: Data Auditing

• Current data auditing rules are blunt instruments for determining the accuracy (or adequacy) of race/ethnicity reporting in statewide hospital databases
  – Provide minimum criteria for flagging hospital data
    • Flags for hospitals exceeding rates for missing or unknown race/ethnicity
    • Flags for extreme variation in reporting (100% or 0% for categories)
Unknown race/ethnicity

- Combined race/ethnicity measure
- 3.4% unknown race/ethnicity (mean across 349 hospitals reporting discharges in 2009)
  - 17 hospitals > 10% unknown
  - 1 hospital > 20% unknown
Candidate Audit Measures

• Reflect self-reported race/ethnicity
• Data are readily available for use by those performing the data audits

• Patient-level record comparisons are best
  – not available for every patient or for every state (e.g. where patient-level data linkages are not routinely performed)
Available Data for Comparison

• Patient-level data
  – Birth certificates (mother self-report; not all hospitals have births)
  – Cancer registry (abstracted with data supplemented by name-based algorithm and death certificates; small subset with self-report)
  – Death certificates (institutional reporting)

• U.S. Census data (self-report)
Measures for Agreement – Patient Level Data

• Patient-level data

\[ \% \text{ agreement} = \frac{\# \text{ agree with GS}}{\# \text{ Total}} \times 100 \]

– This type of measure can be used for single category agreement or overall agreement across all categories

– If GS (gold standard) is truly self-report, then this measure can also be used for validation purposes
Measures for Agreement – Summary Level Data

• Hospital Summary Level Data

Estimated disagreement =

Abs(Reported – Predicted)/2 + % Unknown

Reported = Distribution of race categories

Predicted = Population mean predicted distribution using zip-code level distribution for each patient

* agreement = 100 – disagreement
# PDD vs CCR

<table>
<thead>
<tr>
<th></th>
<th>PDD vs. SR</th>
<th>PDD vs. CCR</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Race</strong></td>
<td>0.91</td>
<td>0.92</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td>0.91</td>
<td>0.95</td>
</tr>
<tr>
<td><strong>Alt Ethnicity (unk = NH)</strong></td>
<td>0.95</td>
<td>0.95</td>
</tr>
<tr>
<td><strong>White-White</strong></td>
<td>0.94</td>
<td>0.93</td>
</tr>
<tr>
<td><strong>Non-White-Non-White Race</strong></td>
<td>0.82</td>
<td>0.91</td>
</tr>
<tr>
<td><strong>Hispanic-Hispanic</strong></td>
<td>0.66</td>
<td>0.66</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>PDD vs. SR</th>
<th>PDD vs. CCR</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Combined Race/Ethnicity</strong></td>
<td>0.90</td>
<td>0.92</td>
</tr>
<tr>
<td><strong>NH White-NH White</strong></td>
<td>0.95</td>
<td>0.96</td>
</tr>
<tr>
<td><strong>Non-White-Non-White</strong></td>
<td>0.79</td>
<td>0.83</td>
</tr>
<tr>
<td><strong>Hispanic-Hispanic</strong></td>
<td>0.66</td>
<td>0.66</td>
</tr>
</tbody>
</table>

N = 16,653
# PDD versus Birth

<table>
<thead>
<tr>
<th>Category</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Race</td>
<td>70.0</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>85.8</td>
</tr>
<tr>
<td>White</td>
<td>71.6</td>
</tr>
<tr>
<td>NW</td>
<td>65.4</td>
</tr>
<tr>
<td>Hisp</td>
<td>89.1</td>
</tr>
<tr>
<td>NH</td>
<td>82.1</td>
</tr>
<tr>
<td>Race/Eth</td>
<td>85.8</td>
</tr>
<tr>
<td>NHW</td>
<td>91.1</td>
</tr>
<tr>
<td>minorities</td>
<td>83.9</td>
</tr>
</tbody>
</table>

N = 513,456
Agreement (maternal) vs Mean Population Agreement

$\rho = 0.5$

$N = 513,456$ at 254 Hospitals where births occurred; Agreement between PDD & birth versus PDD & Census
Other observations

- Too much scatter for good matching for prediction
- Populations are not exact matches (mothers versus all adult patients)
Further work on Data Audits

• Revised metrics
• Match populations for derivation of metrics
• Compare proposed metrics to current insensitive, context-free metrics
• Apply metrics to hospitals across time to see if there have been changes in performance during the observation period
Data Collection and Reporting

- Patient
  - Clerk
  - Registration System
    - Medical Record
      - Record Abstract
        - H.I.T.
          - OSHPD

- CCR Registrar
  - Record Abstract
    - CCR

- Decedent Affairs
  - Record Abstract
    - O. Vital Stats
      - Live Births
Existing Gold Standard for Self-Report

Patient ➔ Clerk ➔ Registration System ➔ Medical Record ➔ Record Abstract ➔ OSHPD

CCR Registrar ➔ Record Abstract ➔ CCR

Clinical Trials Database

Self-reported REL from cancer trials