A New Measure to Assess the Completeness of Case Ascertainment

Barnali Das, Ph.D.
Linda Pickle, Ph.D.
Eric J. (Rocky) Feuer, Ph.D.
Lin Clegg, Ph.D.
Surveillance Research Program,
National Cancer Institute
The Problem

- What % of actual incident cancer cases is a registry reporting? Assess registry quality for certification.
  - Actual number of incident cases in the registry unobserved so have to estimate assumptions model
  - Current method: Assumes the ratio of incidence to mortality is constant across registries. Then
    \[
    \text{Expected Local Incidence} = \left\{ \frac{\text{SEER Incidence}}{\text{US Mortality}} \right\} \{\text{Local Mortality}\}
    \]

- Current method: model is implicit and restricted
New Methodology

- Explicit statistical model to find expected incidence*.
  - Include mortality + demographic + lifestyle factors affecting incidence.
  - Account for spatial variation.
- Extend to account for delay in reporting - use NCI delay model**
- Completeness index constructed by comparing observed incidence rate to model estimate
- Weighting scheme for cancer sites, sex, race unchanged from current method
- Use 20 cancer sites, including prostate cancer

Results: Unadjusted for Registry Differences

- Some correlation
- 100% completeness exceeded by both
- New index does worse in small population areas
Results: Adjusted for Registry Differences

- New Index improves when differences between SEER-NPCR and NPCR only registries are accounted for.
Results: Adjusted for Delay and Registry-specific Differences

- New Index improves further after reporting delay is adjusted for.
## Impact on Certification

(based solely on completeness)

<table>
<thead>
<tr>
<th>New Index</th>
<th>NAACCR Index</th>
<th>Gold</th>
<th>Silver</th>
<th>None</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gold</td>
<td>Gold</td>
<td>20</td>
<td>2</td>
<td>?</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>Silver</td>
<td>3</td>
<td>2</td>
<td>?</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>None</td>
<td>1</td>
<td>1</td>
<td>?</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>24</td>
<td>5</td>
<td>?</td>
<td>29</td>
</tr>
</tbody>
</table>
Variance of the New Index

- Variance of new index (and current index) may be found approximately by statistical asymptotic theory

- Variability in the new index is due to
  - the variability of the observed rates (large)
  - the variability of the predicted rates (small)
  - the correlation between the two (not calculated)

- Calculated variance is conservative (larger) owing to the omission of the third component: more conservative for larger population registries than small population registries
95% Confidence Intervals for New Index by Registry
Incorporating Variability

- For some registries confidence intervals range from uncertified to gold – what should be the certification?

- Alternative: find probabilities that a given registry falls into each certification status

- Award that certification that has highest probability.

<table>
<thead>
<tr>
<th>registry</th>
<th>P(gold)</th>
<th>P(silver)</th>
<th>P(uncertified)</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>XXXXXX</td>
<td>55%</td>
<td>28%</td>
<td>17%</td>
<td>GOLD</td>
</tr>
<tr>
<td>Registry</td>
<td>P(gold)</td>
<td>P(Silver)</td>
<td>P(Uncertified)</td>
<td>Result</td>
</tr>
<tr>
<td>----------</td>
<td>---------</td>
<td>-----------</td>
<td>----------------</td>
<td>--------</td>
</tr>
<tr>
<td>AK</td>
<td>6.94</td>
<td>32.48</td>
<td>60.58</td>
<td>U</td>
</tr>
<tr>
<td>AZ</td>
<td>0.00</td>
<td>43.58</td>
<td>56.42</td>
<td>U</td>
</tr>
<tr>
<td>CA</td>
<td>100.00</td>
<td>0.00</td>
<td>0.00</td>
<td>G</td>
</tr>
<tr>
<td>CT</td>
<td>100.00</td>
<td>0.00</td>
<td>0.00</td>
<td>G</td>
</tr>
<tr>
<td>FL</td>
<td>100.00</td>
<td>0.00</td>
<td>0.00</td>
<td>G</td>
</tr>
<tr>
<td>ATL</td>
<td>98.59</td>
<td>1.41</td>
<td>0.00</td>
<td>G</td>
</tr>
<tr>
<td>HI</td>
<td>98.29</td>
<td>1.71</td>
<td>0.00</td>
<td>G</td>
</tr>
<tr>
<td>ID</td>
<td>99.99</td>
<td>0.01</td>
<td>0.00</td>
<td>G</td>
</tr>
<tr>
<td>IA</td>
<td>99.95</td>
<td>0.05</td>
<td>0.00</td>
<td>G</td>
</tr>
</tbody>
</table>
Incorporating Variability: Results (ctd)

<table>
<thead>
<tr>
<th>Registry</th>
<th>P(gold)</th>
<th>P(Silver)</th>
<th>P(Uncertified)</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>KY</td>
<td>100.00</td>
<td>0.00</td>
<td>0.00</td>
<td>G</td>
</tr>
<tr>
<td>LA</td>
<td>99.76</td>
<td>0.24</td>
<td>0.00</td>
<td>G</td>
</tr>
<tr>
<td>DET</td>
<td>100.00</td>
<td>0.00</td>
<td>0.00</td>
<td>G</td>
</tr>
<tr>
<td>MT</td>
<td>78.95</td>
<td>16.95</td>
<td>4.10</td>
<td>G</td>
</tr>
<tr>
<td>NE</td>
<td>99.99</td>
<td>0.01</td>
<td>0.00</td>
<td>G</td>
</tr>
<tr>
<td>NH</td>
<td>18.33</td>
<td>67.25</td>
<td>14.42</td>
<td>S</td>
</tr>
<tr>
<td>NJ</td>
<td>100.00</td>
<td>0.00</td>
<td>0.00</td>
<td>G</td>
</tr>
<tr>
<td>NM</td>
<td>20.68</td>
<td>74.39</td>
<td>4.93</td>
<td>S</td>
</tr>
<tr>
<td>NC</td>
<td>12.51</td>
<td>87.49</td>
<td>0.00</td>
<td>S</td>
</tr>
</tbody>
</table>
### Incorporating Variability: Results (ctd)

<table>
<thead>
<tr>
<th>Registry</th>
<th>P(gold)</th>
<th>P(Silver)</th>
<th>P(Uncertified)</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>ND</td>
<td>66.64</td>
<td>27.11</td>
<td>6.25</td>
<td>G</td>
</tr>
<tr>
<td>OR</td>
<td>99.93</td>
<td>0.07</td>
<td>0.00</td>
<td>G</td>
</tr>
<tr>
<td>PA</td>
<td>99.99</td>
<td>0.01</td>
<td>0.00</td>
<td>G</td>
</tr>
<tr>
<td>RI</td>
<td>99.99</td>
<td>0.01</td>
<td>0.00</td>
<td>G</td>
</tr>
<tr>
<td>SC</td>
<td>82.32</td>
<td>17.68</td>
<td>0.00</td>
<td>G</td>
</tr>
<tr>
<td>UT</td>
<td>35.89</td>
<td>61.83</td>
<td>2.28</td>
<td>S</td>
</tr>
<tr>
<td>WA</td>
<td>100.00</td>
<td>0.00</td>
<td>0.00</td>
<td>G</td>
</tr>
<tr>
<td>STL</td>
<td>100.00</td>
<td>0.00</td>
<td>0.00</td>
<td>G</td>
</tr>
<tr>
<td>WV</td>
<td>99.88</td>
<td>0.12</td>
<td>0.00</td>
<td>G</td>
</tr>
<tr>
<td>WI</td>
<td>0.02</td>
<td>85.04</td>
<td>14.94</td>
<td>S</td>
</tr>
<tr>
<td>WY</td>
<td>80.17</td>
<td>17.91</td>
<td>1.92</td>
<td>G</td>
</tr>
</tbody>
</table>
Conclusions

- Statistical modeling predicts expected incidence more accurately
- Adjusting for differences between SEER-NPCR and NPCR-only registries and reporting delay helps reduce unrealistic completeness indices
- New index may certify a registry differently – hard to draw firmer conclusions with only certified data....
- Demonstrated the possibility of integrating uncertainty of index in the certification process
New directions

- How can variability be incorporated in the certification process?
  - Certification is unfair to ALL registries (large and small population) without using variability

- Can more races be used?

- Please give us your uncertified data .....
EXTRA SLIDES
Should the Index Include Prostate Cancer?

- Yes – the effect of PSA screening has now stabilized (see plot)
- Prostate cancer is a major cancer in the US – omission cannot be justified if rates are stable
Incidence Model Details

- Modeling based on CINA Deluxe data
- Explanatory variables from Census, BRFSS, Area Resource File and NCHS.
- 20 sites modeled for WBO.
Obtaining Expected Incidence by Modeling

- Regression model predicting county level incidence based on local demographic and lifestyle factors.

- eg. Female lung cancer rate
  - Strong spatial pattern
  - Strongly related to lifestyle: tobacco usage

Details of Model Variables

- **Age**
- **Log mortality rate**
- **Race (W, B, O)**
- **Ethnicity/origin**: % Hispanic, Black, Asian/Pac. Islander, AI/AN
- **Medical facilities**: MD & mammogram screening facility density
- **Household characteristics**: % female head of house, crowded
- **Socioeconomic status**:
  - **Income**: per capita, % < poverty level
  - **Education**: % < 9 years, % 4+ years college
  - **Other**: % unemployment
- **Urban/rural indicators**: urban/rural continuum code, pop. density
- **Geography**: Census Region (NE, MW, S, W), latitude, longitude
- **Lifestyle**: % ever smoked, % obese, % had mammogram last 2 years, % with no health insurance
Current Method: Principle

- **Incidence to mortality ratio based model**
- **Assumption:** \( \frac{\text{incidence}}{\text{mortality}} \) a stable ratio for most sites
- Therefore use model

\[
\frac{\text{Expected Local Incidence}}{\text{Local Mortality}} = \frac{\text{SEER Incidence}}{\text{US Mortality}}
\]

to find expected incidence in an area and compare to the observed incidence
Current Method : Details

- Expected age adjusted incidence rate estimate based on ratio of incidence to mortality rates
  - Ratio assumed constant within race, sex, cancer site groupings across geographic region
  - Mortality adjusted for case fatality

- Completeness index by comparing observed incidence rate to estimate

- Index based on 18 sites (exclude prostate)

- Final completeness estimate adjusted for cancer site, sex and race by weighting
References for NAACCR Method


Cancer Sites Used by NAACCR Index

- Oral Cavity and Pharynx
- Esophagus
- Stomach
- Colon and Rectum
- Liver
- Pancreas
- Lung and Bronchus
- Melanomas of the Skin
- Female Breast
- Cervix
- Corpus and Uterus
- Ovary
- Urinary Bladder
- Kidney and Renal Pelvis
- Brain and Other Nervous System
- Hodgkin’s Disease
- Non-Hodgkins Lymphomas
- Multiple Myeloma
- Leukemias

NOTE: Prostate (WB) and Melanoma (B) omitted
NAACCR Worksheet for Registries

- NAACCR worksheet – adjustments for stability and accuracy
- Three major steps
  - Calculate completeness by site, gender and race (W/B)
  - Weight and combine individual indices to obtain one overall measure by registry
  - Calculate completeness by registry after adjusting for duplicate records
Adjusting for Reporting Delay

- Cases for a given year reported later – outpatient settings

- Extent of delay varies by site: 1998 data shows melanoma (14%), colorectal (3%)

- Ideally should be reflected in expected incidence prediction

- Use NCI delay model to adjust
  