Improving Cause-Specific Survival Estimates: SEER Data

Nadia Howlader, MS
Mathematical Statistician
Data Analysis and Interpretation Branch (DAIB)
Surveillance Research Program (SRP)
National Cancer Institute (NCI)

Methods and Application for Population Based Survival, Frascati, Italy
September 20-21, 2010
Collaboration

- Lynn Ries
- Angela Mariotto
- Marsha Reichman
- Jennifer Ruhl
- Kathy Cronin
Outline

• Background
• Relative survival approach limitations
• Issues with cause of death (COD)
• Algorithm for COD assignment (varies by one and only cancer vs. 1st of multiple cancers)
• Compare relative and cause-specific (using new COD variable) estimates
• Conclusion
Accurate estimates of cancer survival are important. Population-based studies often use relative survival. A ratio of observed to expected survival rates. Observed survival rates (cancer patients). Expected survival rates (US general population life-table).
Background (Cont’d)

• Challenging to estimate relative survival rates for subgroups of population
  • lack of “appropriate” life-tables
  • ethnic minorities, risk factors, socioeconomic status, geographic area
  • “Other-cause” mortality are not always well represented (Hyunsoon’s talk)
• Concord study (2008) developed sex- and geography-specific life tables

• SEER collects cause-of death information from death certificates
Study Aim

• Parallel to improving life-tables, could we use cause of death information to obtain improved estimates of cause-specific survival rates?

• Cause-specific widely used in clinical trials but used with caution in registry data (why?)
Issues with Cause of Death (COD)

• Death certificate errors
  – Metastatic site of the primary cancer diagnosis may be reported as the underlying COD

• How to assign CODs to a primary cancer diagnosis?

• Need to develop an algorithm to identify a single, disease-specific, underlying COD
SEER Cause-Specific Death Classification Variable

- Underlying COD was evaluated
- The algorithm takes into account COD in conjunction with
  - Site of original cancer diagnosis
  - Tumor sequence (Seq 00 vs. Seq 01)
  - Co-morbidities (e.g., HIV/AIDS and/or site-related non-cancer diseases)
SEER Cause-Specific Death Classification Variable (Cont’d)

• Determine the algorithm:
  – Step 1: Careful analysis of all possible COD from death certificates for each cancer site
  – Step 2: Look at all possible COD to assess the chances that was due to the cancer of interest
SEER Cause-Specific Death Classification Variable (Cont’d)

- Step 3: Use heuristic approach to classify each possible COD as “cancer” or “other-cause” death

- Broad categories of COD were identified
Major COD Categories

- Cancer of the same site
- Cancer of the same organ system
- Cancer of any other sites
- HIV alone (varies by cancer site)
- AIDS and cancer
- Site-specific disease
Example: Deaths attributable to lip cancer for cases diagnosed with only lip cancer (Sequence 00)

<table>
<thead>
<tr>
<th>COD Categories</th>
<th>Death attributed to lip cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cancer of the same site</td>
<td>Lip cancer</td>
</tr>
<tr>
<td>Cancer of the same body system</td>
<td>Oral cavity and pharynx cancer</td>
</tr>
<tr>
<td>Cancer of any other site</td>
<td>Skin Cancer</td>
</tr>
<tr>
<td>AIDS and cancer</td>
<td>AIDS complications</td>
</tr>
<tr>
<td>HIV alone*</td>
<td>HIV disease resulting in bacterial infection</td>
</tr>
<tr>
<td>Site –specific disease</td>
<td>Lump in the mouth</td>
</tr>
</tbody>
</table>
Cases that were diagnosed with only one cancer (sequence 00)

*Other cause is treated as censored observation
Cases that were diagnosed with only rectum cancer (sequence 00)

*Other cause is treated as censored observation
Cases that were only diagnosed with only lip cancer (sequence 00)

*Other cause is treated as censored observation
Cases that were only diagnosed with only kaposi sarcoma (sequence 00)

*Other cause is treated as censored observation
Cases with Non-Hodgkin Lymphoma (Sequence 00)

Age at Diagnosis

Survival Probability

Red: Cause-specific (old)
Green: Cause-specific (new)
Yellow: Relative Survival

Note: Five-year Non-Hodgkin lymphoma cancer survival by age at diagnosis, SEER-13, 1992-2004
Cases with more than one cancer (Sequence 01)

• More stringent rule was applied
• Cancer of the same site of 1\textsuperscript{st} diagnosis were attributed
  – COD due to 2\textsuperscript{nd} cancer were treated as censored
• COD for AIDS defining cancer
  – AIDS and cancer
  – HIV alone
How do cause-specific rates using **new COD** variable compare to relative survival rates?
100+ Cancer Sites in SEER, 65+ Ages
Compare Relative and Cause-specific* Estimates (Con’t)

• Relative = Observed/Expected

• If expected rate is overestimated then
  \((\downarrow)\) Relative = Observed/Expected (\(\uparrow\))

• If expected rate is underestimated then
  \((\uparrow)\) Relative = Observed/Expected (\(\downarrow\))

*Using SEER Cause-Specific Death Classification Variable
### Example of Relative Survival Being Problematic: SEER-13, 1992-2004

<table>
<thead>
<tr>
<th></th>
<th>White</th>
<th>AI/AN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Relative</td>
<td>C-S</td>
</tr>
<tr>
<td>Breast, In-situ &amp; 65+</td>
<td>107.5</td>
<td>98.6</td>
</tr>
<tr>
<td>Prostate, L/R &amp; 65+</td>
<td>104.5</td>
<td>94.8</td>
</tr>
<tr>
<td>Lung, All Stage &amp; &lt;65</td>
<td>18.7</td>
<td>20.5</td>
</tr>
<tr>
<td>Oral Cavity, All Stage &amp; &lt;65</td>
<td>67.2</td>
<td>71.6</td>
</tr>
</tbody>
</table>

Note: AI/AN = American Indian/Alaska Native; C-S = Cause-specific; L/R=Localized/Regional
Absent SES-specific life tables, relative survival rates exaggerate disparity

<table>
<thead>
<tr>
<th>2000 Census Tract Poverty</th>
<th>Relative survival rate</th>
<th>Cause-Specific survival rate</th>
<th>Bias in expected survival rate</th>
<th>Bias in relative survival rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hi SES, &lt;10.0%</td>
<td>86.5</td>
<td>84.1</td>
<td>↓</td>
<td>↑</td>
</tr>
<tr>
<td>Medium SES, 10.0% - 19.9%</td>
<td>81.6</td>
<td>81.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low SES, 20.0%+</td>
<td>73.2</td>
<td>75.2</td>
<td>↑</td>
<td>↓</td>
</tr>
<tr>
<td>Difference between High and Low SES</td>
<td><strong>13.3</strong></td>
<td><strong>8.8</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Conclusion

• Developed algorithm associating COD to primary cancer

• Improved cause-specific survival rate estimates

• In most cases, relative survival estimates are in agreement with cause-specific survival estimates

• However, relative survival is not suitable for
  – Heavily screened, different SES, high risk of cancer and other diseases, ethnic minorities
Conclusion (Cont’d)

• Lack of “appropriate” life-table information led to biased survival estimates with relative survival approach

• Cause-specific estimates could be useful for the above subgroups to provide accurate and reliable survival measures

• New COD variable implemented in seer*stat software
For More Information


• Corresponding author
Nadia Howlader
E-mail: howladern@mail.nih.gov
Thank you!