Q&A

- Please submit all questions concerning webinar content through the Q&A panel.

Reminder:

- If you have participants watching this webinar at your site, please collect their names and emails.
  - We will be distributing a Q&A document in about one week. This document will fully answer questions asked during the webinar and will contain any corrections that we may discover after the webinar.
Fabulous Prizes!!!

1. CINA Survival Statistics Preview and Introduction

Chris Johnson
Cancer Data Registry of Idaho
Co-Chair, NAACCR Survival Analysis Task Force
What is Population-based Survival (PBS)

- Clinical trials - highest achievable survival
  - Clinical focus: Value of one treatment vs. another
  - 3-4% of adult patients participate in clinical trials. Participants are younger and healthier.
- Population - survival achieved
  - Impact of cancer control initiatives (across the spectrum of initiatives)
    - Targeting and monitoring cancer control initiatives
  - Policy-setting
    - Effectiveness of healthcare delivery – standard measure of cancer system performance

DURC SAWG → RDU PAN SATF

- The NAACCR Data Use and Research Committee’s Survival Analysis Work Group was formed in 2008 chiefly to evaluate how survival estimates are impacted by “active” follow-up versus ascertainment of deaths only.
  (“Active” follow-up may all be done through record linkage.)
Goal = CINA Survival Statistics

- Under the NAACCR Strategic Management Plan, the Survival Analysis Task Force is responsible for facilitating the development of state and province-specific relative survival data.

- Expected Deliverables:
  - State and province-specific relative survival statistics published in annual CINA publications.

NAACCR Call for Data and Survival

- November 2013 Call for Data
  - NAACCR Prep – survival variables
  - Vital Status Follow-Up Activities Form
  - Active permission from registries to publish registry-specific and aggregated survival estimates
CINA 2006-2010 Included Stage Data
Survival Data coming soon...

SATF has not decided on templates for CINA survival statistics. Meanwhile, SEER CSR examples:

Cancer of the Colon and Rectum (Invasive)
5-Year Relative and Period Survival by Race, Sex, Diagnosis Year, Age and Stage at Diagnosis

5-Year Relative Survival (Percent) by Year of Diagnosis

<table>
<thead>
<tr>
<th>Year of Diagnosis</th>
<th>All Races</th>
<th>White</th>
<th>Black</th>
<th>All Races</th>
<th>White</th>
<th>Black</th>
</tr>
</thead>
<tbody>
<tr>
<td>2004-2010b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Both sexes</td>
<td>Male</td>
<td>Female</td>
<td>Both sexes</td>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td></td>
<td>65.1</td>
<td>66.4</td>
<td>63.7</td>
<td>65.5</td>
<td>67.0</td>
<td>66.2</td>
</tr>
</tbody>
</table>

5-Year Period Survival (Percent)

<table>
<thead>
<tr>
<th>Year</th>
<th>All Races</th>
<th>White</th>
<th>Black</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Both sexes</td>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>2010</td>
<td>65.9</td>
<td>65.2</td>
<td>64.8</td>
</tr>
</tbody>
</table>
SATF has not decided on templates for CINA survival statistics. Meanwhile, SEER CSR examples:

### 5-Year Relative Survival (Percent) 2004-2010: by Age at Diagnosis

<table>
<thead>
<tr>
<th>Age at Diagnosis</th>
<th>All Races</th>
<th>Whites</th>
<th>Blacks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ages &lt;45</td>
<td>67.6</td>
<td>66.2</td>
<td>69.3</td>
</tr>
<tr>
<td>Ages 45-54</td>
<td>70.8</td>
<td>69.6</td>
<td>72.3</td>
</tr>
<tr>
<td>Ages 55-64</td>
<td>68.4</td>
<td>67.8</td>
<td>69.2</td>
</tr>
<tr>
<td>Ages 65-74</td>
<td>67.4</td>
<td>66.8</td>
<td>67.9</td>
</tr>
<tr>
<td>Ages 75+</td>
<td>57.0</td>
<td>57.2</td>
<td>56.8</td>
</tr>
<tr>
<td>Ages &lt;65</td>
<td>69.1</td>
<td>69.2</td>
<td>69.3</td>
</tr>
<tr>
<td>Ages 65+</td>
<td>61.4</td>
<td>62.5</td>
<td>62.9</td>
</tr>
</tbody>
</table>

### 5-Year Relative Survival (Percent) 2004-2010: by Stage at Diagnosis

<table>
<thead>
<tr>
<th>Stage at Diagnosis</th>
<th>All Races</th>
<th>Whites</th>
<th>Blacks</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Stages</td>
<td>64.7</td>
<td>65.5</td>
<td>65.2</td>
</tr>
<tr>
<td>Localized</td>
<td>89.8</td>
<td>89.9</td>
<td>89.8</td>
</tr>
<tr>
<td>Regional</td>
<td>70.5</td>
<td>70.4</td>
<td>71.1</td>
</tr>
<tr>
<td>Distal</td>
<td>12.9</td>
<td>12.4</td>
<td>13.3</td>
</tr>
<tr>
<td>Unstaged</td>
<td>33.2</td>
<td>36.2</td>
<td>30.7</td>
</tr>
</tbody>
</table>

### Cancer of the Colon and Rectum (Invasive)

#### SEER® Relative Survival (Percent) By Year of Diagnosis, All Races, Males and Females

<table>
<thead>
<tr>
<th>Survival Time</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-year</td>
<td>83.5</td>
<td>84.0</td>
<td>83.6</td>
<td>83.9</td>
<td>84.6</td>
<td>84.2</td>
</tr>
<tr>
<td>2-year</td>
<td>76.6</td>
<td>77.0</td>
<td>77.1</td>
<td>77.1</td>
<td>78.2</td>
<td>77.0</td>
</tr>
<tr>
<td>3-year</td>
<td>71.6</td>
<td>72.5</td>
<td>72.1</td>
<td>72.2</td>
<td>73.0</td>
<td></td>
</tr>
<tr>
<td>4-year</td>
<td>68.4</td>
<td>68.8</td>
<td>68.3</td>
<td>68.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5-year</td>
<td>65.8</td>
<td>66.1</td>
<td>65.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-year</td>
<td>63.9</td>
<td>64.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7-year</td>
<td>62.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
What is the hold up?

- *!@%! Life Tables for U.S. States

Topics in Survival Data

Agenda

- 2. Evaluation and selection of registry data for CINA Survival Statistics (Chris Johnson for Hannah Weir)
- 3. Approaches to handling multiple primaries (Chris Johnson for Hannah Weir)
- 4. Types of survival statistics: observed, net, relative, cause-specific, crude probability of death, period, cohort, complete (Diane Nishri, Diana Withrow)
- 5. Age-standardized survival (Diane Nishri, Diana Withrow)
6. Choosing appropriate life tables for relative survival (Angela Mariotto)
7. Cause-specific survival using special variables and comparisons to relative survival (Nadia Howlader)
8. Limitations in interpretability of survival among Hispanics and Asians (Paulo Pinheiro)
9. Active follow-up versus presumed alive survival calculations – which is appropriate for your data? (Chris Johnson)
10. Analyzing your own data. Loading your NAACCR 2014 Call for Data dataset into SEER*Prep, analysis of your data using SEER*Stat. (Chris Johnson)
2. Evaluation and selection of registry data for CINA Survival Statistics

- (Chris Johnson for Hannah Weir)

Data Requirements

- Ascertain all cases in population catchment area
  - Complete and high quality incidence data (e.g., NAACCR Certified)
- Ascertain all deaths among patients in registry
  - Incidence data linked to state and national death certificate databases and/or active follow-up
- Dataset passes survival-related edits
- Complete dates of birth, diagnosis and death
- Data quality indicators available to facilitate interpretation of results (%DCO, %MP, %MC)
Incidence Data

- **Issue:** Survival estimates are influenced by case finding.
- **Scenarios:**
  - Less fatal cancers (e.g., prostate, hematological cancers) may be diagnosed in a doctor’s office.
  - Some cases diagnosed clinically may have poor prognosis.
- **Potential impact:** Survival estimates can be biased (over- or under-estimated) by incomplete incidence data.
- **Assessment:**
  - NAACCR Certification (including edits, %DCO, %case completeness)
  - % MV (SEER 11: 1992-2008) Mean 94.7% (range 92.2% - 95.7%)

“The validity of population-based survival comparisons is clearly dependent on the validity of the incidence data.” Berrino, 2003

Ascertainment of Deaths

- **Issue:** Missing deaths
- **Scenario:**
  - A person diagnosed with cancer moves to another state, and dies there.
- **Potential impact:** Survival estimates can be biased (over-estimated) by incomplete death ascertainment.
- **Assessment:**
  - Follow-up source central
  - Comparisons of survival rates by registry
Data Evaluation

Data Variables and Edits

- Date of diagnosis
- Date of last contact
- Follow-up source central
- Type of report source
- Diagnostic confirmation
- Vital status
- Cause of death
- ICD revision number

Edits associated with vital status variables needed for survival analysis:
- Date of diagnosis
- Date of last contact
- Follow-up source central
- Type of report source
- Diagnostic confirmation
- Vital status
- Cause of death
- ICD revision number

Verify cause of death same on all records for a patient (SEER IR11)
Verify date of follow-up same on all records for a patient (SEER IR08)
Verify vital status same on all records for a patient (SEER IR10)
Upcoming Publication

3. Approaches to handling multiple primaries

- (Chris Johnson for Hannah Weir)

What is a multiple primary?

- Approximately 880,300 of the 11 million cancer survivors living in the US as of January 1, 2005, had been diagnosed with more than one cancer.
- Most of these second or more cancers would be expected to occur even if cancer survivors had the same risk of cancer as the general population.
- Multiple primary cancers can either be diagnosed at the same time (synchronous) or at different times (metachronous).
Risk of Subsequent Cancers

- Although cancer survivors as a group have a small (14%) increased lifetime risk of developing new cancers compared with the general population, some subgroups of patients have a much higher risk.
- The risk of developing subsequent cancers varies by the type of first cancer diagnosed, age at first diagnosis, environmental exposures, genetic factors, treatment, and other factors.

Multiple Primary Rule Sets

- SEER Rules
  - The 2007 Multiple Primary and Histology Coding Rules present the first site-specific multiple primary and histology rules developed to promote consistent and standardized coding by cancer registrars.
  - Complicated!
- IARC/IACR Rules
  - Simpler, result in fewer primaries
Old SEER MP/H Rules (general)

- A cancer of a different site and histologic (microscopic composition of cells and/or tissue) type than the original cancer is considered a separate primary.
- Cancers of different histologic types in the same site are considered separate primaries regardless of whether they are diagnosed at the same or different times.
- A new cancer of the same site or with the same histology as an earlier one is considered the same primary cancer if diagnosed within 2 months or a separate primary cancer if diagnosed after 2 months, unless the medical record specifically states that it is recurrent or metastatic disease.
- If an organ is paired, each member of the pair is generally considered to be a separate site.
- Important exceptions to these general rules include most histological types of cancer in the prostate and urinary bladder, for which multiple tumors are reported as a single primary with the date of the first invasive lesion.
- A different set of rules is used to determine multiple primaries of the lymphatic and hematopoietic systems.

SEER 9 Sequence Distribution, 2007-2011
Percentage of multiple primary cancers (all sites and both sexes combined) by IACR and SEER multiple primary rules (1995–2008)

---

Approaches to handling multiple primaries

- Survival estimates based on first cancers-only exclude a large and increasing number of MP cancers.
- Several recent papers advocate that data on all cancers should be included in the analysis in order to produce clinically and epidemiologically relevant and less biased cancer survival estimates.
- If all first cancers matching the selection criteria are used to produce site-specific survival estimates, then the choice of which MP rule set (SEER or IACR) is used to identify primary cancers has little impact.
Example

Sequence number and primary site category

01 Breast -> 02 Colon -> 03 Breast

Multiple Primary Selection
First Primary Only
First Primary Matching Selection Criteria
Analysis of Breast Cancer Survival
Analysis of Colon Cancer Survival
All Tumors Matching Selection Criteria
Stratified by site*

Analysis

<table>
<thead>
<tr>
<th>All Sites Combined</th>
<th>Breast</th>
<th>Colon</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>01</td>
<td>X</td>
</tr>
</tbody>
</table>

*To be able to produce all three statistics (all, breast, colon) in one run, you need to select "All Tumors Matching Selection Criteria" and stratify by site.

Relevant Publications

- Full dates (day, month, year) should be used in population-based survival studies. Woods LM, Rachet B, Ellis L, Coleman MP. Int J Cancer. 2012 Oct 1;131(7):E1120-4.
Relevant Publications

4. Types of survival statistics: observed, net, relative, cause-specific, crude probability of death, period, cohort, complete

(Diana Withrow)
Acknowledgements & Disclaimers

- **Acknowledgements:** Material for these slides was heavily informed by the presentations of Drs. Paul Dickman and Paul Lambert for their Cancer Survival course held at the 2012 NAACCR annual meeting.

- **Disclaimer:** I am not a biostatistician.

---

**What will I be talking about?**

- **Mortality**
- **Survival**
  - **Net**
  - **Observed**
    - **Cause-specific**
    - **Relative survival**
    - **All-cause**
  
  **Crude probability of death**
Survival vs. Mortality

**Survival**
- The proportion of patients alive after a given time interval has passed since a cancer diagnosis
- E.g. The 5-year survival from breast cancer among women in Canada was 88% in 2013

**Mortality**
- As a cancer statistic:
  - the number of people who have died from cancer in a year
  - E.g. The age-standardized mortality rate for all cancer in Ontario was 145.3 deaths per 100,000 persons in 2013
- Among people diagnosed with cancer:
  - The proportion of patients dead after a given time interval has passed since cancer diagnosis

Survival of cancer patients

Observed vs. Net Survival

- Observed
- Net

Observed vs. Net Survival of cancer patients

- All-cause
- Cause-specific

Net

Relative survival

Mortality

Survival

- Observed
- All-cause

Crude probability of death
Observed survival

- Estimates mortality within a group of persons diagnosed with cancer
- Observed mortality = \( \frac{\text{number of deaths}}{\text{person-time at risk}} \)
- Measures mortality experience of cancer patients rather than mortality specifically due to cancer
- +: minimal data elements required, easy to calculate and interpret
- -: reflects mortality associated with cancer AND background mortality

Mortality

Survival

Net

Observed

Relative survival

All-cause

Cause-specific

Crude probability of death
Cause-specific survival

• Estimates net mortality (mortality associated with a dx of cancer) under certain assumptions
• Deaths attributed to cancer are considered events, deaths due to other causes are censored
• Cause-specific survival = \frac{\text{number of cancer deaths}}{\text{person-time at risk}}
• +: relatively simple to calculate and understand
• -: requires reliably coded cause of death, distinction between cancer and non-cancer death

Cause-specific survival: COD considerations

Figure 1. Categories of death relevant to cause-specific and relative survival/excess mortality estimation of survival

Figure 2. Five-year non-Hodgkin lymphoma cancer survival by age at diagnosis, 1992–2004. Red line = relative survival; blue line = cause-specific survival with non-Hodgkin lymphoma deaths; green line = cause-specific survival with cancer and AIDS death; brown line = cause-specific survival with all malignant cancer deaths.
Relative Survival

- A measure of excess mortality as a result of cancer
- The relative survival ratio (RSR) is the ratio of the observed survival in the patient group to the expected survival of a comparable group from the general population
- +: no COD data required, measures excess mortality irrespective of whether it is directly or indirectly attributable to the cancer
- -: must have an estimate of expected survival in a comparable group from the general population
Relative Survival

- Life tables
- Should be specific to geography and stratified by age, sex, calendar time and race/ethnicity

The Northampton Life Table, from R Price, Observations on reversionary payments: an schemes for providing annuities for widows, and for persons in old age: on the method of calculating the values of assurances on lives: and on the national debt, 4th edn (2 vols), London: T Cadell 1783.

Taken from: http://www.york.ac.uk/depts/maths/histstat/lifework.htm
Crude probability of death

- 1-RSR: proportion of patients who will die of cancer within $i$ years of follow-up in the hypothetical situation where the cancer in question is the only possible cause of death (net probability of death)

- BUT we might want to estimate the proportion of patients who will die of cancer in the presence of competing risks (crude probability of death)
  - While still overcoming cause of death issues

Figure 1. Cumulative probability of death in men with localized prostate cancer over the age of 70. STATISTICS IN MEDICINE Statist. Med. 2000; 19:1729–1740
Interpretation

• In the hypothetical world where it is not possible to die of other causes, the probability of dying of prostate cancer within 15 years of diagnosis is 40%

• In the real world where it is possible to die of other causes, the probability of dying of prostate cancer within 15 years of diagnosis is less than 20%

Resources


Three approaches to survival estimation

1. **Cohort analysis**
   If we are estimating 5 year survival, then all patients must have potential follow-up of 5 years. This means that they must have been diagnosed at least 5 years ago.

2. **Complete analysis**
   If we are estimating 5 year survival, all patients are included, regardless of how much potential follow-up they have. This means that they must have been diagnosed at most 5 years ago.
Three approaches to survival estimation

3. Period analysis

Period analysis exclusively reflects the survival experience of subjects within the most recent calendar period for which the follow-up is available. This is achieved by left truncation of observations at the beginning of this period in addition to censoring at its end.

Survival observed in 20 patients

Ontario colorectal cancer cases, age 15-99

‘Complete’ relative survival, diagnosed 2000-2005, followed to Dec 31, 2005
Patients included in cohort analysis

Ontario colorectal cancer cases, age 15-99

Cohort relative survival, diagnosis year 2000

<table>
<thead>
<tr>
<th>start</th>
<th>end</th>
<th>n</th>
<th>d</th>
<th>w</th>
<th>p</th>
<th>cr_e2</th>
<th>lo_cr_e2</th>
<th>hi_cr_e2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>.5</td>
<td>6651</td>
<td>1083</td>
<td>0</td>
<td>0.8372</td>
<td>0.8521</td>
<td>0.8428</td>
<td>0.8609</td>
</tr>
<tr>
<td>.5</td>
<td>1</td>
<td>5568</td>
<td>425</td>
<td>0</td>
<td>0.9237</td>
<td>0.7992</td>
<td>0.7886</td>
<td>0.8094</td>
</tr>
<tr>
<td>1</td>
<td>1.5</td>
<td>5143</td>
<td>362</td>
<td>0</td>
<td>0.9296</td>
<td>0.7554</td>
<td>0.7438</td>
<td>0.7666</td>
</tr>
<tr>
<td>1.5</td>
<td>2</td>
<td>4781</td>
<td>344</td>
<td>0</td>
<td>0.9280</td>
<td>0.7124</td>
<td>0.7002</td>
<td>0.7244</td>
</tr>
<tr>
<td>2</td>
<td>2.5</td>
<td>4437</td>
<td>268</td>
<td>1</td>
<td>0.9396</td>
<td>0.6814</td>
<td>0.6686</td>
<td>0.6939</td>
</tr>
<tr>
<td>2.5</td>
<td>3</td>
<td>4168</td>
<td>220</td>
<td>0</td>
<td>0.9472</td>
<td>0.6568</td>
<td>0.6436</td>
<td>0.6697</td>
</tr>
<tr>
<td>3</td>
<td>3.5</td>
<td>3948</td>
<td>177</td>
<td>1</td>
<td>0.9552</td>
<td>0.6393</td>
<td>0.6258</td>
<td>0.6526</td>
</tr>
<tr>
<td>3.5</td>
<td>4</td>
<td>3770</td>
<td>181</td>
<td>0</td>
<td>0.9520</td>
<td>0.6199</td>
<td>0.6061</td>
<td>0.6336</td>
</tr>
<tr>
<td>4</td>
<td>4.5</td>
<td>3589</td>
<td>148</td>
<td>0</td>
<td>0.9588</td>
<td>0.6065</td>
<td>0.5923</td>
<td>0.6204</td>
</tr>
<tr>
<td>4.5</td>
<td>5</td>
<td>3441</td>
<td>117</td>
<td>0</td>
<td>0.9660</td>
<td>0.5975</td>
<td>0.5831</td>
<td>0.6118</td>
</tr>
</tbody>
</table>
Ontario colorectal cancer cases, age 15-99

Period relative survival, observed year 2005

<table>
<thead>
<tr>
<th>start</th>
<th>end</th>
<th>n</th>
<th>d</th>
<th>p</th>
<th>cr_e2</th>
<th>lo_cr_e2</th>
<th>hi_cr_e2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>.5</td>
<td>10358</td>
<td>1090</td>
<td>0.8440</td>
<td>0.8589</td>
<td>0.8500</td>
<td>0.8673</td>
</tr>
<tr>
<td>.5</td>
<td>1</td>
<td>9012</td>
<td>460</td>
<td>0.9237</td>
<td>0.8057</td>
<td>0.7955</td>
<td>0.8155</td>
</tr>
<tr>
<td>1</td>
<td>1.5</td>
<td>8114</td>
<td>378</td>
<td>0.9315</td>
<td>0.7634</td>
<td>0.7523</td>
<td>0.7742</td>
</tr>
<tr>
<td>1.5</td>
<td>2</td>
<td>7506</td>
<td>336</td>
<td>0.9333</td>
<td>0.7242</td>
<td>0.7123</td>
<td>0.7357</td>
</tr>
<tr>
<td>2</td>
<td>2.5</td>
<td>6789</td>
<td>234</td>
<td>0.9491</td>
<td>0.6996</td>
<td>0.6872</td>
<td>0.7117</td>
</tr>
<tr>
<td>2.5</td>
<td>3</td>
<td>6434</td>
<td>202</td>
<td>0.9530</td>
<td>0.6784</td>
<td>0.6656</td>
<td>0.6909</td>
</tr>
<tr>
<td>3</td>
<td>3.5</td>
<td>6020</td>
<td>188</td>
<td>0.9536</td>
<td>0.6596</td>
<td>0.6464</td>
<td>0.6726</td>
</tr>
<tr>
<td>3.5</td>
<td>4</td>
<td>5808</td>
<td>179</td>
<td>0.9542</td>
<td>0.6413</td>
<td>0.6278</td>
<td>0.6547</td>
</tr>
<tr>
<td>4</td>
<td>4.5</td>
<td>5503</td>
<td>137</td>
<td>0.9635</td>
<td>0.6304</td>
<td>0.6165</td>
<td>0.6441</td>
</tr>
<tr>
<td>4.5</td>
<td>5</td>
<td>5264</td>
<td>123</td>
<td>0.9648</td>
<td>0.6201</td>
<td>0.6058</td>
<td>0.6341</td>
</tr>
</tbody>
</table>
Pros and cons of different approaches

Cohort:
- Most familiar method and easiest to explain
- Have to wait years for follow-up, so results seem dated

Complete:
- Can require more work to include additional patients
- More patients means smaller standard errors for shorter times, but few patients followed for 5 years

Period:
- Uses the most recent observed data, may be closer to survival that will be experienced by those diagnosed now
- Much harder to explain!
Age-standardization of survival

- People have been age-standardizing incidence and mortality rates since 1844.
- Age-standardization enables the comparison of rates from two populations with different age structures.
- Age-standardization of survival estimates has become increasingly common since Corazziari et al published the International Cancer Survival Standards in 2004.
- Based on analyses of EUROCare-2 data, they defined the smallest number of standard populations that provide age-standardized survival estimates closest to the crude survival estimates for as many cancer sites as possible.

International Cancer Survival Standards

Three age standards were identified based on the age patterns of incidence:

- Cancers mainly of young adults
- Cancers with little variation by age
- Cancers whose incidence increases with age

From De Angelis (2009), with prostate modification
Ontario colorectal survival by age group, with 95% CI

Canadian Cancer Case Standard (Ellison, 2010)

- Each relative survival ratio is standardized WITHIN the site
  - “Age-standardized estimates were calculated using the direct method by weighting age-specific estimates for a given cancer to the age distribution of persons diagnosed with that cancer from 2001 to 2005.” (Canadian Cancer Statistics, 2014)
- Facilitates comparisons within site over time, but not between sites
  - Also prevents comparisons with other jurisdictions, since they are very unlikely to have used the CCCS!
Choosing Appropriate Life Tables for Relative Survival

Angela Mariotto
Surveillance Research Program
Division of Cancer Control and Population Sciences
National Cancer Institute

NAACCR Survival Analysis Webinar
July 10, 2014

Outline

• Life tables and relative survival
• NCHS US and State life tables
• Comparisons relative survival using NCHS US vs. NCHS State life tables
• Current efforts- Modeling life tables by SES
• Discussion/Conclusions
Life Tables

- Life table (mortality table) provide for each age the probability that a person of that age will die before his/her next birthday.
- From the probabilities of death, a number of inferences can be derived, e.g., interval survival, cumulative survival, life expectancy.
- Calculated from mortality rates.

Role of Expected Survival in Relative Survival

- Calculated as the ratio between all-cause survival divided by the expected survival of a comparable group free of cancer (or population)
- Expected survival is calculated from life tables. Also called background survival/mortality
  - Different methods to estimate (Ederer I, Ederer II, Hakulinen and Pohar Perme)
- Assumes life tables represents patients’ mortality/survival for other causes of death.
### Table 1. Life table for total population: United States, 2008

<table>
<thead>
<tr>
<th>Age (years)</th>
<th>Probability of dying between ages x and x + 1</th>
<th>Number surviving to age x</th>
<th>Number dying between ages x and x + 1</th>
<th>Person-years lived between ages x and x + 1</th>
<th>Total person-years lived above age x</th>
<th>Expectation of life at age x</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-1</td>
<td>0.006593</td>
<td>100,000</td>
<td>659</td>
<td>99,345</td>
<td>7,381,733</td>
<td>78.1</td>
</tr>
<tr>
<td>1-2</td>
<td>0.000461</td>
<td>99,341</td>
<td>46</td>
<td>99,295</td>
<td>7,212,964</td>
<td>77.6</td>
</tr>
<tr>
<td>2-3</td>
<td>0.000281</td>
<td>99,295</td>
<td>28</td>
<td>99,267</td>
<td>7,133,646</td>
<td>76.7</td>
</tr>
<tr>
<td>3-4</td>
<td>0.000219</td>
<td>99,267</td>
<td>22</td>
<td>99,245</td>
<td>7,054,265</td>
<td>75.7</td>
</tr>
<tr>
<td>4-5</td>
<td>0.000172</td>
<td>99,245</td>
<td>17</td>
<td>99,227</td>
<td>7,074,109</td>
<td>74.7</td>
</tr>
<tr>
<td>5-6</td>
<td>0.000155</td>
<td>99,228</td>
<td>15</td>
<td>99,212</td>
<td>7,015,872</td>
<td>73.7</td>
</tr>
<tr>
<td>6-7</td>
<td>0.000139</td>
<td>99,213</td>
<td>14</td>
<td>99,206</td>
<td>7,056,531</td>
<td>72.7</td>
</tr>
<tr>
<td>7-8</td>
<td>0.000126</td>
<td>99,199</td>
<td>12</td>
<td>99,181</td>
<td>7,097,444</td>
<td>71.7</td>
</tr>
<tr>
<td>8-9</td>
<td>0.000110</td>
<td>99,187</td>
<td>11</td>
<td>99,171</td>
<td>7,018,252</td>
<td>70.8</td>
</tr>
<tr>
<td>9-10</td>
<td>0.000094</td>
<td>99,176</td>
<td>9</td>
<td>99,167</td>
<td>6,919,071</td>
<td>69.8</td>
</tr>
<tr>
<td>10-11</td>
<td>0.000081</td>
<td>99,167</td>
<td>8</td>
<td>99,162</td>
<td>6,819,900</td>
<td>68.8</td>
</tr>
<tr>
<td>11-12</td>
<td>0.000087</td>
<td>99,158</td>
<td>9</td>
<td>99,154</td>
<td>6,720,738</td>
<td>67.8</td>
</tr>
<tr>
<td>12-13</td>
<td>0.000123</td>
<td>99,150</td>
<td>12</td>
<td>99,144</td>
<td>6,621,583</td>
<td>66.8</td>
</tr>
<tr>
<td>13-14</td>
<td>0.000196</td>
<td>99,138</td>
<td>19</td>
<td>99,128</td>
<td>6,522,440</td>
<td>65.8</td>
</tr>
<tr>
<td>14-15</td>
<td>0.000293</td>
<td>99,118</td>
<td>29</td>
<td>99,104</td>
<td>6,423,312</td>
<td>64.8</td>
</tr>
<tr>
<td>15-16</td>
<td>0.000395</td>
<td>99,089</td>
<td>39</td>
<td>99,070</td>
<td>6,324,208</td>
<td>63.8</td>
</tr>
</tbody>
</table>

| 95-96       | 0.232482                                      | 8,303                    | 1,930                                 | 6,373                                    | 25,780                           | 3.1                           |
| 96-97       | 0.252150                                      | 6,373                    | 1,607                                 | 4,766                                    | 18,442                           | 2.9                           |
| 97-98       | 0.272439                                      | 4,766                    | 1,298                                 | 3,467                                    | 12,873                           | 2.7                           |
| 98-99       | 0.293205                                      | 3,467                    | 1,017                                 | 2,450                                    | 8,575                            | 2.5                           |
| 99-100      | 0.314293                                      | 2,451                    | 770                                   | 1,680                                    | 5,708                            | 2.4                           |
| 100 and over| 1.000000                                      | 1,680                    | 1,680                                 | 1,680                                    | 1,680                            | 2.2                           |

Source: CDC/NCHS

\[ E_x = 1 - q_x \] is what we use to calculate relative survival
Relative Survival Biases if Life Tables Do Not Represent Patients “Expected” Survival

LT Higher 84% > 80%

Relative Survival Underestimated

75% 84% 89%

Expected

94%

All cause

75% 80%

Relative

True Expected is 80%

LT Lower 76% < 80%

Relative Survival Overestimated

75% 76% 99%
Default Life Table in SEER*Stat:
National U.S. 1970-2009 by individual year

- Years: 1970-2009
  - NCHS annual 2001-2009 US life tables
  - Interpolation methods between missing years
- Race: White, Black and Other (AI/API), All Races used for unknown race
  - Standard methods to estimate life tables for other races
- Ages: Maximum age is 99 (previously 119)
  - http://seer.cancer.gov/expsurvival/

Relative Survival from National Life Tables

- Accurate for populations with similar background mortality as the US population, e.g. SEER.
- Not accurate when the background mortality of the study cohort is different than the US. Examples:
  - Special populations: small race/ethnicity groups, by socio-economic status, by comorbidity, by geographic areas
  - May mislead comparisons by registries
  - Cancers with common risk factors with other diseases: smoking and lung cancer
  - Healthy screening effect (early stage breast and prostate cancer)
NCHS State Life Tables

  - Interpolation methods between missing years
  - Available by race (white and black) and sex where the numbers of deaths were sufficient to produce reliable estimates.
  - Blacks not available in 11 states
- Ages: Maximum age 119
- Available at request in SEER*Stat. Blacks US life tables are used when NCHS state life tables are missing

NCHS US vs. State Life tables
2000 White Males

[Graph showing log probability of dying vs. age for different states compared to the US]
Impact of US and State Life Tables on Relative Survival

- Compared 5-year relative survival using US and State life tables for individuals diagnosed with female breast, colorectal, prostate cancers and all cancer sites combined in the SEER-18 areas.
- Compared by cancer site, age, race, and area.

Stroup et al. The Impact of State-Specific Life Tables on Relative Survival. JNCI monograph (in review)
To help interpretation

• Ordered SEER registries/States according to lower → higher mortality

<table>
<thead>
<tr>
<th>Registry</th>
<th>Age-Adjusted Mortality *</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hawaii</td>
<td>615</td>
</tr>
<tr>
<td>Connecticut</td>
<td>669</td>
</tr>
<tr>
<td>California</td>
<td>674</td>
</tr>
<tr>
<td>Seattle</td>
<td>699</td>
</tr>
<tr>
<td>Utah</td>
<td>709</td>
</tr>
<tr>
<td>New Jersey</td>
<td>710</td>
</tr>
<tr>
<td>Iowa</td>
<td>726</td>
</tr>
<tr>
<td>US</td>
<td>764</td>
</tr>
<tr>
<td>New Mexico</td>
<td>770</td>
</tr>
<tr>
<td>Detroit</td>
<td>829</td>
</tr>
<tr>
<td>Georgia</td>
<td>862</td>
</tr>
<tr>
<td>Kentucky</td>
<td>919</td>
</tr>
<tr>
<td>Louisiana</td>
<td>929</td>
</tr>
</tbody>
</table>

US LT Expected Survival < State Expected
US LT Relative Survival (RS)> State LT RS

US LT Expected Survival > State Expected
US LT Relative Survival (RS)< State LT RS

colorectal whites ages 0-79

US LT RS                      State LT RS

All Cancer Sites Whites Ages 0-79

All Cancer Sites Blacks Ages 0-79

Colorectal Whites Ages 0-79

Colorectal Blacks Ages 0-79
5-year Relative Survival and Confidence Intervals
Black Patients Diagnosed with Colorectal Cancer in 2000-2009

Prostate Whites Ages 0-79
Prostate Blacks Ages 0-79
Female Breast Whites Ages 0-79
Female Breast Blacks Ages 0-79
Comparisons for Whites Ages 85+

- Probabilities of dying for older ages lower in State life tables compared to US life table
- Relative survival based on State LT systematically underestimated

What have we learned?

- For younger ages < 80 relative survival using state life tables more accurate because it accounts for some state variability in background mortality.
- State life tables biased for older ages and should not be used
- Differences in relative survival due to different life tables in general very small
What have we learned? (2)

- Relative survival differences between registries for the black population can be quite large due to small black populations in some registries.
- Example: colorectal cancer in Utah and Hawaii.
- Important to report variability of relative survival in comparisons, e.g., confidence intervals.

How can we improve State Life Tables

- State Life Tables better than US Life Tables for younger ages.
- Corrections for older ages:
  - We could use US life tables for ages 80+
  - or
  - Modeling……using Poisson regression models
Current Efforts

- Poisson regression model using age as a spline function.
- On going effort in collaboration with Bin Huang and Hannah Weir.
- Uses mortality and socio-economic status data at county level
  - Socio-economic status (SES) is measured as Yost composite index (Mandi et al, Cancer Causes and Control, 2014)
- Model by sex, race and geography area (to be defined) and use spline of age and SES as covariate
- Investigate excluding cancer death as cause of death
Discussion/Conclusions

• There is an advantage in using NCHS life tables
  • Has the “imprimatur” of NCHS
  • Well documented and available to all
• Modeling is an alternative however it requires:
  • Validation and method to identify the “best” life table for estimating relative survival
  • Balance between modeling all important factors (race, sex, geography, SES) and allowing for enough data so that estimates are reliable

Acknowledgements

Bin Huang (Kentucky CR)
Nan Stroup (New Jersey CR)
Hannah Weir (CDC)
Steve Scoppa, Jou Zou (IMS)
Hyunsoon Cho (NCC Korea)
Thank you

Questions

mariotta@mail.nih.gov
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
• Conclusions

Background

• 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

Background (Cont’d)

• 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
-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>HIV disease with other malignant neoplasms</td>
</tr>
<tr>
<td>HIV alone*</td>
<td>HIV disease with infectious diseases</td>
</tr>
<tr>
<td>Site –specific disease</td>
<td>Mouth disease</td>
</tr>
</tbody>
</table>

*Only for HIV/AIDS associated cancers

Cases that were diagnosed with only one cancer (sequence 00)

![Chart showing the percentage of cases per COD category](chart.png)

- Breast
- Oral cavity & pharynx

*Other cause is treated as censored observation

National Cancer Institute
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 with Non-Hodgkin Lymphoma (Sequence 00)

Note: Five-year Non-Hodgkin lymphoma cancer survival by age at diagnosis, SEER-13, 1992-2004

National Cancer Institute
Cases with Non-Hodgkin Lymphoma (Sequence 00)

- 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 1st diagnosis were attributed
  - COD due to 2nd cancer were treated as censored
- COD for AIDS defining cancer
  - AIDS and cancer
  - HIV alone
Overview of COD Algorithm

<table>
<thead>
<tr>
<th>Cause of death groups</th>
<th>Death attributed to the specific cancer site</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sequence 00 - one &amp; only one primary</td>
</tr>
<tr>
<td>Cancer of the same site</td>
<td>Yes</td>
</tr>
<tr>
<td>Cancer of the same body system</td>
<td>Yes</td>
</tr>
<tr>
<td>Cancer of any other site</td>
<td>Yes</td>
</tr>
<tr>
<td>AIDS and cancer (B21)</td>
<td>Yes</td>
</tr>
<tr>
<td>HIV alone (B20)</td>
<td>HIV/AIDS associated cancers*</td>
</tr>
<tr>
<td>Site –specific disease</td>
<td>Selective</td>
</tr>
</tbody>
</table>

*HIV/AIDS associated cancers= Oral Cavity and Pharynx, Cervix, Anus Cancer, and Lymphomas and Kaposi Sarcoma

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
  (↓) Relative = Observed/Expected (↑)

• If expected rate is underestimated then
  (↑) Relative = Observed/Expected (↓)

*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></th>
<th>AI/AN</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Relative</td>
<td>C-S</td>
<td>Diff</td>
<td>Relative</td>
</tr>
<tr>
<td>Breast, In-situ &amp; 65+</td>
<td>107.5</td>
<td>98.6</td>
<td>8.9</td>
<td>95.8</td>
</tr>
<tr>
<td>Prostate, L/R &amp; 65+</td>
<td>104.5</td>
<td>94.8</td>
<td>9.8</td>
<td>87.4</td>
</tr>
<tr>
<td>Lung, All Stage &amp; &lt;65</td>
<td>18.7</td>
<td>20.5</td>
<td>-1.8</td>
<td>16.7</td>
</tr>
<tr>
<td>Oral Cavity, All Stage &amp; &lt; 65</td>
<td>67.2</td>
<td>71.6</td>
<td>-4.4</td>
<td>51.6</td>
</tr>
</tbody>
</table>

Note: AI/AN = American Indian/Alaska Native; C-S = Cause-specific; L/R=Localized/Regional
# Example of Relative Survival Being Problematic: SEER-13, 1992-2004

<table>
<thead>
<tr>
<th></th>
<th>White</th>
<th></th>
<th></th>
<th>AI/AN</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Relative</td>
<td>C-S</td>
<td>Diff</td>
<td>Relative</td>
<td>C-S</td>
<td>Diff</td>
</tr>
<tr>
<td>Breast, In-situ &amp; 65+</td>
<td>107.5</td>
<td>98.6</td>
<td>8.9</td>
<td>95.8</td>
<td>99.0</td>
<td>-3.2</td>
</tr>
<tr>
<td>Prostate, L/R &amp; 65+</td>
<td>104.5</td>
<td>94.8</td>
<td>9.8</td>
<td>87.4</td>
<td>91.3</td>
<td>-3.8</td>
</tr>
<tr>
<td>Lung, All Stage &amp; &lt;65</td>
<td>18.7</td>
<td>20.5</td>
<td>-1.8</td>
<td>16.7</td>
<td>19.7</td>
<td>-3.1</td>
</tr>
<tr>
<td>Oral Cavity, All Stage &amp; &lt; 65</td>
<td>67.2</td>
<td>71.6</td>
<td>-4.4</td>
<td>51.6</td>
<td>58.0</td>
<td>-6.4</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>
</tbody>
</table>

Difference between High and Low SES: 13.3


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


• http://seer.cancer.gov/causespecific/

• Contact information: howladern@mail.nih.gov

Thank you!
IN THIS PRESENTATION

- Whites (non-Hispanic)
- Blacks (non-Hispanic)
- Asians (non-Hispanic)
- Hispanics

POPULATION GROUPS IN THE US 2012

Whites 54%
Hispanics 16%
Blacks 13%
Asians 0%
Native Americans <1%

Source: http://www.census.gov
WE NEED BETTER CANCER INDICATORS FOR:

- HISPANICS
- ASIANS
- And that includes survival!

TWO TYPES OF FOLLOW-UP IN US REGISTRIES: PASSIVE VS. ACTIVE FOLLOW-UP

<table>
<thead>
<tr>
<th>PASSIVE FOLLOW-UP</th>
<th>“ACTIVE” FOLLOW-UP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ascertainment of deaths</td>
<td>idem + Last alive contact</td>
</tr>
<tr>
<td>Presumed Alive</td>
<td>Reported Alive</td>
</tr>
<tr>
<td>NPCR</td>
<td>SEER</td>
</tr>
<tr>
<td>Mary Pumpkin Diagnosed in 2005, not matched in NDI 2010:</td>
<td>Mary Pumpkin Diagnosed in 2005, not matched in NDI 2010, last known as alive in 2007:</td>
</tr>
<tr>
<td>Survival time = 5 years</td>
<td>Survival time = 2 years</td>
</tr>
<tr>
<td>The “Lost to follow-up” are presumed as alive</td>
<td>The “Lost to follow-up” and their length of time under observation, e.g. be it 1 month, 4 years or 10 years, is taken into account</td>
</tr>
</tbody>
</table>
ACTIVE VS PASSIVE FOLLOW-UP

• Does it make any difference when calculating 5-year survival rates (active vs passive follow-up)?

**5-YEAR SURVIVAL PROPORTIONS WITH ACTIVE AND PASSIVE FOLLOW-UP. SEER-17, 2000-2008. LAST FUP 2009**

<table>
<thead>
<tr>
<th>Site</th>
<th>Race/ethnicity</th>
<th>Passive follow-up</th>
<th>Active follow-up</th>
<th>Survival difference (in percent points)</th>
</tr>
</thead>
<tbody>
<tr>
<td>colorectal</td>
<td>Whites</td>
<td>56.2 (56.0-56.5)</td>
<td>56.0 (55.7-56.2)</td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td>Blacks</td>
<td>46.7 (46.6-46.8)</td>
<td>46.2 (45.9-46.5)</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>Hispanics</td>
<td>54.1 (53.9-54.4)</td>
<td>54.0 (53.7-54.4)</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td>Asians</td>
<td>60.9 (60.6-61.0)</td>
<td>60.8 (60.5-61.0)</td>
<td>1.1</td>
</tr>
<tr>
<td>female breast</td>
<td>Whites</td>
<td>78.7 (78.7-78.8)</td>
<td>78.7 (78.5-78.8)</td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td>Blacks</td>
<td>65.1 (65.0-65.4)</td>
<td>65.0 (64.8-65.4)</td>
<td>0.3</td>
</tr>
<tr>
<td></td>
<td>Hispanics</td>
<td>76.8 (76.1-77.6)</td>
<td>76.5 (75.7-77.4)</td>
<td>0.9</td>
</tr>
<tr>
<td></td>
<td>Asians</td>
<td>82.4 (81.7-83.1)</td>
<td>81.9 (81.4-82.4)</td>
<td>0.5</td>
</tr>
<tr>
<td>lung</td>
<td>Whites</td>
<td>14.9 (14.7-15.1)</td>
<td>14.7 (14.5-14.9)</td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td>Blacks</td>
<td>10.9 (10.5-11.2)</td>
<td>10.6 (10.3-10.9)</td>
<td>0.3</td>
</tr>
<tr>
<td></td>
<td>Hispanics</td>
<td>14.1 (13.7-14.5)</td>
<td>14.0 (13.6-14.4)</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>Asians</td>
<td>18.1 (17.4-18.8)</td>
<td>18.0 (17.4-18.5)</td>
<td>0.2</td>
</tr>
<tr>
<td>prostate</td>
<td>Whites</td>
<td>84.3 (83.8-84.9)</td>
<td>84.3 (83.8-84.9)</td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td>Blacks</td>
<td>75.5 (75.2-76.0)</td>
<td>75.2 (74.8-75.6)</td>
<td>0.3</td>
</tr>
<tr>
<td></td>
<td>Hispanics</td>
<td>82.8 (82.2-83.5)</td>
<td>82.6 (82.2-83.2)</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td>Asians</td>
<td>85.3 (84.8-85.8)</td>
<td>84.8 (84.4-85.4)</td>
<td>0.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Site</th>
<th>Race/ethnicity</th>
<th>Passive follow-up</th>
<th>Active follow-up</th>
<th>Survival difference (in percent points)</th>
</tr>
</thead>
<tbody>
<tr>
<td>gall bladder</td>
<td>Whites</td>
<td>15.5 (15.4-15.6)</td>
<td>15.4 (15.3-15.5)</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>Blacks</td>
<td>11.6 (11.4-11.8)</td>
<td>11.5 (11.3-11.7)</td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td>Hispanics</td>
<td>10.7 (10.5-10.9)</td>
<td>10.7 (10.5-10.9)</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>Asians</td>
<td>21.3 (21.0-21.6)</td>
<td>21.3 (21.0-21.6)</td>
<td>0.2</td>
</tr>
<tr>
<td>liver</td>
<td>Whites</td>
<td>11.7 (11.5-11.9)</td>
<td>11.6 (11.4-11.8)</td>
<td>0.3</td>
</tr>
<tr>
<td></td>
<td>Blacks</td>
<td>7.1 (6.9-7.3)</td>
<td>7.0 (6.8-7.2)</td>
<td>0.3</td>
</tr>
<tr>
<td></td>
<td>Hispanics</td>
<td>15.2 (14.9-15.5)</td>
<td>15.1 (14.8-15.4)</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>Asians</td>
<td>20.4 (20.1-20.7)</td>
<td>20.1 (19.8-20.4)</td>
<td>0.3</td>
</tr>
<tr>
<td>pancreas</td>
<td>Whites</td>
<td>6.1 (5.9-6.4)</td>
<td>5.9 (5.7-6.1)</td>
<td>0.3</td>
</tr>
<tr>
<td></td>
<td>Blacks</td>
<td>3.1 (2.8-3.4)</td>
<td>3.0 (2.7-3.3)</td>
<td>0.3</td>
</tr>
<tr>
<td></td>
<td>Hispanics</td>
<td>8.6 (8.3-8.9)</td>
<td>8.5 (8.3-8.8)</td>
<td>0.3</td>
</tr>
<tr>
<td></td>
<td>Asians</td>
<td>18.5 (18.2-18.9)</td>
<td>18.3 (18.0-18.7)</td>
<td>0.2</td>
</tr>
<tr>
<td>stomach</td>
<td>Whites</td>
<td>22.9 (22.6-23.3)</td>
<td>22.8 (22.5-23.1)</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>Blacks</td>
<td>21.2 (20.9-21.5)</td>
<td>21.1 (20.8-21.4)</td>
<td>0.4</td>
</tr>
<tr>
<td></td>
<td>Hispanics</td>
<td>29.6 (29.3-30.0)</td>
<td>29.4 (29.1-29.7)</td>
<td>0.3</td>
</tr>
<tr>
<td></td>
<td>Asians</td>
<td>42.7 (42.4-43.1)</td>
<td>42.5 (42.2-42.8)</td>
<td>0.3</td>
</tr>
</tbody>
</table>

* denotes statistically significant difference
FINDINGS

• 1. Larger/Significant differences between the two types of follow-up for Hispanics and Asians

• 2. Larger differences for cancers of poor prognosis

PROPORTION OF LOST TO FOLLOW-UP BY TIME AFTER DIAGNOSIS. LUNG CANCER 2000-2003. SEER-17
FINDING

- Hispanics and Asians are much more likely to be lost to follow-up than Whites or Blacks

ARE THOSE WHO ARE LOST TO FOLLOW-UP THE SAME AS THOSE WITH COMPLETE FOLLOW-UP?

<table>
<thead>
<tr>
<th>Population</th>
<th>SEER stage</th>
<th>N</th>
<th>Hazard Ratio*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whites</td>
<td>localized/regional</td>
<td>462,414</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>distant</td>
<td>109,965</td>
<td>0.99 (0.86-1.14)</td>
</tr>
<tr>
<td></td>
<td>unstaged/unknown</td>
<td>29,256</td>
<td></td>
</tr>
<tr>
<td>Blacks</td>
<td>localized/regional</td>
<td>63,023</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>distant</td>
<td>17,954</td>
<td>1.36 (1.06-1.75)</td>
</tr>
<tr>
<td></td>
<td>unstaged/unknown</td>
<td>4,386</td>
<td></td>
</tr>
<tr>
<td>Hispanics</td>
<td>localized/regional</td>
<td>40,815</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>distant</td>
<td>8,878</td>
<td>1.69 (1.45-1.96)</td>
</tr>
<tr>
<td></td>
<td>unstaged/unknown</td>
<td>2,531</td>
<td></td>
</tr>
<tr>
<td>Asians</td>
<td>localized/regional</td>
<td>29,674</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>distant</td>
<td>7,383</td>
<td>1.77 (1.48-2.11)</td>
</tr>
<tr>
<td></td>
<td>unstaged/unknown</td>
<td>1,362</td>
<td></td>
</tr>
</tbody>
</table>

*Adjusted for Age, Year of diagnosis, SEER Registry, Gender, and Cancer Site
FINDING

- Hispanics and Asians who are lost to follow-up are more likely to have bad prognoses
- This will bias our survival statistics

ESTIMATED PROPORTION OF MISSED DEATHS

<table>
<thead>
<tr>
<th>POPULATION</th>
<th>COUNTRY OF ORIGIN</th>
<th>Missed deaths / Likely deaths</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>WHITES</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1% (0.11-0.15)</td>
<td></td>
</tr>
<tr>
<td>BLACKS</td>
<td></td>
<td>0.3% (0.26-0.46)</td>
<td></td>
</tr>
<tr>
<td>HISPANICS</td>
<td></td>
<td>2.8% (2.41-3.21)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mexican</td>
<td>4.7% (3.83-5.81)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Puerto Rican</td>
<td>0.3% (0.05-1.49)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cuban</td>
<td>1.3% (0.43-3.67)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Central and South American</td>
<td>9.4% (7.24-12.12)</td>
<td></td>
</tr>
<tr>
<td>ASIANS</td>
<td>Chinese</td>
<td>2.6% (2.18-2.98)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Japanese</td>
<td>0.3% (0.10-0.89)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Filipino</td>
<td>4.2% (3.25-5.30)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hawaiian</td>
<td>0.3% (0.05-1.51)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Korean</td>
<td>1.3% (0.59-2.77)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vietnamese</td>
<td>2.2% (1.28-3.71)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>South Asian (India, Pakistan, Sri Lanka)</td>
<td>6.4% (3.78-10.64)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>all other Asians</td>
<td>3.2% (2.07-4.87)</td>
<td></td>
</tr>
</tbody>
</table>
WE ARE NOT DETECTING 100% OF DEATHS

- Why?
  - cause they go back to their countries to die? (salmon bias)
  - cause they came for treatment only, followed by a quick return home, and are recorded as US de-facto residents?
  - names? Changed, complicated, unfamiliar, increased likelihood of typos
  - SSN? About 11 million fake SSNs being used (9 million Hispanics//1.3 millions Asians)

Sources and Refs: [www.census.gov](http://www.census.gov); Federal Trade Commission Report on SSNs and ID theft; Pew Hispanic Center

HISPANIC NAMES

- Juan (1) Manuel (2) Garcia (3) Perez (4)
- Garcia=father’s surname
- Perez=mother’s surname
- #3 is the true last name as in US standards
When in the US:
Some will have #3 as last name
Some will have #4 as last name
Some will have #3 and #4 together as a single name, others as a double-barrelled name
WHAT IS THE TAKE HOME MESSAGE?

For Asians and Hispanics:

- Death Linkages are somewhat deficient:
  - Reported Alive (SEER) is somewhat deficient
  - Presumed Alive (NPCR) is more deficient
- It is a **data quality issue**, it is **problem of the Foreign-born**
- It is not a methods issue

IS THIS CURRENTLY A PROBLEM?

**Cancer Registries**

- Very small problem

  - SURVIVAL RATES
  - RARELY ON HISPANICS AND ASIANS (no life tables)
  - OVERESTIMATION OF RATES AMONG WHITES (TX, FL, CA states with a large proportion of white latinos)

**Outside Researchers**

- Substantial problem

  - MULTIVARIATE ANALYSES
  - LARGE STATES:
    - FL, TX (PASSIVE follow-up) biased Hispanic advantages
    - CA (ACTIVE follow-up), potential problem to a lesser degree
  - NY (no reports on survival among Hispanics and Asians)
NEED A PUBLICATION? PRESS COVERAGE? TRY THE HISPANIC PARADOX

WHAT CAN REGISTRIES DO? CURRENTLY NOT MUCH 😞

- [DATA RELEASE TO RESEARCHERS]
  1. Recommend caution to outside researchers when analyzing survival for Hispanics and Asians
  2. Avoid use of the variable birthplace for all survival calculations in registry or in datasets for outside researchers
- [NDI LINKAGES]
  3. If names or dates are problematic then submit multiple name sequences (or dates) possibilities for the NDI linkage, e.g. Cristina Garcia-Perez, submit Cristina Garcia, Cristina Perez, Cristina Garcia-Perez
- [SSDI LINKAGES]
  4. Hopefully available in the near future, this linkage will inform on "invalid SSNs" and make possible more efforts based on name and DOB with the NDI linkages for these records
ANALYSIS – RULE OF THUMB

- 5- Calculate survival rates for distant stage lung cancer, if the differences between races are significant and very different between races, then that will point toward a problem with the death detection in your registry data

- Thank you!

- 702 - 895 5717
- paulo.pinheiro@unlv.edu
9. Active follow-up versus presumed alive survival calculations – which is appropriate for your data?

Chris Johnson
Cancer Data Registry of Idaho
Co-Chair, NAACCR Survival Analysis Task Force

Follow-up procedures vary among cancer registries in North America

- SEER registries ascertain vital status and date of last contact to meet follow-up standards.
- NPCR and Canadian registries primarily conduct linkages with local and national death records to ascertain deaths.
Active Follow-Up vs. Passive Follow-Up

- To calculate survival statistics, we are chiefly interested in obtaining follow-up information on vital status (alive or dead) and date of last contact or death.
- Active follow-up is “any activity which involves direct contact with the patient, the patient’s family, or the patient’s physician, in order to encourage contact between the patient and the health provider.”
- “Passive follow-up refers to methods which do not require contact with hospitals, physicians, or patients.”
  - Central Cancer Registries: Design, Management, and Use (1994)

Passive Follow-Up

- Passive methods are used mostly to determine the vital status of the patient and a more current date last seen alive, or date of death.
- Sources of information/linkages with:
  - State and national death files
  - Motor vehicles records
  - Hospital discharge data
  - Health care claims data
  - Voter registration
SEER Follow-Up Calculation
Feb 2013 Submission

The last year of data being submitted is 2011. The percent of patients diagnosed during the years prior to 2011 who have current follow-up is defined as

\[ P = 100(D + A)/T \]

where D is the number who died prior to January 1, 2012, A is the number with follow-up dates on or after January 1, 2012 (includes both alive and dead patients), and T is equal to A + D + the number of patients who were last known to be alive with follow-up dates prior to January 1, 2012. P can be calculated for individual years of diagnosis up through 2010 and for all years combined prior to 2011.

For all invasive cancers and calendar years specified by NCI, P shall be at least 95 percent, but must not be below 90 percent.

For patients ages 20-64, and under age 20, P shall be at least 90 percent in each case, but must not be below 80 percent.

---

Active Follow-Up vs. “Active” Follow-Up

- What is in a name?

- “Reported Alive” versus “Presumed Alive”
Presumed Alive

- Because not all central cancer registries conduct active patient follow-up, it is necessary to have an option for calculating survival times based on the assumption that the registry has ascertained all available deaths, and persons not known to be deceased are presumed to be alive as of the last date for which complete death ascertainment is available.

Follow-Up Time Under Active Follow-Up vs. Ascertainment of Deaths Only

<table>
<thead>
<tr>
<th>Patient 1 - Active FUP</th>
<th>Patient 2 - Active FUP</th>
<th>Patient 3 - Active FUP</th>
<th>Patient 4 - Active FUP</th>
<th>Patient 5 - Active FUP</th>
<th>Patient 1 - Deaths only</th>
<th>Patient 2 - Deaths only</th>
<th>Patient 3 - Deaths only</th>
<th>Patient 4 - Deaths only</th>
<th>Patient 5 - Deaths only</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan-07</td>
<td>Jan-12</td>
<td>Jan-07</td>
<td>Jan-12</td>
<td>Jan-07</td>
<td>Jan-12</td>
<td>Jan-07</td>
<td>Jan-12</td>
<td>Jan-07</td>
<td>Jan-12</td>
</tr>
</tbody>
</table>

7/15/2014
**The Importance of Death Ascertainment**


**OBJECTIVE:** designed to measure the impact of variation in patient follow-up on survival statistics.

**METHODS:** SEER data used to construct datasets simulated scenarios of complete (SEER), incomplete, and no follow-up (NPCR) of alive patients; and complete and incomplete death ascertainment.

**CONCLUSIONS:**
- Complete death ascertainment important for producing accurate cancer survival statistics, and
- Ascertainment of deaths only should generally be sufficient for survival analysis.
Follow-Up Requirements

**Alive Status**
- SEER Program requires all SEER registries to follow alive patients
  - 95% patients (minimum of 90%) have current follow-up
- NPCR registries are not required to follow patients
  - impute follow-up date to be the end of study (e.g., Dec 31 2011)

**Death Status**
- All Registries conduct death clearance with state DC
- SEER and NPCR provide support for registries to link with the National Death Index and the Social Security Death Index

Reported Alive or Presumed Alive?

- The “at risk” interval for living patients should be based on the best available information.
  - Use **reported alive** method if data meet SEER standards for follow-up.
  - Otherwise, use **presumed alive** method.
  - Reports of cancer survival should clearly indicate whether the results are based on the reported alive method, the presumed alive method, or a combination of both.
10. Analyzing your own registry data using SEER*Stat

Chris Johnson
Cancer Data Registry of Idaho
Co-Chair, NAACCR Survival Analysis Task Force
Proposed Model for Preparing Registry Data for Survival Analysis - Overview

1. Start with your NAACCR Call for Data file (that includes day components of date, cause of death, and county of residence at dx).

2. Process it with SEER Data Viewer software to create new fields for survival (pre-calculated survival fields based on days, SEER Cause-specific death classification and SEER Other cause of death classification).

3. Run the new file through SEERPrep using latest .dd file.

4. Analyze with SEERStat.

Step 1 – Start with your NAACCR Call for Data file

- Where is that pesky file, anyway?
- If your CFD data file doesn’t already include day components of dates, cause of death, and county of residence at dx, make one that does.

**IMPORTANT** – make certain the your data file is sorted by Addr at DX--state

   Patient ID
   Sequence number—central

so that survival time variables will be calculated correctly and person selection features will be available in SEER*Stat.
Step 2. Process your Data File with SEER Data Viewer software

- Creates new fields for survival:
  - 7 pre-calculated survival fields based on days
  - SEER Cause-specific death classification
  - SEER Other cause of death classification

- In future version, we anticipate that County at Diagnosis will be used to match to life tables for relative survival

SEER Survival Variable Resources

- 7 pre-calculated survival fields based on days
  - Additional information about the algorithm and what specific values are assigned in given missing date situations are available here: http://seer.cancer.gov/survivaltime/.

- SEER Cause-specific Death Classification
SEER Data Viewer

Version 1.4 released May 10, 2014

The SEER Data Viewer can be used to view and manage data that are stored in text files. It is primarily designed for managing cancer incidence data files formatted according to the NAACCR Data Specifications (Volume II). The file layouts and coding manuals for NAACCR versions 1.2 and 1.3 are integrated into the system. This software can also be used to process data files containing NAACCR data items in comma-separated value (CSV) text files, and the system can be configured so that it can be used with any data stored in a fixed column or CSV file format.

The SEER Data Viewer is easy to use:
- Anyone can open a file and view data in a table (no programming required)
- You can select fields by NAACCR item number or field name. You do not need to know the column location.
- You can use the SEER Data Viewer to:
  - View records in a table
  - Calculate derived fields and create a new data file
  - Provide other staff with saved searches or saved template algorithms
  - Fix common problems like line length, end of line char, etc.
  - Explore the data using SQL

For more information, see the SEER Data Viewer documentation on the SEER Data Viewer website.
SEER Data Viewer Instructions

- File->Preferences and select Survival Time on the left and specify your study-cut off year (by default it would be current year – 22 months or 2012 currently).
  - For Nov 2013 CFD, study cutoff year was 2011 – the last year you could have ascertained all deaths for.
- To have the data viewer add the survival time fields and SEER cause-specific and other cause of death classifications:
  - Select a file to view/process (needs to be NAACCR 13)
  - Go to Output Options and select the option to “Create a copy of the input file” and specify a target name
  - Click the Process Data File Now button
Step 3. Run the New File through SEER*Prep

- Run the new file through SEER*Prep using latest.dd file.
What is SEER*Prep?

- SEER*Prep software converts ASCII text data files to the SEER*Stat database format, allowing you to analyze your cancer data using SEER*Stat.
- SEER*Prep performs two main functions:
  - it converts text data to the specific binary format required by SEER*Stat,
  - it creates the SEER*Stat data dictionary.
NAACC R 2013-2014 Webinar Series
Step 4. Analyze Your Data with SEER*Stat

What is SEER*Stat?

- SEER*Stat is a statistical package created for the analysis of SEER and other cancer databases.
- It was developed by Information Management Services, Inc. in consultation with the SEER Program of the National Cancer Institute (NCI).
- The SEER*Stat statistical software provides a convenient, intuitive mechanism for the analysis of SEER and other cancer-related databases.
- It is a powerful PC tool to view individual cancer records and to produce statistics for studying the impact of cancer on a population.
Overview of SEER*Stat

- SEER*Stat allows you a great deal of freedom to request the cancer statistics/values/methods you want for your analysis.

- Part 1: Session
- Part 2: Execute
- Part 3: Matrix
Overview of SEER*Stat

Part 1: Session

- The analysis is set up in the session window. Each session consists of tabs on which you select the database subset, statistics, and appearance of your output matrix.
- You should work through each tab in order from left to right and from top to bottom to ensure that all options have been considered.
  - However, changes can be made in any order.
  - It is possible to work on multiple sessions simultaneously.

Part 2: Execute

- Once the session is set up, you are ready to execute it as a job.
- While the job is executing, you can change the session or begin a new one without affecting the original job.
- It is possible to execute more than one job at a time.
Overview of SEER*Stat

- Part 3: Matrix

- When the job has finished executing, the output matrix you requested is displayed.

- You can change the appearance of the output matrix, print it, copy it to the Windows clipboard, and/or export the statistics/values so they may be used in another application.
Conclusions: SEER*Stat

- Advantages of SEER*Stat over other statistical tools:
  - Simple to use GUI
  - Facilitates comparisons with SEER data
  - Can paste results into other Windows programs
  - SEER/NCI is responsible for keeping it updated and standardized
  - Well supported by IMS
Proposed Model for Preparing Registry Data for Survival Analysis - Overview

1. Start with your NAACCR Call for Data file (that includes day components of date, cause of death, and county of residence at dx).
2. Process it with SEER Data Viewer software to create new fields for survival (pre-calculated survival fields based on days, SEER Cause-specific death classification and SEER Other cause of death classification).
3. Run the new file through SEERPrep using latest .dd file.
4. Analyze with SEERStat.

Questions?
Coming Up…

- Collecting Cancer Data: Lung
  - August 7, 2014
- Coding Pitfalls
  - September 11, 2014
- Registration is open for 2014-2015 Cancer Registry & Surveillance Webinar Series

And the winners are........
CE Certificate Quiz/Survey

- Phrase
- Link

Thank You!!!!

Please send any questions to:
Jim Hofferkamp [jhofferkamp@naaccr.org](mailto:jhofferkamp@naaccr.org)
Shannon Vann [svann@naaccr.org](mailto:svann@naaccr.org)