An Overview of Survival Statistics in SEER*Stat

Angela Mariotto, NCI
Nadia Howlader, NCI
Steve Scoppa, IMS
Don Green, IMS

NCI Analytic Tools SEERies

October 28, 2021
1. Overview of survival
2. Relative Survival Rates
   - Period Survival
   - Conditional Survival
3. Demos of relative survival and the period method
4. Cause-specific Survival
5. Crude survival (Crude probabilities of death)
6. Demo of cause-specific survival
## What Makes a Survival Analysis Different

<table>
<thead>
<tr>
<th>SURVIVAL</th>
<th>INCIDENCE/MORTALITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>It is the study of:</td>
<td>Occurrence of one event</td>
</tr>
<tr>
<td>Time between 2 events (e.g. diagnosis and death)</td>
<td></td>
</tr>
<tr>
<td>Universe/denominator</td>
<td>The whole population</td>
</tr>
<tr>
<td>Cohort of cancer patients</td>
<td></td>
</tr>
<tr>
<td>It is usually reported as</td>
<td>Rates per 100,000 population</td>
</tr>
<tr>
<td>Proportion or percent of patients surviving/dying at a given time</td>
<td></td>
</tr>
<tr>
<td>Censoring</td>
<td></td>
</tr>
<tr>
<td>Event cannot be observed for some patients (e.g. lost to follow-up, occurrence of a competing event, end of study)</td>
<td>-</td>
</tr>
</tbody>
</table>
Survival is not a one size fits all statistic
Different statistics to answer different questions

- **Patient 1:** I have just been diagnosed with ovarian cancer. What are my chances of surviving this cancer?
- **Patient 2:** I have cardiovascular disease and have been diagnosed with localized breast cancer, what are my chances of dying of breast cancer in the next 5 years?
- **Science Writer:** How has survival of prostate cancer changed over time? How do you expect it to change in the future?
- **Congressperson:** What is the most recent estimate of 5-year survival for breast cancer? How does it differ by race/ethnicity?
- **Cancer Survivor:** I have survived five years after diagnosis with colorectal cancer. What is the possibility that I am cured? What are my chances of not dying of cancer in the next 5 years?
- **Researcher:** Do cancer patients have higher risks of death for other causes (in the absence of cancer death) compared to the general population?
Competing events and censoring

- 32,216 colorectal cancer patients diagnosed in 2010-2015 at ages 75-84 years. At the end of the first year at diagnosis
  - 6,538 die of cancer \(\rightarrow\) event of interest
  - 2,072 die of other causes \(\rightarrow\) competing event
  - 599 are lost to follow-up \(\rightarrow\) censored
- Depending on how we deal with competing events we can estimate:
  - **Net cancer survival** (eliminate risks of competing events)
  - **Crude survival** (include risks of competing events)
SEER*Stat Survival Measures-Statistic TAB

- **Observed Survival**
  - Probability of surviving (any causes of death)

- **Net Survival**
  - Probability of surviving cancer in the absence of other causes of death

- **Crude Probability of Death**
  - Probability of dying of cancer and other causes and surviving
Net Survival

*Probability of surviving cancer in the absence of other causes of death*
How might we measure the prognosis of cancer patients?

- Total mortality (among cancer patients)
  - All cause survival or observed survival (event is death)
- Interest has been typically in survival associated with a diagnosis of cancer (not affected by the chances of dying of other causes) → Net survival
  - Relative survival: standard cancer registry method that does not use cause of death information
  - Cancer-specific survival: uses cause of death information
- It is important to note that net survival is interpreted in a hypothetical world where competing risks are assumed to be eliminated
Cause of death information

- Patients are enrolled at the registry after being diagnosed with cancer and followed-up
- Unlike a clinical trial (detailed review of the medical record to ascertain the cause of death), registries depend on death certificates to obtain cause of death information
- Cause of death information obtained from the death certificates may not be reliable (misclassification errors) or may not be available for some of the registries → Relative survival
Relative survival (Net survival)

- Standard method of estimating net cancer survival from cancer registry data
- Does not use cause of death information (cause of death is usually not available or unreliable)
- Relative survival is the ratio of observed survival in the patient group divided by the expected survival of a comparable group from the general population.

\[
\text{relative survival ratio} = \frac{\text{observed survival proportion}}{\text{expected survival proportion}}
\]

- Measure of excess mortality experienced by cancer patients

\[
\text{excess mortality} = \text{observed mortality} - \text{expected mortality}
\]
How do we estimate expected survival from comparable group?

- Usually from nationwide (or statewide) population life tables stratified by age, sex, calendar time, and race.
- Life tables are matched by age, sex, calendar year, and race (and geography if applicable) to each cancer patient in the cohort.
- In SEER*Stat there are 2 sets of life tables:
  - US life tables from 1970+, race (white, black and other)
  - US by geography and socioeconomic status at the county of residency (5 race/ethnicity groups)
- [https://seer.cancer.gov/expsurvival/](https://seer.cancer.gov/expsurvival/)
Method to estimate expected survival

- Expected survival is calculated by matching each patient in the cancer cohort general population life tables by age, sex, period and other covariates (if appropriate).
- Two most used methods
  - Ederer II (default)
  - Net (Pohar-Perme)
Method to estimate expected survival (cont.)

- Ederer II: the matched individuals are considered to be at risk until the corresponding cancer patient dies or is censored. (Default)

- The Pohar-Perme approach estimates net survival directly, without explicitly estimating expected survival (Pohar-Perme et al, 2012, Biometrics)
  - We used the method developed for life table calculations (Coviello, Dickman et al. 2015, Stata)
  - Deaths are weighted with the inverse of the expected probability of surviving. Older people carry higher weight.

- Comparison of Ederer II and Pohar-Perme
Choice of expected survival method-statistic tab

- Ederer II: Default
- Net/Pohar-Perme
  - Most commonly used in international studies
  - Internal age-standardization (larger confidence interval)

Both Ederer I and Hakulinen are no longer used
Parameter TAB - Survival Time (length) Calculation

• For the SEER Research files only the Pre-Calculated Duration option is available:
  - Pre-calculated survival is calculated at each registry using complete dates and submitted to SEER/NCI. Day components of dates are NOT sent to SEER.
  - Calculate from Dates option can be available for users’ databases (using SEER*Prep), if appropriate.
• **Standard Life**: Show standard life table (actuarial) output for the interval calculation (default is in single month)
• User can change to any interval, e.g. 6 months, 12 months etc…

• **Case Listing**: Provides case listing for the selected cohort. Includes selected variables from the Table Tab and survival variables
Survival variables automatically added to Case Listing. For example

- Fields used to match to expected rate table (e.g. race, sex, year)
- Vital status recode (study cut-off used)
- End Calc Vital Status (Adjusted): calculated at the end of the duration, e.g. 5 years.
- Number of Intervals (Calculated)
- Cumulative Expected (Calculated)
- Final Interval Expected (12 month)
- Final Interval Year (Calculated)
- In Cause-Specific Survival, additional fields related to defining the events is included in the output
Cumulative Summary - Show results for the intervals defined below
1-year observed survival and relative survival (life page)

Relative Survival = \frac{63.5\%}{97.3\%} = 65.3\%

<table>
<thead>
<tr>
<th>Alive at</th>
<th>Observed</th>
<th>Expected</th>
<th>Relative</th>
<th>SE Obs</th>
<th>SE Rel</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Start</td>
<td>Died</td>
<td>Follow-up</td>
<td>Interval</td>
<td>Cum</td>
</tr>
<tr>
<td>&lt; 1 mo</td>
<td>1,113,189</td>
<td>90,795</td>
<td>2,621</td>
<td>91.0%</td>
<td>91.0%</td>
</tr>
<tr>
<td>1-&lt;2 mo</td>
<td>1,021,143</td>
<td>60,047</td>
<td>5,002</td>
<td>94.0%</td>
<td>94.0%</td>
</tr>
<tr>
<td>2-&lt;3 mo</td>
<td>954,394</td>
<td>50,113</td>
<td>5,002</td>
<td>94.7%</td>
<td>94.7%</td>
</tr>
<tr>
<td>3-&lt;4 mo</td>
<td>898,479</td>
<td>34,668</td>
<td>4,641</td>
<td>96.1%</td>
<td>96.1%</td>
</tr>
<tr>
<td>4-&lt;5 mo</td>
<td>859,149</td>
<td>29,366</td>
<td>5,311</td>
<td>96.5%</td>
<td>96.5%</td>
</tr>
<tr>
<td>5-&lt;6 mo</td>
<td>823,972</td>
<td>24,604</td>
<td>4,737</td>
<td>97.0%</td>
<td>97.0%</td>
</tr>
<tr>
<td>6-&lt;7 mo</td>
<td>794,631</td>
<td>23,380</td>
<td>4,664</td>
<td>97.0%</td>
<td>97.0%</td>
</tr>
<tr>
<td>7-&lt;8 mo</td>
<td>766,387</td>
<td>18,311</td>
<td>4,503</td>
<td>97.5%</td>
<td>97.5%</td>
</tr>
<tr>
<td>8-&lt;9 mo</td>
<td>742,233</td>
<td>17,433</td>
<td>4,330</td>
<td>97.6%</td>
<td>97.6%</td>
</tr>
<tr>
<td>9-&lt;10 mo</td>
<td>721,510</td>
<td>17,621</td>
<td>4,241</td>
<td>97.6%</td>
<td>97.6%</td>
</tr>
<tr>
<td>10-&lt;11 mo</td>
<td>698,646</td>
<td>14,018</td>
<td>3,902</td>
<td>97.0%</td>
<td>97.0%</td>
</tr>
<tr>
<td>11-&lt;12 mo</td>
<td>651,325</td>
<td>15,975</td>
<td>3,947</td>
<td>97.8%</td>
<td>97.8%</td>
</tr>
<tr>
<td>12-&lt;13 mo</td>
<td>661,506</td>
<td>12,377</td>
<td>3,497</td>
<td>97.6%</td>
<td>97.6%</td>
</tr>
<tr>
<td>13-&lt;14 mo</td>
<td>645,820</td>
<td>13,310</td>
<td>3,578</td>
<td>97.9%</td>
<td>97.9%</td>
</tr>
<tr>
<td>14-&lt;15 mo</td>
<td>629,932</td>
<td>10,614</td>
<td>3,322</td>
<td>98.3%</td>
<td>98.3%</td>
</tr>
<tr>
<td>15-&lt;16 mo</td>
<td>614,956</td>
<td>10,452</td>
<td>3,264</td>
<td>98.3%</td>
<td>98.3%</td>
</tr>
<tr>
<td>16-&lt;17 mo</td>
<td>610,186</td>
<td>10,431</td>
<td>3,244</td>
<td>98.3%</td>
<td>98.3%</td>
</tr>
</tbody>
</table>
5-year Observed, expected and relative survival (summary page)

<table>
<thead>
<tr>
<th>Time (mo)</th>
<th>N</th>
<th>Median Obs</th>
<th>Median Rel</th>
<th>Observed</th>
<th>Expected</th>
<th>Relative</th>
<th>SE Obs</th>
<th>SE Rel</th>
<th>Obs Cum CIs Lower</th>
<th>Obs Cum CIs Upper</th>
<th>Rel Cum CIs Lower</th>
<th>Rel Cum CIs Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>1,118,189</td>
<td>27.60</td>
<td>35.69</td>
<td>63</td>
<td>97.3</td>
<td>65.3</td>
<td>0</td>
<td>0</td>
<td>65.4%</td>
<td>63.6%</td>
<td>65.2%</td>
<td>65.4%</td>
</tr>
<tr>
<td>24</td>
<td>1,118,189</td>
<td>27.60</td>
<td>35.69</td>
<td>52.3</td>
<td>94.7</td>
<td>55.2</td>
<td>0</td>
<td>0.1</td>
<td>52.2%</td>
<td>52.4%</td>
<td>55.1%</td>
<td>55.3%</td>
</tr>
<tr>
<td>36</td>
<td>1,118,189</td>
<td>27.60</td>
<td>35.69</td>
<td>45.9</td>
<td>92.0</td>
<td>45.8</td>
<td>0</td>
<td>0.1</td>
<td>45.8%</td>
<td>46.0%</td>
<td>44.8%</td>
<td>46.0%</td>
</tr>
<tr>
<td>48</td>
<td>1,118,189</td>
<td>27.60</td>
<td>35.69</td>
<td>41.7</td>
<td>89.3</td>
<td>41.6</td>
<td>0</td>
<td>0.1</td>
<td>41.6%</td>
<td>41.8%</td>
<td>41.8%</td>
<td>41.8%</td>
</tr>
<tr>
<td>60</td>
<td>1,118,189</td>
<td>27.60</td>
<td>35.69</td>
<td>36.4</td>
<td>86.5</td>
<td>44.3</td>
<td>0</td>
<td>0.1</td>
<td>44.3%</td>
<td>44.5%</td>
<td>44.3%</td>
<td>44.5%</td>
</tr>
</tbody>
</table>

Actuarial method. Ederer II method used for cumulative expected. Confidence interval: Log(Log) transformation. The level is 95%.

* The relative cumulative survival is over 100 percent and has been adjusted.
# The relative cumulative survival increased from a prior interval and has been adjusted.
Figure shows Cumulative Observed (blue), Expected (orange) and relative survival (red).
User can customize Y-axis.
Graph of Survival – Bar Graph
Other SEER*Stat survival methods

1. Cohort, period and complete
2. Age-standardization in SEER*Stat
3. Conditional survival
Cohort - Includes calendar years for which all cases have potential follow-up for the survival duration. For example, the cohort method can only include patients diagnosed in 2008.

Complete Analysis - Includes all patients diagnosed in the most recent years spanning the maximum duration to be estimated.

Period - Uses only the most recent interval survival estimate of cases diagnosed in different calendar years (cross-sectional estimate of survival).

Observed survival by year of follow-up (1 to 5) and year of diagnosis (2008 to 2012).
For stability purposes, SEER uses additional years of diagnosis for reporting Complete and Period methods.

<table>
<thead>
<tr>
<th>Year</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>96.6%</td>
<td>96.6%</td>
<td>96.8%</td>
<td>96.7%</td>
<td>96.8%</td>
<td>96.8%</td>
<td>96.7%</td>
</tr>
<tr>
<td>2</td>
<td>94.8%</td>
<td>95.1%</td>
<td>95.3%</td>
<td>95.0%</td>
<td>95.2%</td>
<td>95.3%</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>94.7%</td>
<td>94.7%</td>
<td>95.0%</td>
<td>94.7%</td>
<td>94.8%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>95.1%</td>
<td>95.1%</td>
<td>95.1%</td>
<td>95.2%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>95.5%</td>
<td>95.3%</td>
<td>95.1%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 2: Observed Survival by Year of Follow up and Year of Diagnosis, Regional Female Breast Cancer, SEER 18 Registries (2006-2012) - Method Implemented by SEER**
Why to age-standardize?

- Survival generally depends on age at diagnosis, and the age distribution of cancer patients may vary over time or differ among geographical areas.
- Age-standardized survival is used to compare survival in different cancer populations with different age distributions.
- Available for “net” survival measures (relative and cause-specific).
- Large literature on age-standardization.
- In SEER*Stat direct external age-standardization is implemented.
Survival is calculated for each age group

Age-standardized survival is the weighted sum of age-specific survival

\[
AgeStd\ Surv = \sum_{age} w_{age} \ Surv_{age}
\]

The standards provided are the International Cancer Survival Standard (ICSS) derived in Corazziari et al. (2004) for the adult population (ages 15+)

Users can also define their own standards through SEER*prep
Age Standards for Survival

SEER now provides age-standard adult cancer populations (ages 15+) to calculate age-standardized survival. Age-standardized survival is used to compare survival across time or different cancer populations with different age distributions. The standards provided are the International Cancer Survival Standard (ICSS) derived in Coronel et al. (2004) for three broad groups of cancer sites with similar patterns of incidence by age. The idea is that by using the appropriate standard, the age-standardized survival would be similar to the raw (un-weighted) survival.

Standard Populations for Survival

The three standards can be generally described as:

- International Cancer Survival Standard (ICSS) 1 - For cancer sites with increasing incidence by age (most cancer sites)
- International Cancer Survival Standard (ICSS) 2 - For cancer sites with broadly constant incidence by age
- International Cancer Survival Standard (ICSS) 3 - For cancer sites that mainly affect young adults

Cancer Site for Which Each of the Three International Cancer Survival Standards (ICSS) Apply

<table>
<thead>
<tr>
<th>ICSS Standard</th>
<th>Cancer Sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Lip, tongue, salivary glands, oral cavity, oropharynx, hypopharynx, head &amp; neck, oesophagus, stomach, small intestine, colon, rectum, liver, biliary tract, pancreas, nasal cavity, larynx, lung, pleura, breast, corpus uteri, ovary, vagina &amp; vulva, penis, bladder, kidney, choroid melanoma, non-Hodgkin lymphomas, multiple myeloma, chronic lymphatic leukemia, acute myeloid leukemia, chronic myeloid leukemia, leukemia, all cancers, prostate</td>
</tr>
<tr>
<td>2</td>
<td>Nasopharynx, soft tissues, melanoma, cervix uteri, brain, thyroid gland, bone</td>
</tr>
<tr>
<td>3</td>
<td>Testis, Hodgkin's disease, acute lymphatic leukemia</td>
</tr>
</tbody>
</table>

https://seer.cancer.gov/stdpopulations/survival.html
Notes About Age-Standardized Survival

- Useful in comparisons: international, by registry or by calendar period
- Not so useful for representing survival for the specific populations
  - In some instances, it is better to show survival for different age groups
- Age-standardized survival (direct method) cannot be calculated if survival is not available for some specific age group
Statistic TAB

- **Period survival**
- **Age-standardized. Select standard for the specific cancer sites**
- **Selection of life-tables**
Conditional survival

- Conditional survival is the probability of surviving x more years/months given alive after y years/months from diagnosis.
- For example, the probability of surviving 10 years from diagnosis given alive 5 years from diagnosis is calculated as the ratio

\[
\frac{\text{Probability of surviving 10 years from diagnosis}}{\text{Probability of surviving 5 years from diagnosis}}
\]
Parameter TAB - Intervals

A screen shot of the parameter tab in SEER*Stat. The conditional survival is entered as 13-48, 25-60.

- **e.g., 13-48,25-60** (this is based on monthly survival intervals)

13-48: Given that you have survived 1 year (entering 13th interval), what is the probability that you will survive 3 additional years

25-60: Given you have survived 2 years, what is the probability that you will survive an additional 3 years
Because relative survival is a ratio it can be over 100%. This happens when the overall mortality of the cancer patients is lower than the general population (e.g., localized breast cancer, healthy screening factor). By default, SEER*Stat adjusts these estimates to not be over 100% or increase from prior cumulative interval. Users can uncheck the boxes. Option to export files that can be read by other programs (e.g. CANSURV and JPSurv software).
Demo 1 and 2
References (1)


References (2)


References (3)


4. Lambert PC, Dickman PW, Rutherford MJ. Comparison of different approaches to estimating age standardized net survival. Bmc Medical Research Methodology 2015;15.


