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

NIH NATIONAL CANCER INSTITUTE

SEER

Agenda

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

	SURVIVAL	INCIDENCE/MORTALITY
It is the study of:	Time between 2 events (e.g. diagnosis and death)	Occurrence of one event
Universe/ denominator	Cohort of cancer patients	The whole population
It is usually reported as	Proportion or percent of patients surviving/dying at a given time	Rates per 100,000 population
Censoring	Event cannot be observed for some patients (e.g. lost to follow-up, occurrence of a competing event, end of study)	-

Survival is not a one size fits all statistic
Different statistics to answer different questions

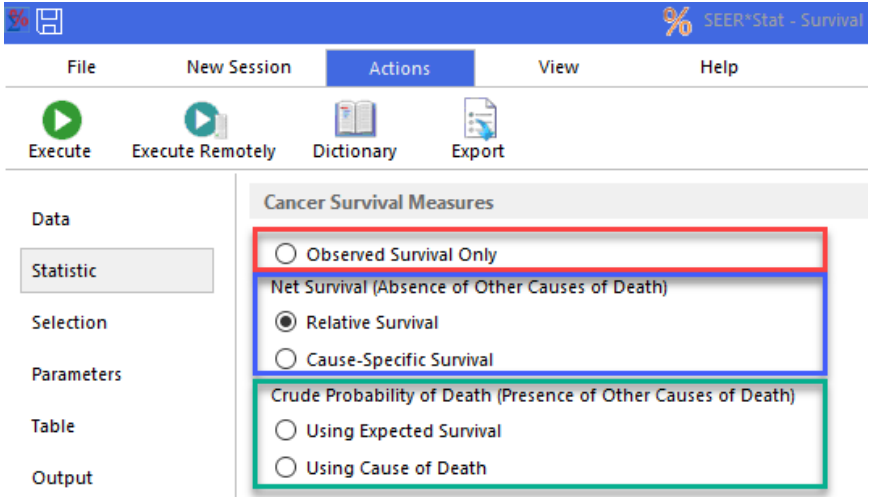
- Patient 1: I have just been diagnosed with ovarian cancer. What are my chances of surviving this cancer?
- Patient 2: I have cardiovascular disease and have been diagnosed with localized breast cancer, what are my chances of dying of breast cancer in the next 5 years?
- Science Writer: How has survival 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 → event of interest
  - 2,072 die of other causes → competing event
  - 599 are lost to follow-up → 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

$$\begin{matrix}\text{excess} \\ \text{mortality}\end{matrix} = \begin{matrix}\text{observed} \\ \text{mortality}\end{matrix} - \begin{matrix}\text{expected} \\ \text{mortality}\end{matrix}$$

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/

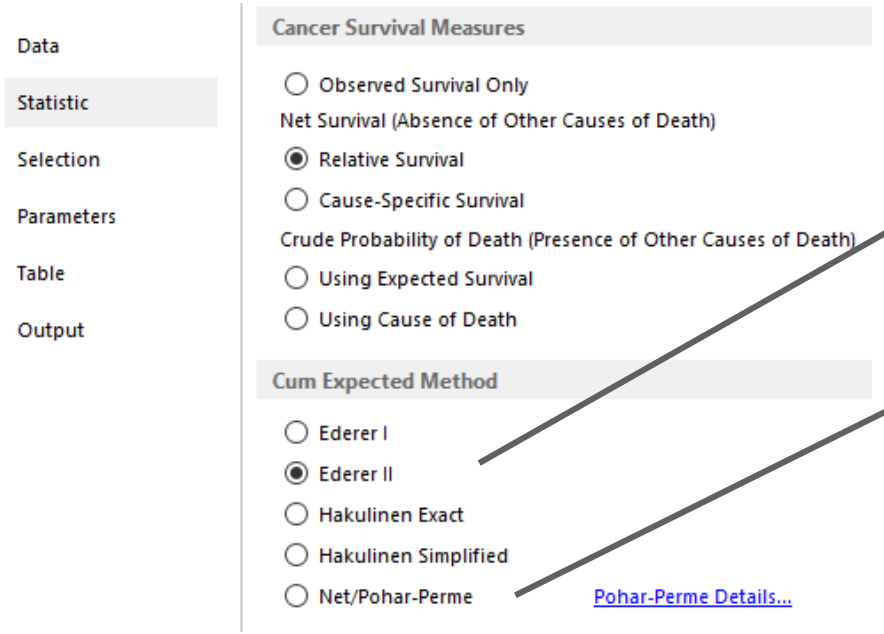
Method to estimate expected survival

- 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
  - Lambert PC, Dickman PW, Rutherford MJ. Comparison of approaches to estimating age-standardized net survival. BMC Med Res Methodol 2015;15:64.
  - Seppa K, Hakulinen T, Laara E, Pitkaniemi J. Comparing net survival estimators of cancer patients. Stat Med 2016;35:1866-1879.

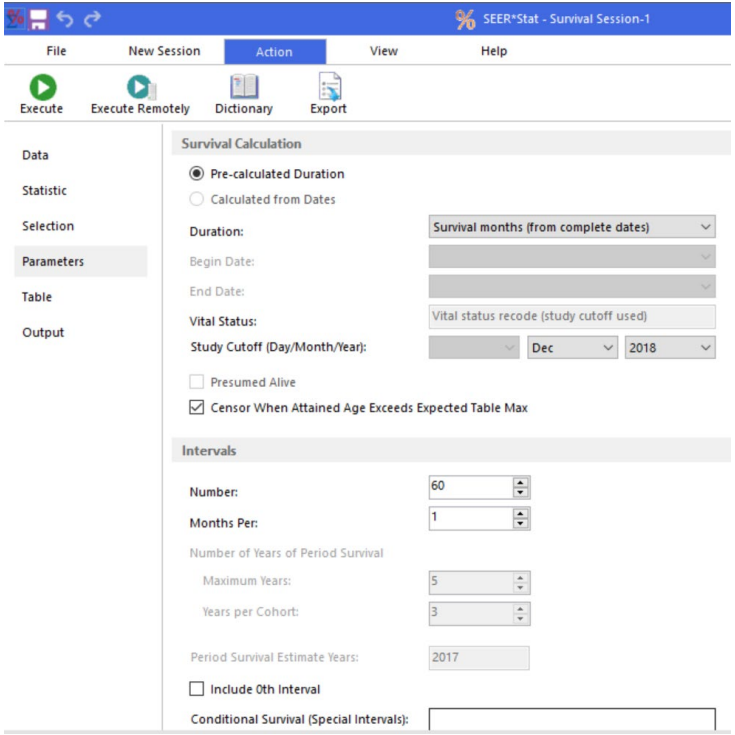
Choice of expected survival method-statistic tab



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

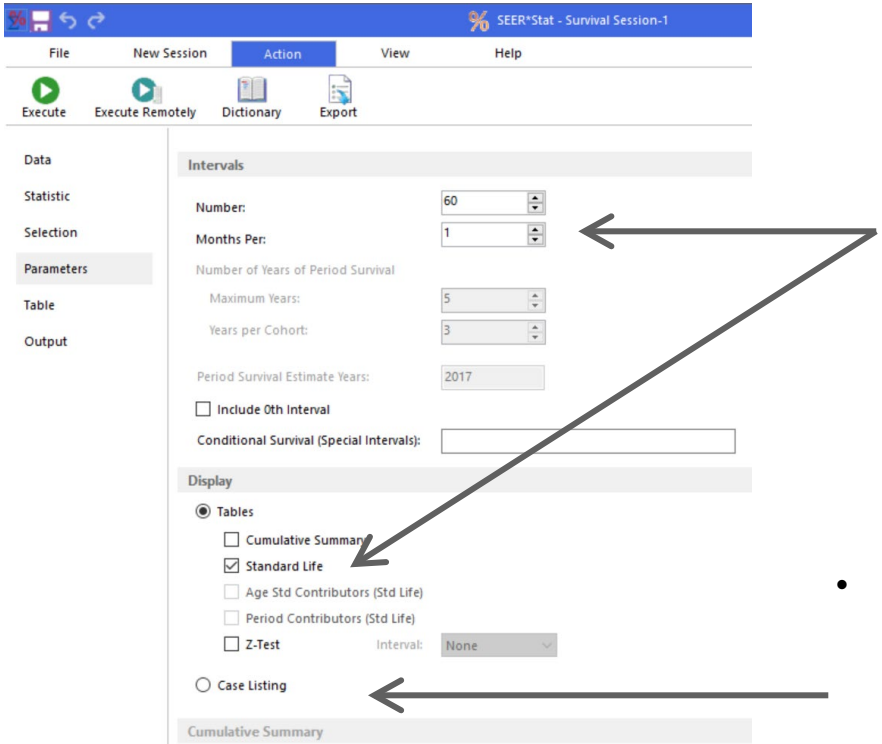
Both Ederer I and Hakulinen are no longer used

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.

Parameter TAB- Outputs Definition- Standard life and Case Listing



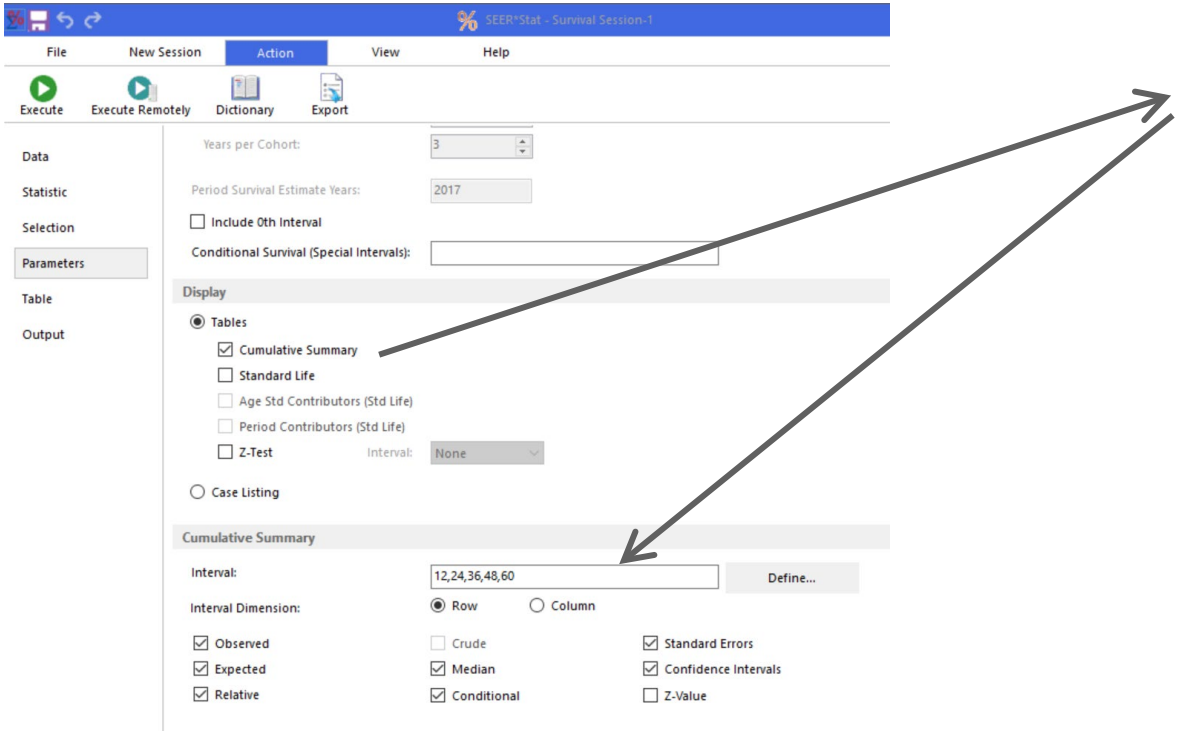
- **Standard Life**-Show standard life table (actuarial) output for the interval calculation (default is in single month)
- User can change to any interval, e.g. 6 months, 12 months etc…

- **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

Output in the Survival session



Cumulative Summary- Show results for the intervals defined below

1-year observed survival and relative survival (life page)

Observed survival

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

Expected survival

Page: Life

		Alive at Start	Died	Lost to Follow-up	Observed Interval	Observed Cum	Expected Interval	Expected Cum	Relative Interval	Relative Cum	SE Obs Interval	SE Obs Cum	SE Rel Interval	SE Rel Cum
1	< 1 mo	1,118,189	90,795	6,251	91.9%	91.9%	99.7%	99.7%	92.1%	92.1%	0.0%	0.0%	0.0%	0.0%
2	1-<2 mo	1,021,143	60,847	5,902	94.0%	86.4%	99.8%	99.5%	94.3%	86.8%	0.0%	0.0%	0.0%	0.0%
3	2-<3 mo	954,394	50,113	5,802	94.7%	81.8%	99.8%	99.2%	95.0%	82.4%	0.0%	0.0%	0.0%	0.0%
4	3-<4 mo	898,479	34,689	4,641	96.1%	78.7%	99.8%	99.0%	96.4%	79.4%	0.0%	0.0%	0.0%	0.0%
5	4-<5 mo	859,149	29,866	5,311	96.5%	75.9%	99.8%	98.8%	96.7%	76.8%	0.0%	0.0%	0.0%	0.0%
6	5-<6 mo	823,972	24,604	4,737	97.0%	73.6%	99.8%	98.6%	97.2%	74.7%	0.0%	0.0%	0.0%	0.0%
7	6-<7 mo	794,631	23,380	4,664	97.0%	71.5%	99.8%	98.4%	97.3%	72.7%	0.0%	0.0%	0.0%	0.0%
8	7-<8 mo	766,587	18,811	4,503	97.5%	69.7%	99.8%	98.1%	97.8%	71.0%	0.0%	0.0%	0.0%	0.0%
9	8-<9 mo	743,273	17,433	4,330	97.6%	68.1%	99.8%	97.9%	97.9%	69.5%	0.0%	0.0%	0.0%	0.0%
10	9-<10 mo	721,510	17,621	4,241	97.6%	66.4%	99.8%	97.7%	97.8%	68.0%	0.0%	0.0%	0.0%	0.0%
11	10-<11 mo	699,648	14,418	3,902	97.9%	65.0%	99.8%	97.5%	98.1%	66.7%	0.0%	0.0%	0.0%	0.0%
12	11-<12 mo	681,328	15,975	3,747	97.6%	63.5%	99.8%	97.3%	97.9%	65.3%	0.0%	0.0%	0.0%	0.0%
13	12-<13 mo	661,606	12,377	3,409	98.1%	62.3%	99.8%	97.1%	98.4%	64.2%	0.0%	0.0%	0.0%	0.0%
14	13-<14 mo	645,820	13,310	3,578	97.9%	61.0%	99.8%	96.8%	98.2%	63.0%	0.0%	0.0%	0.0%	0.0%
15	14-<15 mo	628,932	10,614	3,322	98.3%	60.0%	99.8%	96.6%	98.5%	62.1%	0.0%	0.0%	0.0%	0.0%
16	15-<16 mo	614,996	10,452	3,264	98.3%	59.0%	99.8%	96.4%	98.5%	61.2%	0.0%	0.0%	0.0%	0.0%
17	16-<17 mo	601,280	10,543	3,484	98.2%	57.9%	99.8%	96.2%	98.5%	60.2%	0.0%	0.0%	0.0%	0.0%

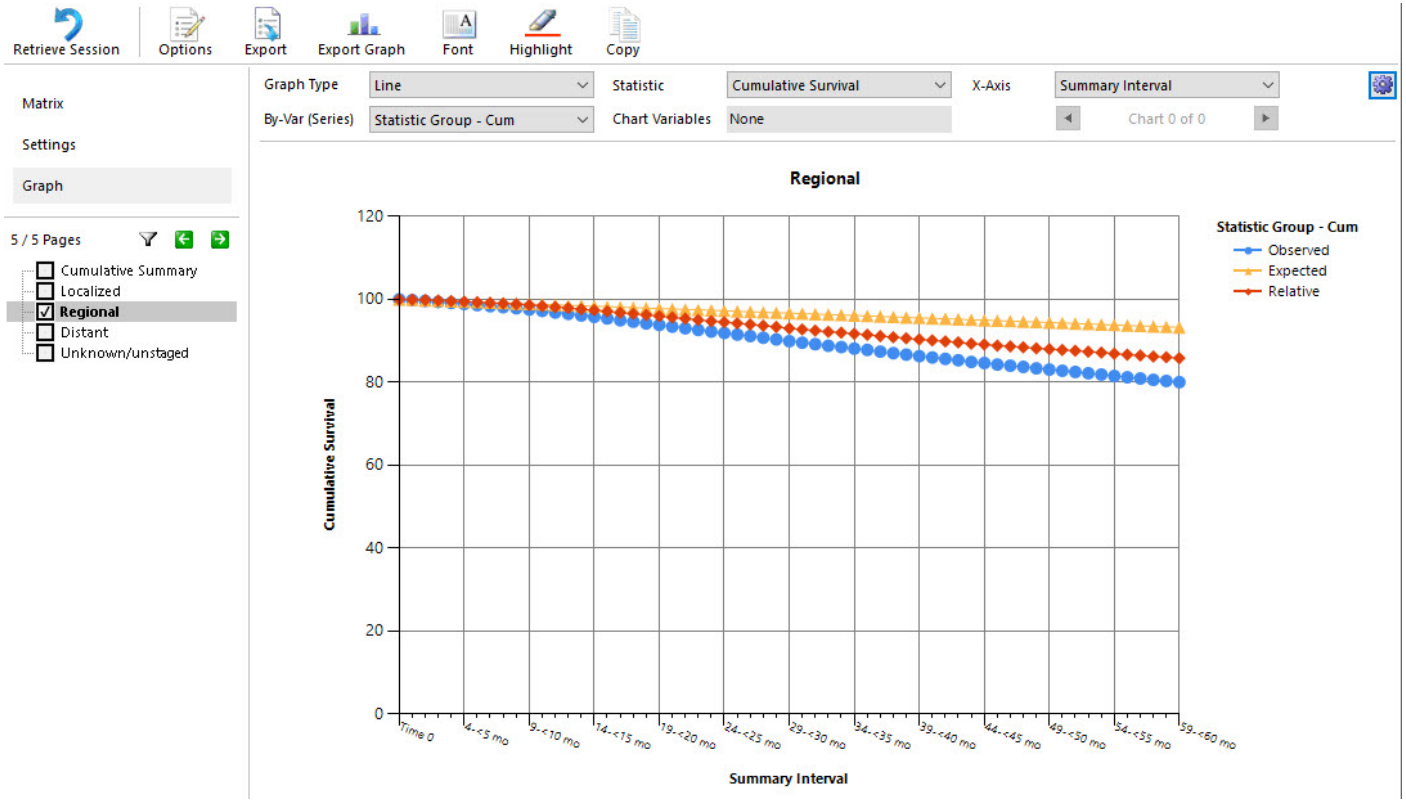
5-year Observed, expected and relative survival (summary page)

Page: Cumulative Summary

	N	Median Obs	Median Rel	Observed	Expected	Relative	SE Obs	SE Rel	Obs Cum CIs Lower	Obs Cum CIs Upper	Rel Cum CIs Lower	Rel Cum CIs Upper
12 mo	1,118,189	27.60	35.69	63.5%	97.3%	65.3%	0.0%	0.0%	63.4%	63.6%	65.2%	65.4%
24 mo	1,118,189	27.60	35.69	52.3%	94.7%	55.2%	0.0%	0.1%	52.2%	52.4%	55.1%	55.3%
36 mo	1,118,189	27.60	35.69	45.9%	92.0%	49.9%	0.0%	0.1%	45.8%	46.0%	49.8%	50.0%
48 mo	1,118,189	27.60	35.69	41.7%	89.3%	46.7%	0.0%	0.1%	41.6%	41.8%	46.6%	46.8%
60 mo	1,118,189	27.60	35.69	38.4%	86.5%	44.3%	0.0%	0.1%	38.3%	38.5%	44.2%	44.5%

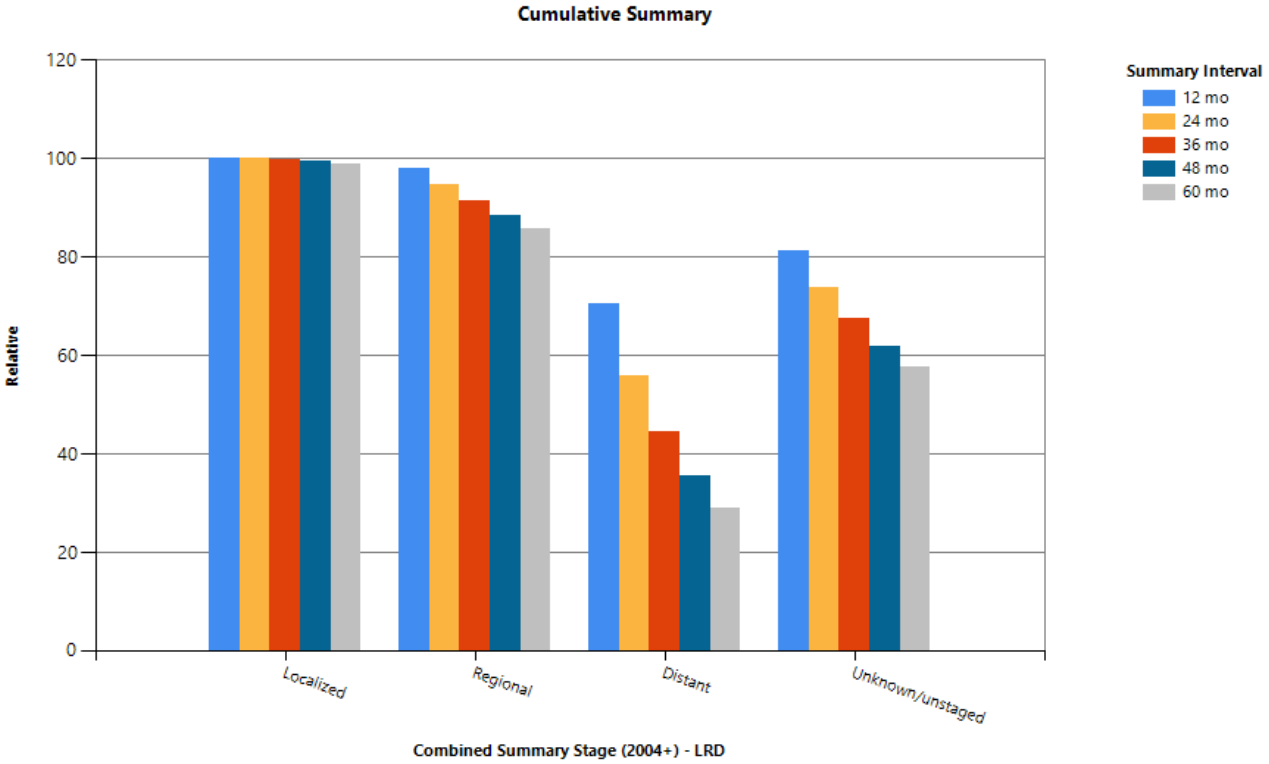
Actuarial method. Ederer II method used for cumulative expected.
Confidence interval: Log(-Log()) Transformation. The level is 95%.
* The relative cumulative survival is over 100 percent and has been adjusted.
\# The relative cumulative survival increased from a prior interval and has been adjusted.

Figure shows Cumulative Observed (blue), Expected (orange) and relative survival (red)
User can customize Y-axis



SEER*Stat Graphs

Graph of Survival – Bar Graph

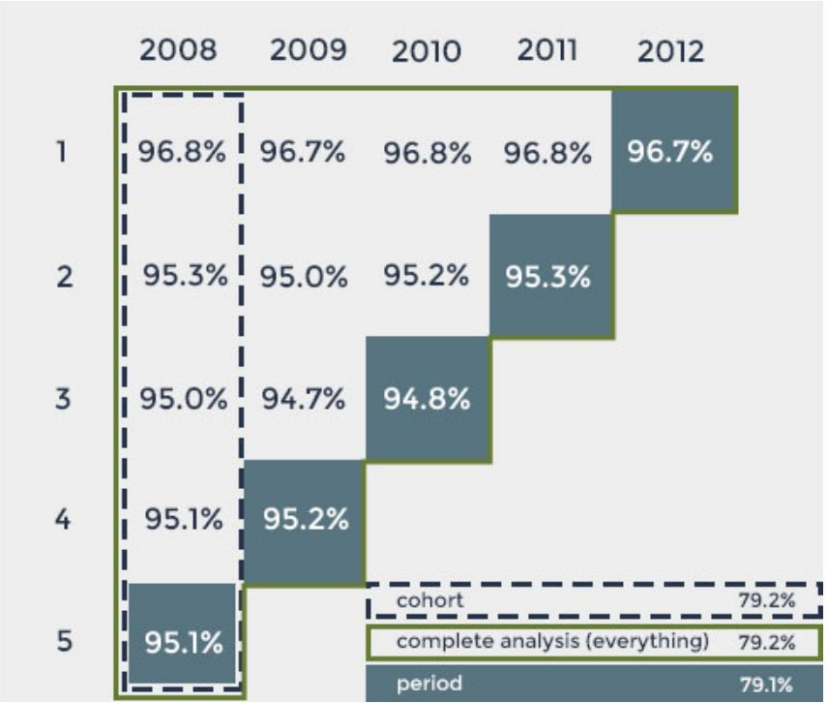


Other SEER*Stat survival methods

1. *Cohort, period and complete*
2. *Age-standardization in SEER*Stat*
3. *Conditional survival*

https://surveillance.cancer.gov/survival/cohort.html

Figure 1: Observed Survival by Year of Follow up and Year of Diagnosis, Regional Female Breast Cancer, SEER 18 Registries (2008-2012)

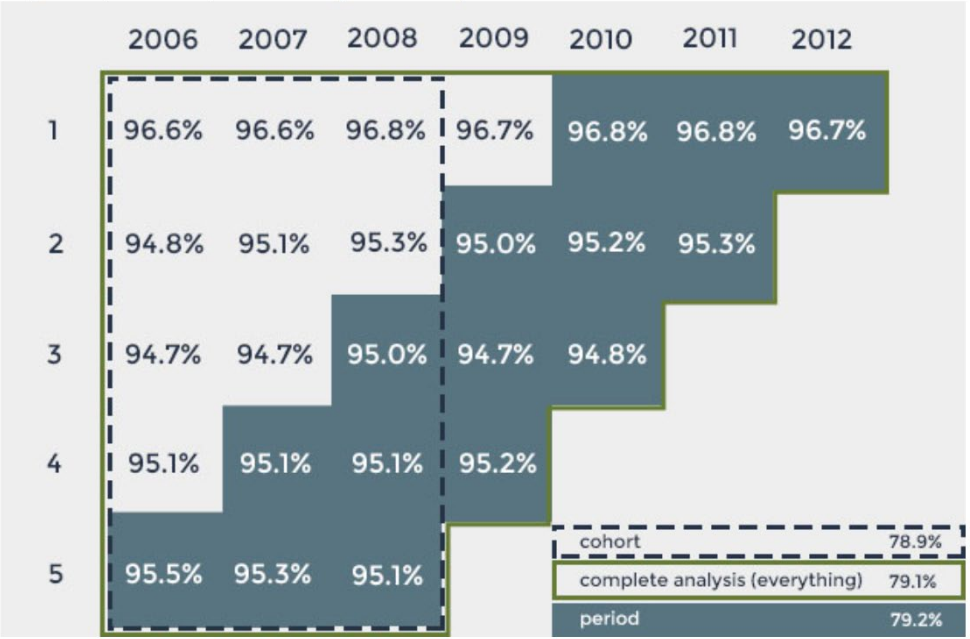


- Cohort - Includes calendar years for which all cases have potential follow-up for the survival duration. For example, the cohort method can only include patients diagnosed in 2008.
- Complete Analysis - Includes all patients diagnosed in the most recent years spanning the maximum duration to be estimated.
- 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

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

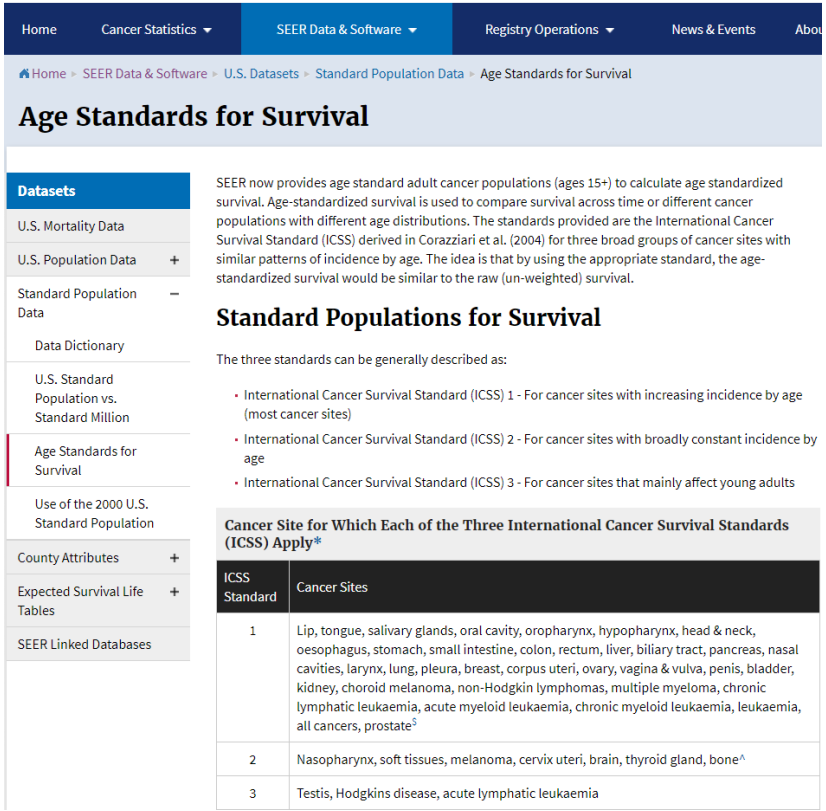
Direct (or External) Age-standardization

- 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

https://seer.cancer.gov/stdpopulations/survival.html



Home | Cancer Statistics | SEER Data & Software | Registry Operations | News & Events | Abou

Home ▸ SEER Data & Software ▸ U.S. Datasets ▸ Standard Population Data ▸ Age Standards for Survival

Age Standards for Survival

Datasets

- U.S. Mortality Data
- U.S. Population Data +
- Standard Population Data −
  - Data Dictionary
  - U.S. Standard Population vs. Standard Million
  - Age Standards for Survival
  - Use of the 2000 U.S. Standard Population
- County Attributes +
- Expected Survival Life Tables +
- SEER Linked Databases

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 Corazziari 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

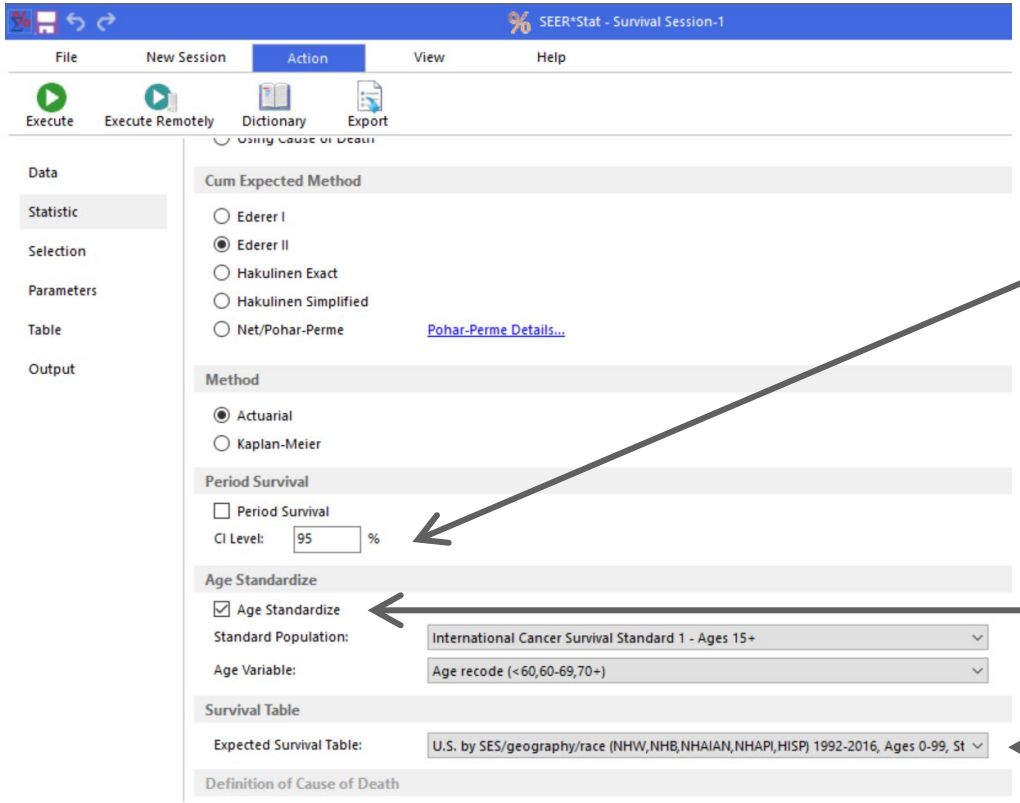
Cancer Site for Which Each of the Three International Cancer Survival Standards (ICSS) Apply*

ICSS Standard	Cancer Sites
1	Lip, tongue, salivary glands, oral cavity, oropharynx, hypopharynx, head & neck, oesophagus, stomach, small intestine, colon, rectum, liver, biliary tract, pancreas, nasal cavities, larynx, lung, pleura, breast, corpus uteri, ovary, vagina & vulva, penis, bladder, kidney, choroid melanoma, non-Hodgkin lymphomas, multiple myeloma, chronic lymphatic leukaemia, acute myeloid leukaemia, chronic myeloid leukaemia, leukaemia, all cancers, prostate§
2	Nasopharynx, soft tissues, melanoma, cervix uteri, brain, thyroid gland, bone^
3	Testis, Hodgkins disease, acute lymphatic leukaemia

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

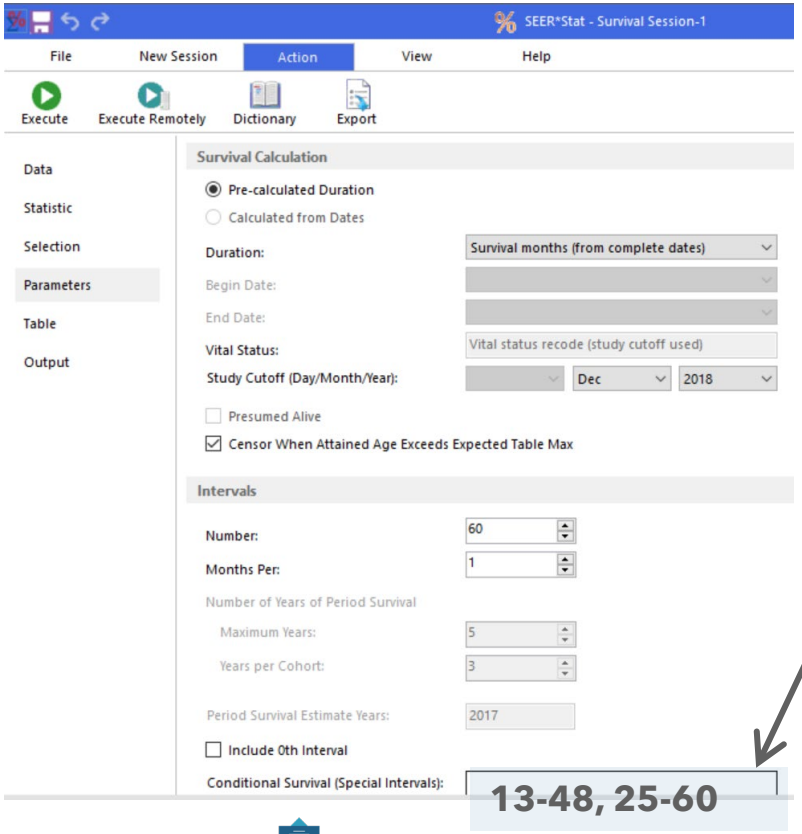
Conditional survival

- 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

$$\frac{Probability\ of\ surviving\ 10\ years\ from\ diagnosis}{Probability\ of\ survivig\ 5\ years\ from\ diagnosis}$$

Parameter TAB -Intervals

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

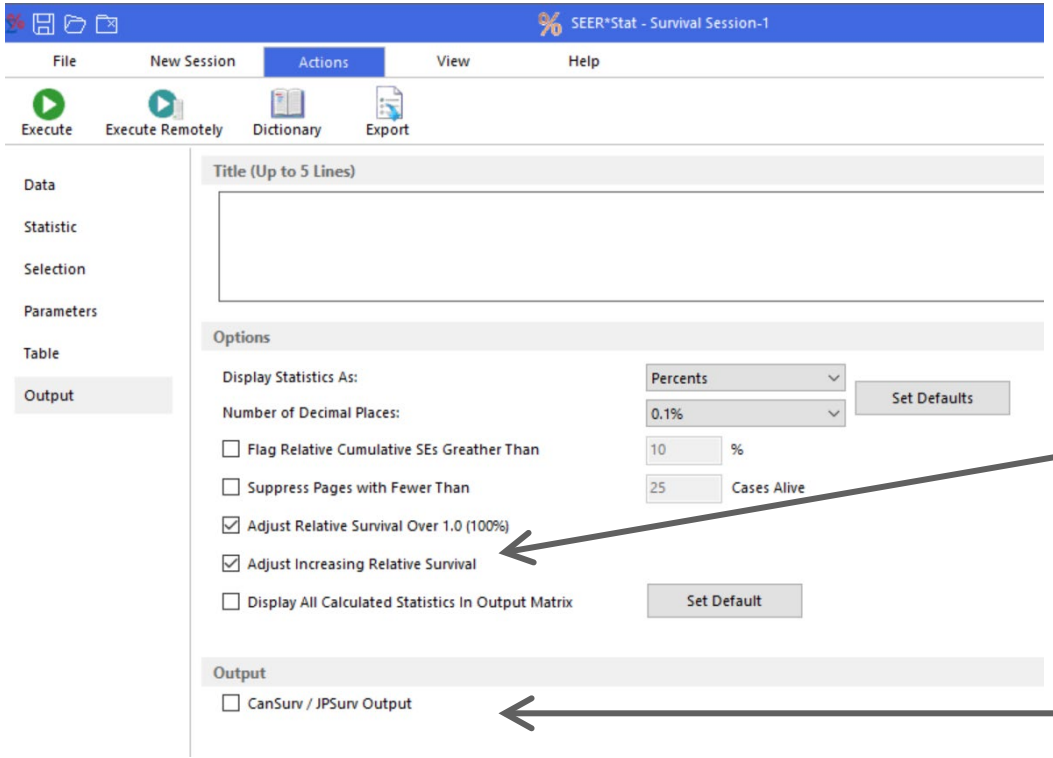


e.g., 13-48,25-60 (this is based on monthly survival intervals)

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

OUTPUT TAB OPTIONS



- 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)

- Cho H, Howlader N, Mariotto AB, Cronin KA (2011). "Estimating relative survival for cancer patients from SEER using expected rates based on Ederer I versus Ederer II method", Surveillance Research Program, National Cancer Institute; 2011. Technical Report #2011-01. Available from: http://surveillance.cancer.gov/reports/
- Ederer, F., Axtell, L. M., and Cutler, S. J. (1961), "The Relative Survival Rate: A Statistical Methodology," *Natl Cancer Inst Monogr*, 6, 101-121.
- Ederer, F. and Heise, H. (1959), "Instructions to Ibm 650 Programmers in Processing Survival Computations," Technical, End Results Evaluation Section, National Cancer Institute.
- Hakulinen, T. (1982), "Cancer Survival Corrected for Heterogeneity in Patient Withdrawal," Biometrics, 38, 933-942
- Hakulinen T, Abeywickrama KH. A computer program package for relative survival analysis. *Computer Programs in Biomedicine* 1985;19:197-207
- Hakulinen, T. (1977), "Long-Term Relative Survival Rates," Journal of Chronic Diseases, 30, 431-443
- Perme, M.P., Stare, J., and Esteve, J. (2012), "On Estimation in Relative Survival", Biometrics, 68, 113-120.
- Cronin, K. A. and Feuer, E. J. (2000), "Cumulative cause-specific mortality for cancer patients in the presence of other causes: a crude analogue of relative survival", Stat Med;19:1729-1740.

References (2)

- Marubini E, Valsecchi MG. Analyzing Survival Data from Clinical Trials and Observational Studies. John Wiley & Sons, England, 1995.
- Schairer C, Mink PJ, Carroll L, Devesa SS. (2004), "Probabilities of Death From Breast Cancer and Other Causes Among Female Breast Cancer Patients". J Natl Cancer Inst, 96:1311-21.
- Howlader, N., Ries, L., Mariotto, A., Reichman, M. Ruhl, J., Cronin, A. (2010) "Improved Estimates of Cancer-Specific Survival Rates from Population-Based Data." *Journal of National Cancer Institute*, Vol. 102, Issue 20.
- Brenner H, Hakulinen T. Advanced detection of time trends in long-term cancer patient survival: experience from 50 years of cancer registration in Finland. *Am J Epidemiol* 2002 Sep 15;156(6):566-77.
- Mariotto et. al. Estimates of long-term survival for newly diagnosed cancer patients: a projection approach. Cancer 2006; 106(9):2039-50
- Howlader N, Mariotto AB, Woloshin S, Schwartz LM. Providing Clinicians and Patients With Actual Prognosis: Cancer in the Context of Competing Causes of Death. J Natl Cancer Inst Monogr 2014;49:255–264

References (3)

- 1. Perme MP, Stare J, Esteve J. On estimation in relative survival. Biometrics 2012;68(1):113-20.
- 2. Coviello E, Dickman PW, Seppa K, et al. Estimating net survival using a life-table approach. Stata Journal 2015;15(1):173-185.
- 3. Dickman PW, Lambert PC, Coviello E, et al. Estimating net survival in population-based cancer studies. International journal of cancer Journal international du cancer 2013;133(2):519-21.
- 4. Lambert PC, Dickman PW, Rutherford MJ. Comparison of different approaches to estimating age standardized net survival. Bmc Medical Research Methodology 2015;15.
- 5. Forjaz G, Howlader N, Scoppa S, et al. Subsequent impact of including second and later cancers in cause-specific survival estimates using population-based registry data. Cancer 2021; 10.1002/cncr.33940.
- 6. Forjaz de Lacerda G, Howlader N, Mariotto AB. Differences in Cancer Survival with Relative versus Cause-Specific Approaches: An Update Using More Accurate Life Tables. Cancer Epidemiol Biomarkers Prev 2019;28(9):1544-1551.
- 7. Mariotto AB, Zou Z, Johnson CJ, et al. Geographical, racial and socio-economic variation in life expectancy in the US and their impact on cancer relative survival. PLoS One 2018;13(7):e0201034.

NATIONAL
CANCER
INSTITUTE

www.cancer.gov

www.cancer.gov/espanol