*Statistical Research and Applications Branch, NCI, Technical Report # 2003-02-A*

# Differences Between Brenner et al. and NCI Methods for Calculating Period Survival[1]

**Kathleen Cronin, Angela Mariotto, Steve Scoppa, Don Green, Lynn Clegg**

The application of period survival to the SEER data led to several modifications to the methodology proposed for calculating period survival by Brenner et al.(1,2) These modifications were introduced to address data issues specific to SEER. The differences between the Brenner et al. and NCI methods have a minor effect on the resulting period survival estimates. This report summarizes the difference between the two methods and presents a few examples demonstrating the impact on the estimates for period survival.

We use the following terminology in this report.

Last data year – the most recent year for which SEER data is available. This lags behind the last calendar year by 3 years. For example in calendar year 2003, the last year of data included in SEER reporting is 2000.

Time scales:

- Survival time – time since diagnosis for each patient
- Calendar time – calendar year

## Data

NCI

Survival estimates produced by the SEER program do not include patients diagnosed in the last data year. For example, in calendar year 2003 and last data year 2000 no cases diagnosed in 2000 would be included in the survival estimate. However follow-up information on cases diagnosed before the year 2000 would be included in the survival estimate. To calculate 10-year period survival, data would be used for cases diagnosed in 1990 to 1999 with follow-up information through 2000. Exclusion of cases diagnosed in the last data year is based on the fact that for these cases death information arrives earlier than information that they are alive. Thus, for cases diagnosed in the last data year a higher percent is lost to follow-up biasing estimates of 1-year survival. For example using

[1] This report corrects and updates:

Cronin K, Mariotto A, Scoppa S, Green D, Clegg L. Differences Between Brenner et al. and NCI Methods for Calculating Period Survival. Technical Report # 2003-02, Statistical Research and Applications Branch, National Cancer Institute 2003 (http://srab.cancer.gov/reports).

When citing this report use:

Cronin K, Mariotto A, Scoppa S, Green D, Clegg L. Differences Between Brenner et al. and NCI Methods for Calculating Period Survival. , Statistical Research and Applications Branch , NCI, Technical Report # 2003-02-A (http://srab.cancer.gov/reports).

the November 2002 submission of the SEER data , publicly released in April 2003, and last data year 2000, the 1-year relative survival for cases diagnosed in 1997,1998, 1999 and 2000 are respectively, 78.2%, 78.5%, 78.9% and 73.5%, which shows a bias in the 2000 1-year survival estimate.

Brenner et al.
Cases diagnosed in the last data year (i.e. 2000) are included in the estimation of one-year survival.

**Inclusion Criterion**

NCI
To estimate the conditional probability of surviving *x*-years conditioned on surviving to the beginning of year *x*, we consider all cases diagnosed in

$$\text{Calendar year} = \text{Last data year} - x$$

and include only those cases that survive ($x$-1) years from diagnoses.

Note that although we consider all cases diagnosed within a particular calendar year, the inclusion criteria is based on survival time from diagnosis not calendar year. For the last data year of 2000, cases diagnosed in 1999 would be included in the estimate for 1-year survival. Similarly cases diagnosed in 1998 that survived until the beginning of their second year past diagnosis would be included in the estimate of 2-year survival.

Figure 1 illustrates calculation of 1-year survival for the last data year 2000. The horizontal axis represent year and month at diagnosis and vertical axis survival time in months. Each diagonal line represents survival experience of a patient with arrow end representing alive and square end representing death. Although the last data year is 2000 we also represent survival experience in year 2001, and future survival in grey to show that survival experiences of patients diagnosed in 1999 are the same as those of patients diagnosed in 2000.

The NCI method considers only patients diagnosed in calendar year 1999 (cases 1 through 7) and count deaths within 1 year from diagnosis (cases 1,3,4 and 7). Patient 5 is not considered as death because he died after 14 months after diagnosed. Thus, the NCI 1-year survival is 4/7.

Brenner et al.
This approach includes all cases that survive until the last data year. Therefore the inclusion criteria in based on calendar time of their last follow-up time rather than survival time. Each individual that dies during the last data year counts as a death in the interval *x* where

| | |
|---|---|
| $x$ = (last data year – diagnosis year) | if the month of death is before the month of diagnosis and in the interval |
| $x$ = (last data year – diagnosis year +1) | if the month of death is after the month of diagnosis. |

Figure 1 illustrates the calculation of 1-year survival in last data year 2000 using Brenner et al. method. All cases with last follow-up in calendar year 2000 enter the calculation (cases 2,5,6,7,8,9,10, 11,12,13 and 14). The way they contribute to the 1-year survival is given in Table 1 below. Note that in this case patients 7,8, 10 and 11 count as a death in the 1-year survival interval while patient 5 would count in the 2-year survival interval. All cases alive, independently, of when they were diagnosed 1999 or 2000 count 0.5 to the number of person-years as if in calendar year 2000 we can observed on average half year of the survival time. Cases who died within 1-year from diagnosis contribute 1 to death and 1 or 0.5 to the number of person-years if diagnosed in 2000 or 1999, respectively.

Figure 1. Diagram to illustrate how 1-year survival is calculated from cases diagnosed in 1999 and 2000 using the NCI and Brenner period algorithms. The horizontal axis represent year an month and vertical axis represent survival time in months. A patient survival experience is represented with a diagonal line. The beginning of each diagonal line represents a diagnosis. The arrow end represents that the patient is alive and a square end represents death. Year 2001 is shown with future hypothetical observed survival.

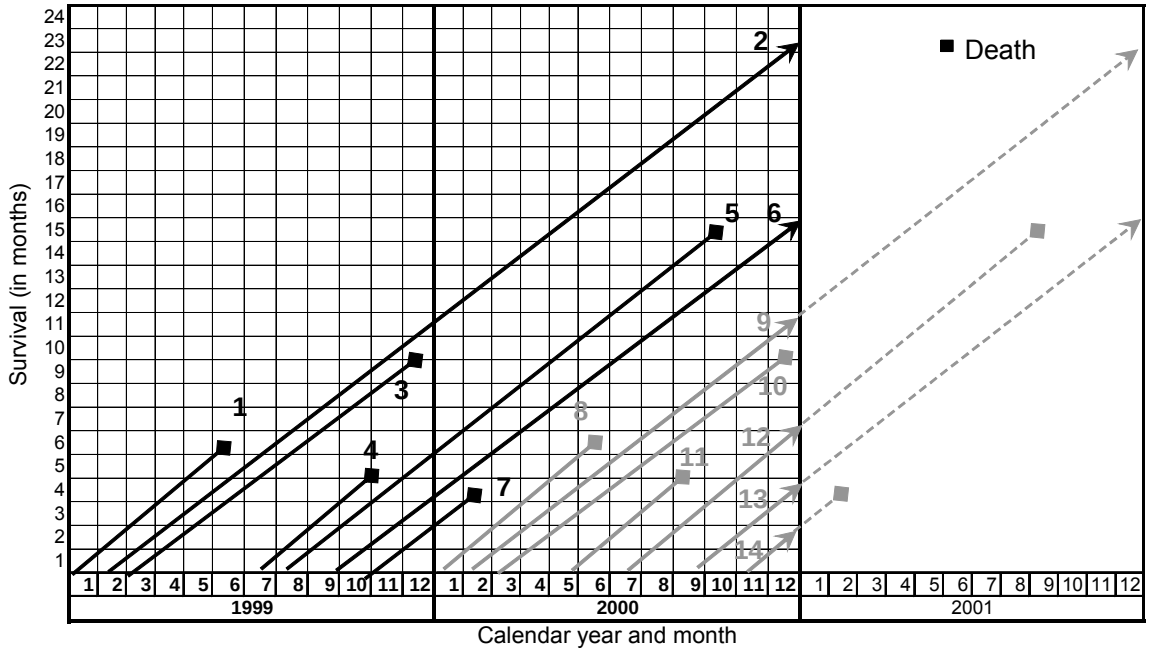

**Table 1**. Person-months and death contributions to Brenner et al. and NCI period 1-year survival using data from Figure 1.

| | NCI method | | | | Brenner method | | |
|---|---|---|---|---|---|---|---|
| **Patients** | **Included** | **Person-years** | **Deaths** | | **Included** | **Person-years** | **Deaths** |
| 1 | Yes | 1 | 1 | | No | - | - |
| 2 | Yes | 1 | - | | Yes | 0.5 | - |
| 3 | Yes | 1 | 1 | | No | - | - |
| 4 | Yes | 1 | 1 | | No | - | - |
| 5 | Yes | 1 | - | | Yes | 0.5 | - |
| 6 | Yes | 1 | - | | Yes | 0.5 | - |
| 7 | Yes | 1 | 1 | | Yes | 0.5 | 1 |
| 8 | No | - | - | | Yes | 1 | 1 |
| 9 | No | - | - | | Yes | 0.5 | - |
| 10 | No | - | - | | Yes | 1 | 1 |
| 11 | No | - | - | | Yes | 1 | 1 |
| 12 | No | - | - | | Yes | 0.5 | - |
| 13 | No | - | - | | Yes | 0.5 | - |
| 14 | No | - | - | | Yes | 0.5 | - |
| Total | | 7 | 4 | | | 7 | 4 |

Comparison of Approaches

Figures 1 demonstrates the differences between the two methods by classifying the same 14 patients under both methods. Because survival experiences were identical for patients diagnosed in 1999 and 2000 both methods give the same identical results.

With the NCI method cases diagnosed in 1999 would only contribute to 1-year survival estimates in last data year 2000, partial year follow-up information is ignored (i.e. patients diagnosed in 1999 would have on average 1.5 years of follow-up information by the end of calendar year 2000, but information after 1 year from diagnosis does not enter the calculation). The Brenner et al. method utilized partial year follow-up information. For example patients 2 and 6 in Figure 1 survive into their second year after diagnosis and therefore contribute to the number at risk for 2-year survival. The Brenner et al. method takes advantage of this partial year follow-up to get estimates of survival that are slightly more up to date.

Finally, Brenner et al. uses patients diagnosed in 1999 and followed up through 2000 in both the 1-year and 2-year estimates. Patient 7 in Figure 1 is an event in calculating 1-year survival since they survived less than 1 year. Patient 5 is counted as at risk in year 1 and an event in year 2 since they survived the first year and died in the second year. The NCI method would also count patient 7 as an event in year 1. However patient 5 would

only count at risk for the first year and not as an event in the 2-year estimates, since the 2-year estimates are based only on patients diagnosed in calendar year 1998.

**Calculation of Expected Survival**

Let $E_{ix}$ be the cumulative expected survival from diagnosis through the end of survival year x for individual i.

NCI

The calculation of expected survival for interval x is based on all individuals that had the potential to contribute to x-year survival. As described above, all individuals diagnosed in (last calendar year – x) can potentially contribute, although only individuals that survive to the beginning of the $x^{th}$ -year after diagnosis actually contribute to the numerator of the relative survival calculation.

Expected survival for interval x = $\sum_{i=1}^{n} E_{ix} \; / \sum_{i=1}^{n} E_{i(x-1)}$

Where n represents all cases diagnosed in (last data year –x) whether or not they survived until the beginning of the $x^{th}$-year after diagnosis. This definition of interval expected survival has the property that for a given cohort of patients the product of interval expected survival equals the cumulative expected survival estimate for any year. The method is referred to as *backed out expected survival* because the interval survival is backed out from the cumulative expected survival estimates.

Brenner et al.

When calculating expected survival for year *x*, only individuals that survive to the last data year are included in the calculation.

Expected survival for interval x = $\sum_{i=1}^{n} \left( E_{ix} \; / \; E_{i(x-1)} \right)$

Where there are *n* individuals included in the calculation of x-year interval relative survival, therefore they have survived up until the beginning of the last data year.

**Conclusion**

The difference described above has little effect on the estimate of period survival. To demonstrate the magnitude of difference between the NCI and Brenner et al. approaches we present a table of the two approaches applied to breast, colon and rectum, and prostate

cancer. Table 2 shows 10 year period survival estimates using the NCI and Brenner et al. methods by historical stage for the cancer sites considered.

Table 2. Comparison of NCI and Brenner et al. methods for calculating 10 year period survival using SEER 2003 submission for the SEER 9

| **Cancer site/stage** | **NCI Estimate** | **Brenner et al. Estimate** |
|---|---|---|
| Female breast / all stages | 0.805 | 0.806 |
| Female breast / localized | 0.936 | 0.934 |
| Female breast / regional | 0.672 | 0.668 |
| Female breast / distant | 0.124 | 0.122 |
| Prostate / all stages | 0.978 | 0.963 |
| Prostate / local & regional | 1.000 | 1.000 |
| Prostate / distant | 0.197 | 0.193 |
| Colon and rectum / all stages | 0.562 | 0.583 |
| Colon and rectum / localized, male and female | 0.810 | 0.831 |
| Colon and rectum / regional, male and female | 0.585 | 0.593 |
| Colon and rectum / distant, male and female | 0.070 | 0.083 |

**References**

Brenner H, Gefeller O. An alternative approach to monitoring cancer patient survival. Cancer 1996;78:2004-10

Brenner H,Gefeller O. Deriving more up-to-date estimates of long-term patient survival. J Clin Epidemiol 1997;50:211-216