# How the NCDB Data Collection Cycle Affects Survival Calculations

American College of Surgeons
*Inspiring Quality: Highest Standards, Better Outcomes*

Jerri Linn Phillips, M.A., C.T.R.
Andrew Stewart, M.A.
National Cancer Data Base, Commission on Cancer, American College of Surgeons

Commission on Cancer®

## Abstract

### Background

When the NCDB originated, a 5-year data collection cycle was adopted in order to obtain 5-year follow-up data for survival rate calculation without requiring resubmission of all diagnosis years annually. NCDB based its survival calculations on cases that had been resubmitted after the 5-year initial submission because of that cycle. However, a substantial portion of reports lacked 5 years' worth of follow-up at the time data were submitted during the 5th calendar year after the year of diagnosis.

In 2011 NCDB implemented a data collection cycle in which all new and updated cases diagnosed since the program's Reference Date are submitted annually. This study was conducted to evaluate effects of the new submission cycle on case-censoring due to lack of follow-up information and stage-specific survival rates.

### Approach

NCDB receives over 1 million case reports per diagnosis year from Commission on Cancer accredited programs. Programs are required to follow their cases annually. The reports for cases diagnosed 2004-2009 that were added or updated since the last NCDB Call for Data along with the reports for unchanged cases already in the NCDB database constitute the case pool.

Eight sites were selected to represent short- and long-term survival and relatively rare and common disease: tongue, stomach, rectum, liver, non-small cell carcinoma of lung, melanoma of skin, breast, prostate. Sites are SEER Recode sites, except for the limited lung histologies. Stage is based on the AJCC 6th edition stage group as assigned by the managing physician.

### Analysis

Two aspects of the analysis are presented here for selected sites:

(1) Percent of living cases (as of the most current NCDB data) that were followed 1, 2, 3, 4 and 5 years, and
(2) Observed survival rates based on 2004 diagnoses only, and 2004 cases plus cases diagnosed in 2005, 2006, 2007, 2008.

## The Problem

Prior to 2011, only a small portion of cases diagnosed in the "most recent" diagnosis year had even one year of follow-up recorded at submission. Over the next 4 off-cycle years, additional more recent reports trickled in for a variety of reasons, but not in systematic fashion.

Follow-up rates prior to the next on-cycle submission are shown on the table below. The orange cells depict the most recent on-cycle diagnosis year; the blue cells depict off-cycle accumulation. Using 2004 diagnoses as a timing example, cases were submitted in the fall of 2005 with a clean resubmission deadline in 2006; the NCDB analytic file copy was created after all resubmitted cases were processed. Of the eight sites evaluated, liver had the most limited follow-up; breast had the best.

| Site | | | Analytic File Copy Date | | | | |
|---|---|---|---|---|---|---|---|
| (N Living) | DX Year | Span | 2006 | 2007 | 2008 | 2009 | 2010 |
| Liver | 2004 | 1 Year | 18.2% | 23.8% | 25.6% | 26.3% | 30.8% |
| (1,833) | | 2 Years | | 3.2% | 6.2% | 7.5% | 9.4% |
| | | 3 Years | | | 1.4% | 3.1% | 4.6% |
| | | 4 Years | | | | 1.1% | 2.3% |
| | | 5 Years | | | | | 1.0% |
| (2,099) | 2005 | 1 Year | | 17.7% | 23.0% | 24.8% | 30.1% |
| | | 2 Years | | | 2.1% | 4.9% | 6.6% |
| | | 3 Years | | | | 1.6% | 2.9% |
| | | 4 Years | | | | | 1.0% |
| | | 5 Years | | | | | |
| (2,765) | 2006 | 1 Year | | | 16.5% | 22.0% | 27.4% |
| | | 2 Years | | | | 2.9% | 5.1% |
| | | 3 Years | | | | | 1.2% |
| | | 4 Years | | | | | |
| | | 5 Years | | | | | |
| (3,575) | 2007 | 1 Year | | | | 15.3% | 24.3% |
| | | 2 Years | | | | | 3.0% |
| | | 3 Years | | | | | |
| | | 4 Years | | | | | |
| | | 5 Years | | | | | |
| (4,989) | 2008 | 1 Year | | | | | 17.7% |
| | | 2 Years | | | | | |
| | | 3 Years | | | | | |
| | | 4 Years | | | | | |
| | | 5 Years | | | | | |
| Female | 2004 | 1 Year | 24.1% | 39.3% | 43.9% | 45.9% | 52.5% |
| Breast | | 2 Years | | 9.9% | 16.6% | 19.9% | 24.6% |
| (146,902) | | 3 Years | | | 3.5% | 7.7% | 11.8% |
| | | 4 Years | | | | 2.1% | 5.5% |
| | | 5 Years | | | | | 1.9% |
| (153,642) | 2005 | 1 Year | | 25.6% | 35.5% | 28.6% | 46.4% |
| | | 2 Years | | | 5.0% | 9.9% | 14.0% |
| | | 3 Years | | | | 2.6% | 5.8% |
| | | 4 Years | | | | | 1.8% |
| | | 5 Years | | | | | |
| (165,485) | 2006 | 1 Year | | | 21.1% | 30.9% | 39.0% |
| | | 2 Years | | | | 4.4% | 8.7% |
| | | 3 Years | | | | | 2.2% |
| | | 4 Years | | | | | |
| | | 5 Years | | | | | |
| (179,479) | 2007 | 1 Year | | | | 20.6% | 33.7% |
| | | 2 Years | | | | | 4.7% |
| | | 3 Years | | | | | |
| | | 4 Years | | | | | |
| | | 5 Years | | | | | |
| (194,816) | 2008 | 1 Year | | | | | 24.3% |
| | | 2 Years | | | | | |
| | | 3 Years | | | | | |
| | | 4 Years | | | | | |
| | | 5 Years | | | | | |

An early indicator that submission cycle affects follow-up was the finding that follow-up rates declined in 2008 when NCDB moved the submission date 2-3 weeks earlier in the calendar year in order to post data closer to the date of diagnosis.

## The Solution

Beginning with cases collected in 2011, NCDB made two adjustments to its data collection cycle. First, each year any cases added or modified after the previous Call for Data began are to be submitted; unmodified cases are not submitted. The result is that, in the future, all changes to a hospital's registry will be transmitted to NCDB within a year, rather than 5 or more years later. Second, the collection period was moved 3 months later in order to capture more delayed treatments in the initial collection. The positive effect on follow-up rate is notable in the final column below. The orange cells depict the most recent follow-up that could have been reported in 2011 (2010 contact); white cells depict follow-up that should be fully recorded for the respective cohort.

| Site | | | Analytic File Copy Date | | | | |
|---|---|---|---|---|---|---|---|
| (N Living) | DX Year | Span | 2006 | 2007 | 2008 | 2009 | 2010 | 2012 |
| Stomach | 2004 | 1 Year | 18.2% | 24.9% | 26.9% | 27.8% | 32.8% | 77.3% |
| (3,899) | | 2 Years | | 3.5% | 6.0% | 7.6% | 9.4% | 69.9% |
| | | 3 Years | | | 1.1% | 2.8% | 4.4% | 66.9% |
| | | 4 Years | | | | 0.8% | 1.9% | 62.7% |
| | | 5 Years | | | | | 0.7% | 56.0% |
| (4,016) | 2005 | 1 Year | | 19.9% | 24.1% | 25.5% | 32.8% | 81.2% |
| | | 2 Years | | | 1.9% | 4.3% | 5.7% | 74.4% |
| | | 3 Years | | | | 1.3% | 2.5% | 71.4% |
| | | 4 Years | | | | | 0.7% | 66.1% |
| | | 5 Years | | | | | | 50.0% |
| (4,527) | 2006 | 1 Year | | | 16.9% | 22.1% | 27.6% | 81.9% |
| | | 2 Years | | | | 2.1% | 4.0% | 75.2% |
| | | 3 Years | | | | | 1.1% | 69.1% |
| | | 4 Years | | | | | | 47.4% |
| | | 5 Years | | | | | | |
| (5,481) | 2007 | 1 Year | | | | 15.3% | 24.1% | 77.6% |
| | | 2 Years | | | | | 2.2% | 68.0% |
| | | 3 Years | | | | | | 42.3% |
| | | 4 Years | | | | | | |
| | | 5 Years | | | | | | |
| (6,542) | 2008 | 1 Year | | | | | 18.3% | 74.2% |
| | | 2 Years | | | | | | 38.2% |
| | | 3 Years | | | | | | |
| | | 4 Years | | | | | | |
| | | 5 Years | | | | | | |
| (9,050) | 2009 | 1 Year | | | | | | 35.5% |
| | | 2 Years | | | | | | |
| | | 3 Years | | | | | | |
| | | 4 Years | | | | | | |
| | | 5 Years | | | | | | |
| Melanoma | 2004 | 1 Year | 19.1% | 28.0% | 31.0% | 31.8% | 36.9% | 85.4% |
| of Skin | | 2 Years | 4.8% | 8.7% | 9.8% | 11.7% | 80.2% |
| (29,917) | | 3 Years | | 1.9% | 3.2% | 4.5% | 77.1% |
| | | 4 Years | | | 0.6% | 1.7% | 72.9% |
| | | 5 Years | | | | 0.6% | 64.3% |
| (33,218) | 2005 | 1 Year | 20.1% | 26.2% | 27.6% | 33.7% | 88.0% |
| | | 2 Years | | 2.7% | 4.5% | 6.2% | 83.3% |
| | | 3 Years | | | 0.8% | 1.9% | 79.6% |
| | | 4 Years | | | | 0.7% | 73.1% |
| | | 5 Years | | | | | 53.4% |
| (34,080) | 2006 | 1 Year | | 17.3% | 22.8% | 28.8% | 87.6% |
| | | 2 Years | | | 1.9% | 3.9% | 81.8% |
| | | 3 Years | | | | 1.2% | 74.8% |
| | | 4 Years | | | | | 50.2% |
| | | 5 Years | | | | | |
| (26,721) | 2007 | 1 Year | | | | 15.5% | 25.9% | 85.5% |
| | | 2 Years | | | | | 3.1% | 75.9% |
| | | 3 Years | | | | | | 44.2% |
| | | 4 Years | | | | | | |
| | | 5 Years | | | | | | |
| (38,877) | 2008 | 1 Year | | | | | 19.6% | 79.8% |
| | | 2 Years | | | | | | 37.4% |
| | | 3 Years | | | | | | |
| | | 4 Years | | | | | | |
| | | 5 Years | | | | | | |
| (41,383) | 2009 | 1 Year | | | | | | 37.9% |
| | | 2 Years | | | | | | |
| | | 3 Years | | | | | | |
| | | 4 Years | | | | | | |
| | | 5 Years | | | | | | |

In each instance where annual follow-up would have fallen in 2010 if it were conducted on the anniversary of the diagnosis date (orange cells), the follow-up rates decline.

## Why is follow-up so low in the orange cells?

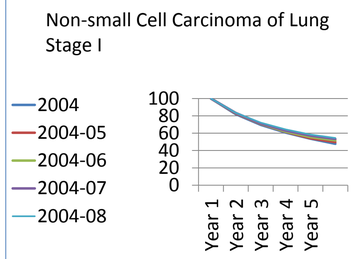
There are two reasons why follow-up may not occur on the anniversary of the date of initial diagnosis. First, if the patient reported back to the facility during the months following diagnosis – or the program's first contact with the patient was some time after diagnosis – annual follow-up may begin at a later time. Second, the Commission on Cancer survey protocol allows up to 15 months to follow the patient and record the results in the registry; both must have occurred before NCDB receives the data. NCDB examined failure to follow 2008 diagnosed disease for two years after its 2011 submissions by month of diagnosis. Patients diagnosed later in 2008 were substantially more likely not to have two years of follow-up recorded in the 2011 submissions.

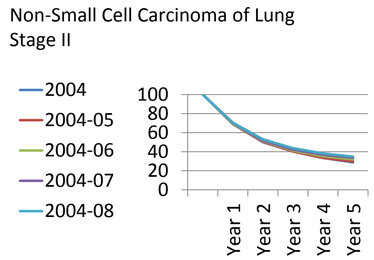
For 2008 diagnoses, percent of living patients NOT followed 2 years by month of diagnosis after 2011 submissions to NCDB

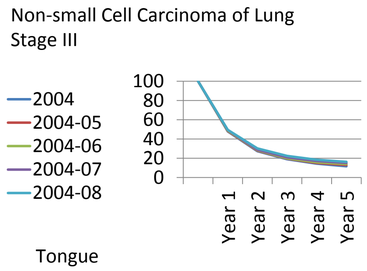
| | Tongue | Stomach | Rectum | Liver | Lung | Melanoma | Breast | Prostate |
|---|---|---|---|---|---|---|---|---|
| Jan | 26.6 | 33.2 | 29.9 | 38.7 | 32.3 | 37.0 | 22.4 | 29.1 |
| Feb | 29.9 | 37.7 | 31.5 | 41.3 | 34.1 | 39.6 | 25.0 | 32.5 |
| Mar | 35.3 | 44.4 | 37.6 | 48.4 | 37.9 | 45.3 | 29.4 | 38.4 |
| Apr | 39.0 | 51.4 | 41.9 | 50.1 | 44.2 | 49.9 | 34.5 | 43.0 |
| May | 47.5 | 53.6 | 49.3 | 53.4 | 49.4 | 55.7 | 39.7 | 48.5 |
| Jun | 51.3 | 54.8 | 51.2 | 57.5 | 55.7 | 61.1 | 45.7 | 55.0 |
| Jul | 56.2 | 66.7 | 56.8 | 72.4 | 59.7 | 66.9 | 51.3 | 59.4 |
| Aug | 62.9 | 67.8 | 68.3 | 74.7 | 64.5 | 68.4 | 56.8 | 66.3 |
| Sep | 66.6 | 72.4 | 68.7 | 72.9 | 68.7 | 74.3 | 63.5 | 70.8 |
| Oct | 71.5 | 79.0 | 75.1 | 78.6 | 74.3 | 78.6 | 69.1 | 77.1 |
| Nov | 74.6 | 81.1 | 78.9 | 85.6 | 80.0 | 84.3 | 75.5 | 82.8 |
| Dec | 85.3 | 87.0 | 83.7 | 89.1 | 85.1 | 88.0 | 82.4 | 86.7 |

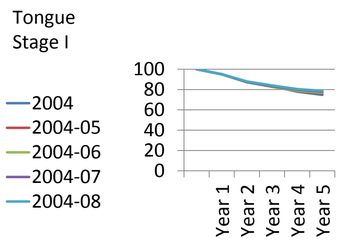
## Implications for Computing Survival Rates

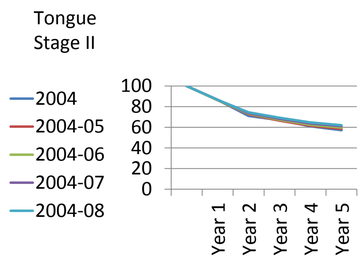
As seen in "The Problem", calculation of 5-year survival rates was not possible prior to re-collection of the diagnosis year 5 years after the initial submission. With the exception of the most recent update periods (the orange cells), the question arose whether cases diagnosed in subsequent years could be incorporated in life table or Kaplan-Meier rates. The following examples are based on observed survival rates computed using life-table methods with a cut-off date of December 31, 2009. Observed survival was selected because it is basic to most survival calculations. In the large NCDB database, patient mix changes very little from year to year and is not likely to affect these results. These graphs compare stage-specific results using 5 input populations: 2004 only, 2004-05, 2004-06, 2004-07 and 2004-09, to the same cut-off date for stages I - III.

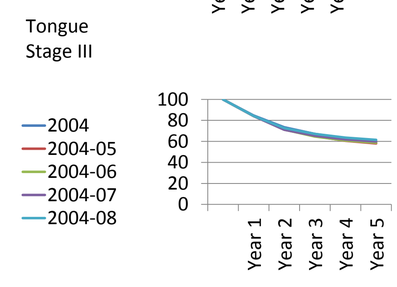



As these examples illustrate, adding cases from subsequent years did not change the calculations much. However, there was a persistent tendency for each added year to increase the calculated survival rate by year 5. The amount of increase ranges from about 1 percentage point to several. In these examples, the 5-year Stage I Non-small cell lung carcinoma survival rate increases from 47.8 percent to 54.5 when 2005 through 2008 diagnoses are taken into account. By contrast, Stage III Tongue cancer survival increases only from 59.7 percent to 61.4. The results from other sites and stages are similar.

## Conclusions

- The change in submission cycle vastly improved NCDB's access to annual follow-up results.
- Follow-up for the most recent year collected lags behind due to hospital and NCDB processes.
- Using more recent years to supplement 5-year follow-up increases the number of cases in the early years of survival calculations (not shown).
- NCDB will recalculate survival rates for the same years, using the same cut-off date, to establish the stability of these results using one additional year's submissions in the coming months. Until then, the jury is still out for including the more recent years in these calculations.