

# Data You Can Use - Topics in Cancer Control and Prevention

1/25/2018 - *Accessing Online Statistics for Childhood Cancer Incidence in Pennsylvania and Net Cancer Survival in Pennsylvania*

## Questions and Answers

1. *Do you have anything specifically to discuss for pediatric cancers- prevention, symptoms, and treatment methods? How does the new PA law for terminally ill people getting unregulated experimental and/or biological drug treatments affect them?*

**Presenter – Dr. Aplenc**

**Answer-** In general, pediatric cancers are less related to environmental factors such as smoking than adult cancers. As a result, there are fewer cancer prevention options in pediatric cancer than adult cancer. Cancer symptoms are similar in children and adults. In general, pediatric cancers can be treated more intensively as children tolerate intensive chemotherapy better than adults. However, the very newest cancer therapies are usually tested in adults first rather than children. The recent Pennsylvania law is not, in my personal practice experience, having a substantial effect on pediatric cancer care.

2. *I am interested in learning about public dataset opportunities and the proportion of the population that relates to older adults (65 years and older). Additionally, I would like to know if the data collected involves deNovo cases of cancer.*

**Presenter – Alan Fleegal – Wendy Aldinger, Pennsylvania Cancer Registry (PCR) provided this answer**

**Answer-** The Pennsylvania Cancer Registry is an incidence based registry. We collect demographic, cancer identification (site, laterally histology), staging information and first course of treat all at the time of diagnosis. We do not collect follow-up information or post cancer information.

Attached is our reporting manual that outlines what we collect.

To access our incidence statistics can be found

at: <http://www.statistics.health.pa.gov/HealthStatistics/CancerStatistics>. From there you can access our interactive on-line tool, EDDIE, to generate reports

3. *The data always seems to have a lag time of at least three years. With the advent of modern technology and the potential of being in "real time", why does the lag occur? Do other states have the same lag time? How can we improve the report out stats?*

**Presenter – Nathan Werth – Wendy Aldinger, PCR provided this answer**

**Answer-** Having data available within 24 months is the gold standard nationwide for reporting cancer incidence data. Hospitals have 180 days after the patient is discharged from their hospital to submit the case to the registry. Quality assurance procedures are done to ensure all cases are reported by the hospitals. In addition, cases from non-hospital facilities are also reported and verified. After all cases are received, the registry then must complete the consolidation of these records. The registry receives over 120,000 cases a year that get consolidated into about 80,000 incidence cases. After the cases are

consolidated, additional quality procedures are done to ensure the data is complete and accurate. The below diagram shows the process of collecting and completing a file.

**Timeline-Collection and Completion of Pennsylvania Cancer Registry Dataset**

Calendar Year	2016				2017				2018				
Task / Quarter	1	2	3	4	1	2	3	4	1	2	3	4	
Receive 2016 abstracts		[Yellow]											
Process 2016 abstracts		[Blue]											
Consolidate 2016 abstracts			[Orange]										
Close out 2016								[Grey]					
Reconcile 2016													
Conduct 2016 path clearance									[Yellow]				
Conduct 2016 death clearance										[Green]			
Prepare 2016 incidence file											[Grey]		

4. *Is it possible to get a breakdown of age groups <20? Such as 0-4, 5-9, 10-14, etc.?*

**Presenter – Alan Fleegal**

**Answer-** The Childhood Cancer Incidence dataset in EDDIE does not have break outs for specific age groups. We suppress rates when they are statistically unreliable and counts when they risk patient identification. Because of how infrequently childhood cancers are diagnosed, we could not justify including finer age groups in the EDDIE Childhood Cancer dataset.

However, the main Cancer incidence dataset in EDDIE has statistics for specific age groups (e.g., < 1, 1 to 4, 5 to 9, 10 to 14, etc.) at the county and state level. To see these age groups, the user must select anything other than “Age-Adjusted Rate” in the rate/ratio menu (e.g., try selecting “Crude/age-Specific Rate”). Be aware these data will often be suppressed.

5. *Does the chart versus US SEER get generated within EDDIE? And if yes, is it also possible to generate a chart with 3 bars: specific county, all of PA, and US SEER?*

**Presenter – Alan Fleegal**

**Answer-** No. The cancer statistics in EDDIE are limited to Pennsylvania residents and do not include statistics for the U.S. from the SEER program. To obtain U.S. SEER data please visit their website:

<https://seer.cancer.gov/>

6. *Are there any LIVESTRONG locations in Pittsburgh or Western PA? If so, how does this location differ to the Philadelphia location?*

**Moderator – Joanna Stoms**

**Answer-**There are LIVESTRONG at the Y programs throughout Pennsylvania. The nearest program can be found by searching by [ZIP Code](#).

7. *Is there a way that researchers or scientists could add potential measures or counts to be considered for the EDDIE databases or net survival registries? What is this likelihood?*

**Presenter – Nathan Werth**

**Answer-** If you would like to provide data to the Department of Health, we would be happy to accept it. Please contact us, and we'll discuss what can be done.

The net survival statistics are all from our analysis using the Pennsylvania Cancer Registry's incidence records. The laws behind the registry aim to give us complete capture of all health care professionals' encounters with a neoplasm, excluding certain morphologies the registry asks not to be reported. If you would like to provide data we could include in our analysis (for example, the way neighborhood poverty is used), please contact us.

8. *Where can I find a data dictionary or similar documentation for the PA Cancer Registry data?*

**Presenter – Alan Fleegal**

**Answer-** The registry's incidence records use the fields defined by the North American Association of Central Cancer Registries (NAACCR). The current data dictionary is version 16. You can read about the fields in this link: <http://datadictionary.naacr.org/?c=10>. The registry maintains some state-specific fields, but these are mostly internal codes with little value outside the Department of Health.

9. *Can net survival be used to estimate the actual survival chances for a patient?*

**Presenter – Nathan Werth**

**Answer -** No. Net survival only shows the risk of death added by a cancer diagnosis. The purpose of net survival rates is to assist in comparing the effect of cancer between different groups.

10. *What is "lead time bias?" Did you control for it?*

**Presenter – Nathan Werth**

**Answer-** Lead-time is how long a cancer existed before its diagnosis. Because all net cancer survival rates are measured in time since diagnosis, differences in lead time can affect the estimates. An earlier diagnosis means there will be more time alive since diagnosis.

In general, we did not control for lead time bias. This allows comparison between our rates and those calculated by the Surveillance, Epidemiology, and End Results program (SEER). To lessen the effect of lead time bias, one can either use stage-specific net survival rates or look at long-term rates, such as 10 years after diagnosis.