

SEER Variables

Annie Noone, Nadia Howlader, Lynn Ries, Jennifer Ruhl, Gretchen Flynn, Steve Scoppa

NCI Analytic Tools SEERies

November 2, 2021

1. *Overview of variables*
 - *Demographics*
 - *Tumor characteristics*
 - *Treatment*
 - *Follow-up and dates*
 - *County attributes*
2. *Cancer type recodes*
3. *Cancer type specific variables*
4. *Staging definitions*
 - *Caution when using over time*

Overview of variables

The Surveillance, Epidemiology, and End Results (SEER) Program

- Funded by NCI in 1973 *to support cancer surveillance and research* on the diagnosis, treatment and outcomes of cancer
 - 1971 National Cancer Act
- Mission:
 - Collect data on *all cancer cases* diagnosed within a described geographic area → *population-based*
 - Monitor cancer trends and support research on the diagnosis, treatment and outcomes of cancer.

Data collection

Notification from facilities required by law

Treatment facilities (Hospitals) identify cancer cases – abstract information

Labs, radiology facilities send electronic documents (pathology reports)



Population-based registries (SEER*DMS platform)

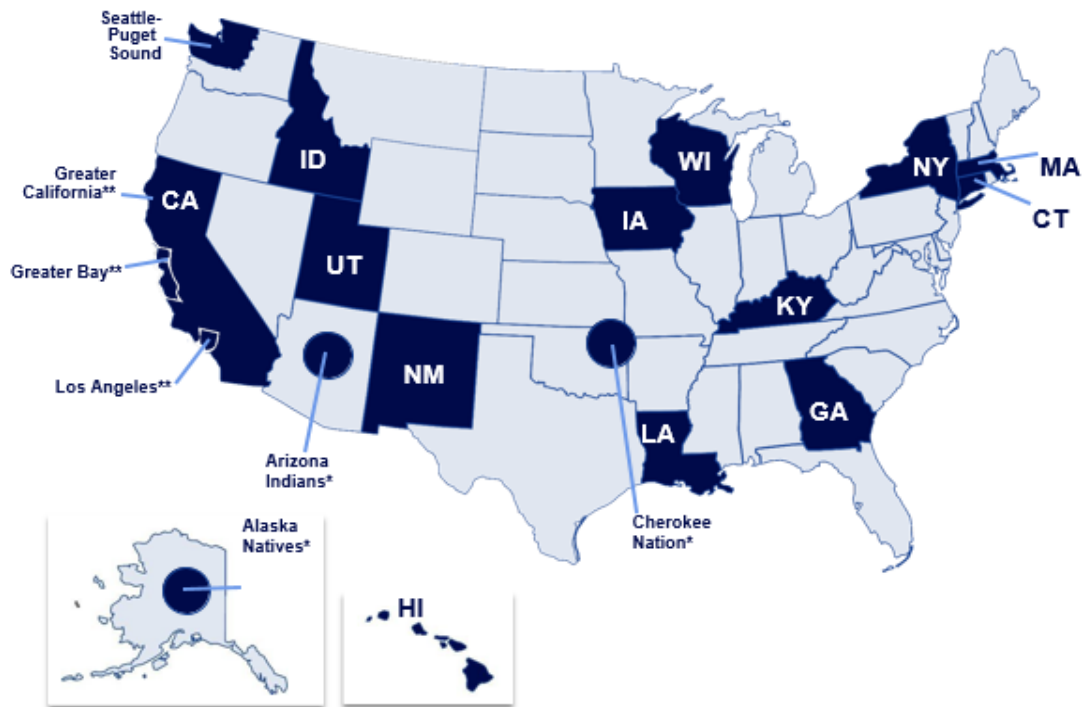
- Data abstracted and consolidated by trained cancer registrar working at hospital or registries
- Follow standards and rules
- Collect, verify and consolidate cancer cases and their cancer information



Funding organizations: NCI, CDC, NAACCR

- Certify and analyze data for public health, research, surveillance and policy use

SEER geographic areas



*Subcontract under New Mexico

**Three regions represent the state of California: Greater Bay, Los Angeles, and Greater California

Demographic variables

- Age at diagnosis
- Sex
- State / county at diagnosis
- Race and ethnicity
- Marital status

Tumor diagnosis

- Date of diagnosis (month and year)
- Tumor sequence number
 - People can be diagnosed with more than one tumor
 - 00=the only tumor of a person, 01= first tumor, 02 =second tumor
 - Over 16% of all new cases are diagnosed among individuals with prior tumors
- Type of Reporting Source
 - E.g. Hospital, radiation or oncology centers, physician's office, nursing home, autopsy/death certificate, other hospital outpatient
- Diagnostic confirmation
 - E.g. Positive microscopic confirmation, positive lab test/marker, radiology, clinical diagnosis only, unknown

Tumor characteristics: Primary site

- Most tumors are classified by body location
 - Taken from ICD-10 codes for malignant cancers
- 4 character code refers to specific location
 - All location codes are in the format:

C_{xx}._x
site subsite

- Lung is C₃₄
- Upper lobe of lung is C_{34.1}

Tumor characteristics: Histology

- Describes the appearance of the tumor cells
- A 4-digit code between 8000 and 9999 (not all values are valid)
- Generally used for selecting subsets of cancers from a specific site/location
 - “non-small cell” lung cancer
- Very useful for hematologic malignancies
 - Distinguishes different types of lymphoma

Tumor characteristics: Behavior

- Reflects the way a tumor acts within the body
- Generally preceded by a "/" to separate it from the morphology code

Registrars generally use:

/2 malignant but still growing in place
(noninvasive or in situ)

/3 malignant and can invade surrounding
tissues (malignant, primary site)

Other codes:

/0 a tumor without the potential for
spread (benign)

/1 uncertain whether benign or malignant

/6 malignant, metastatic site (borderline)

/9 malignant, uncertain whether primary
or metastatic site

Tumor characteristics: Extent of disease

- Tumor size, extension, nodal involvement, metastases
- Many different staging systems have been used
 - Historic stage
 - SEER summary stage
 - American Joint Committee on Cancer (AJCC)
- Complex data to use especially over time

<https://seer.cancer.gov/analysis/stage.html>

Tumor characteristics: Grade

- Reflects how mature (developed) the cancer cells are in a tumor
 - Differentiated tumor cells resemble normal cells and grow at a slower rate
 - Undifferentiated tumor cells, which lack the structure and function of normal cells and grow uncontrollably.

Codes

1 Well differentiated, NOS

2 Moderately, moderately well,
intermediate

3 Poorly

4 Undifferentiated, anaplastic

9 Grade was not specified

Codes for lymphoma or leukemia

5 T-cell

6 B-cell, pre-B, B-precursor

7 Null cell, Non T-non-B

8 NK, Natural killer cell

Expanded tumor characteristics

- Genomics/Biomarkers
 - Breast cancer: ER, PR, HER2, multigene assay
 - Prostate: PSA
 - CRC: CEA, KRAS
 - Etc.

First course of treatment

- All methods of treatment recorded in the treatment plan and administered to the patient before disease progression or recurrence
- Surgery
- Radiation therapy
- Systemic treatment (chemotherapy, immunotherapy, hormonal)
 - “Not received” and “Unknown if received” grouped together

<https://seer.cancer.gov/tools/codingmanuals/index.html>

<https://seer.cancer.gov/seerstat/variables/seer/surgery/>

Noone AM, Lund JL, Mariotto AB et al. (2016) Comparison of SEER treatment data with Medicare claims. *Medical Care*

Follow-up and survival information

- Death and cause of death collected through linkages with CDC's National Death Index database
- SEER also conducts linkages with other administrative data sources to verify if patient is still alive
 - Social Security Administration
 - Motor Vehicle Administration
 - Voter registry

County attributes

- Static attributes
 - Data collected at one point in time, linked regardless of diagnosis year/death
 - E.g. Percentage of families below poverty level
- Time-dependent
 - Data collected at multiple points in time, linked by matching with diagnosis year/death year
 - E.g. Socioeconomic status index
- Tutorials

<https://seer.cancer.gov/seerstat/tutorials/county.html>

Cancer type recodes

Primary site recode

- SEER converts data all data to ICD-O-3
 - Includes data originally coded in a previous version
- Recode variables are based on primary site and histology data fields
 - Provided as a research convenience for commonly reported cancer site/histology groups

<https://seer.cancer.gov/analysis/incidence.html>

Primary site recode: Example

Leukemia			
Lymphocytic Leukemia	ICD-O-3 site	ICD-O-3 Histology	Recode
Acute Lymphocytic Leukemia		9826, 9835-9836	35011
	C420, C421, C424	9811-9818, 9837	
Chronic Lymphocytic Leukemia	C420, C421, C424	9823	35012
Other Lymphocytic Leukemia		9820, 9832-9834, 9940	35013

Childhood cancer recode

- Classification for children focuses more on morphology rather than primary site
- International Classification for Childhood Cancer (ICCC) provides definitions based on site and morphology
- These are available as part of the SEER data

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

ICCC recode: Example

Site group	ICD-O-3 Histology	ICD-O-3 Primary Site	ICD-O-3 Behavior	Extended Recode	Regular
IV. Neuroblastoma And Other Peripheral Nervous Cell Tumors					
(a) Neuroblastoma and ganglioneuroblastoma	9490, 9500	000-809	3	033	041
(b) Other peripheral nervous cell tumors	8680-8683, 8690-8693, 8700, 9520-9523	000-809	3	034	042
	9501-9504	000-699, 739-768, 809	3	034	042

Other recodes

- Adolescent and young adult (AYA)
- Brain/CNS subtype
- Lymphoma subtype
- Rare cancer classification

Site-specific factors (SSFs)

- In 2004, a system was introduced called Collaborative Stage (CS)
 - A new way of defining cancer site using **schemas**
 - Collection of detailed variables called site-specific factors
 - Collected until 2018 then transition to Site-specific Data Items (SSDI)
 - Collaborative Stage staging system
 - Replaced by EOD in 2018

<https://seer.cancer.gov/seerstat/databases/ssf/>



**NATIONAL
CANCER
INSTITUTE**

www.cancer.gov

www.cancer.gov/espanol