

SEER Breast Cancer Dataset

Abstract: This dataset of breast cancer patients was obtained from the 2017 November update of the SEER Program of the NCI, which provides information on population-based cancer statistics. The dataset involved female patients with infiltrating duct and lobular carcinoma breast cancer (SEER primary cites recode NOS histology codes 8522/3) diagnosed in 2006-2010. Patients with unknown tumor size, examined regional LNs, regional positive LNs, and patients whose survival months were less than 1 month were excluded; thus, 4024 patients were ultimately included.

AGE

SEER*Stat Name: *Age at diagnosis*

Field Description: This data item represents the age of the patient at diagnosis **for this cancer**. The code represents the patient's actual age in years.

RACE

SEER*Stat Name: *Race recode (White, Black, Other)*

Field Description: Race recode is based on the race variables and the American Indian/Native American IHS link variable. This recode should be used to link to the populations for white, black and other. It is independent of Hispanic ethnicity. For more information, see : http://seer.cancer.gov/seerstat/variables/seer/race_ethnicity.

1	White
2	Black
3	Other (American Indian/AK Native, Asian/Pacific Islander)
4	Other unspecified (1991+)
5	Unknown

MARITAL STATUS

SEER*Stat Name: *Marital status at diagnosis*

Field Description: This data item identifies the patient's marital status at the time of diagnosis for the reportable tumor.

1	Single (never married)
2	Married (including common law)
3	Separated
4	Divorced
5	Widowed

T STAGE

SEER*Stat Name: Breast Adjusted AJCC 6th T (1988+)

Field Description: Created from merged EOD 3rd Edition and Collaborative Stage disease information. Currently only available for Breast schema. For more information see <http://seer.cancer.gov/seerstat/variables/seer/ajcc-stage/6th>.

1	T1
2	T2
3	T3
4	T4

N STAGE

SEER*Stat Name: Breast Adjusted AJCC 6th N (1988+)

Field Description: Created from merged EOD 3rd Edition and Collaborative Stage disease information. Currently only available for Breast schema. For more information see <http://seer.cancer.gov/seerstat/variables/seer/ajcc-stage/6th>.

1	N1
2	N2
3	N3

6TH STAGE

SEER*Stat Name: Breast Adjusted AJCC 6th Stage (1988+)

Field Description: Created from merged EOD 3rd Edition and Collaborative Stage disease information. Currently only available for Breast schema. For more information see <http://seer.cancer.gov/seerstat/variables/seer/ajcc-stage/6th>.

1	IIA
2	IIB
3	IIIA
4	IIIB
5	IIIC

GRADE

SEER*Stat Name: Grade

Field Description: Grading and differentiation codes of 1-4, 9 are defined in ICD-O-2; 1992. Grade information may be incomplete for cases diagnosed before 1977. In the early 1980's, additional codes specifying T-cell, B-cell, or null cell involvement in lymphomas and leukemias (histologies M9590-9940) were introduced by SEER. Because the reporting requirements and medical terminology have changed over time, care should be exercised when analyzing this information.

1	Grade I; grade i; grade 1; well differentiated; differentiated, NOS
2	Grade II; grade ii; grade 2; moderately differentiated; moderately differentiated; intermediate differentiation
3	Grade III; grade iii; grade 3; poorly differentiated; differentiated
4	Grade IV; grade iv; grade 4; undifferentiated; anaplastic

A STAGE

SEER*Stat Name: SEER historic stage A

Field Description: Derived from Collaborative Stage (CS) for 2004+ and Extent of Disease (EOD) from 1973-2003. It is a simplified version of stage: in situ, localized, regional, distant, & unknown. Over time several different EOD schemes have been used. Thus caution should be used when doing trend analysis. For more information including sites and years for which it isn't calculated, see <http://seer.cancer.gov/seerstat/variables/seer/lrd-stage>.

1	Regional — A neoplasm that has extended 1) beyond the limits of the organ of origin directly into surrounding organs or tissues; 2) into regional lymph nodes by way of the lymphatic system; or 3) by a combination of extension and regional lymph nodes.
2	Distant — A neoplasm that has spread to parts of the body remote from the primary tumor either by direct extension or by discontinuous metastasis (e.g., implantation or seeding) to distant organs, issues, or via the lymphatic system to distant lymph nodes.

TUMOR SIZE

SEER*Stat Name: CS tumor size (2004+)

Field Description: Information on tumor size. Available for 2004+. Earlier cases may be converted and new codes added which weren't available for use prior to the current version of CS. Each indicates exact size in millimeters. For more information, see <http://seer.cancer.gov/seerstat/variables/seer/ajcc-stage>.

ESTROGEN STATUS

SEER*Stat Name: ER Status Recode Breast Cancer (1990+)

Field Description: Created by combining information from Tumor marker 1 (1990-2003) (NAACCR Item #=1150), with information from CS site-specific factor 1 (2004+) (NAACCR Item #=2880). This field is blank for non-breast cases and cases diagnosed before 1990.

1	Positive
2	Negative

PROGESTERONE STATUS**SEER*Stat Name: PR Status Recode Breast Cancer (1990+)**

Field Description: Created by combining information from Tumor marker 2 (1990-2003) (NAACCR Item #=1150), with information from CS site-specific factor 2 (2004+) (NAACCR Item #=2880). This field is blank for non-breast cases and cases diagnosed before 1990.

1	Positive
2	Negative

REGIONAL NODES EXAMINED**SEER*Stat Name: Regional nodes examined (1988+)**

Field Description: Records the total number of regional lymph nodes that were removed and examined by the pathologist.

REGIONAL NODES POSITIVE**SEER*Stat Name: Regional nodes positive (1988+)**

Field Description: Records the exact number of regional lymph nodes examined by the pathologist that were found to contain metastases.

SURVIVAL MONTHS**SEER*Stat Name: Survival Months**

Field Description: Created using complete dates, including days, therefore may differ from survival time calculated from year and month only. For more information, see <http://seer.cancer.gov/survivaltime>.

STATUS

SEER*Stat Name: Vital status recode (study cutoff used)

Field Description: Any patient that dies after the follow-up cut-off date is recoded to alive as of the cut-off date.

1	Alive
2	Dead