



## Datasets

### Open Access

# SEER Breast Cancer Data

Race	60-69	1280 (31.8%)		
	White	348 (67.5%)	1.04 (0.94-1.14)	
	Black	201 (7.2%)	1.09 (0.89-1.31)	
	Other	320 (7.9%)		
Marital status	Divorced	485 (12%)	0.98 (0.94-1.04)	
	Married	2643 (65.7%)	1	
	Separated	45 (1.1%)	1.02 (0.95-1.09)	
	Single	615 (15.3%)	1.04 (0.91-1.13)	
	Widowed	235 (5.8%)	1.07 (0.87-1.31)	
T stage	T1	1602 (39.74%)	0.62 (0.54-0.70)	
	T2	1786 (44.4%)	1	
	T3	533 (13.2%)	1.62 (1.42-1.84)	
	T4	103 (2.48%)	2.64 (2.04-3.42)	
N stage	N0	2732 (67.9%)		
	N1	800 (19.4%)	1.46 (1.08-1.95)	
	N2	472 (11.7%)	2.13 (1.17-3.82)	
	N3			
Path stage	IIA	1305 (32.4%)	1	
	IIB	1130 (28%)	0.70 (0.60-0.81)	
	IIIA	1050 (26%)	0.49 (0.36-0.67)	
	IIIB	67 (1.67%)	0.34 (0.22-0.54)	
	IIIC	472 (11.73%)	0.24 (0.13-0.41)	
Grade	Well differentiated; Grade I	543 (13.5%)	0.90 (0.81-0.99)	
	Moderately differentiated; Grade II	2351 (58.4%)	1	
	Poorly differentiated; Grade III	1111 (27.6%)	1.11 (1.00-1.23)	
	Undifferentiated; anaplastic; Grade IV	19 (0.47%)	1.23 (1.11-1.51)	
A stage	Regional	3932 (97.7%)	1	
	Distant	92 (2.28%)	0.33 (0.25-0.43)	
Tumor size	<2 mm	250 (6%)	1-140	30.47
	26 mm - 70 mm	846 (21%)		
	71 mm - 105 mm	186 (4.6%)		
	>105 mm	36 (0.89%)		
Estrogen status	Positive	3755 (93.3%)	1	
	Negative	269 (6.68%)	1.38 (1.03-1.94)	
Progesterone status	Positive	3326 (82.7%)	1	
	Negative	698 (17.3%)	1.11 (0.88-1.4)	
Regional nodes	Total examined	4024 (100%)	1-61	14.35

Citation Author(s): JING TENG (North China Electric Power University)

Submitted by: [jing teng](#)

Last updated: Tue, 07/29/2025 - 16:04

DOI: [10.21227/a9qy-ph35](https://doi.org/10.21227/a9qy-ph35)Data Format: [.csv](#)Links: [SEER](#)

35340 views

1828 downloads

11 citations

Categories: [Biomedical and Health Sciences](#)  
[Cancer Data](#)Keywords: [Bayesian inference](#); [Breast cancer](#); [Lymph node ratio](#);  
[Prognostic](#)

Average: 5 (4 votes)

[CITE](#)[SHARE/EMBED](#)

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.

## INSTRUCTIONS:

### 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](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>.

Regional — A neoplasm that has extended 1) beyond the limits of the organ of origin  
1 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.

Distant — A neoplasm that has spread to parts of the body remote from the primary tumor either by direct extension or by discontinuous 2 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

Good morning, I am Alessandra Borba breast cancer surgeon and PhD student at PUCRS -Brazil. I would like to know if it is possible to access the TNM CLINICAL staging of this dataset , in addition to the pathological already described. I would like to research some patients cN0 and pN+. And, if it is possible

access the type of breast surgery and axillary surgery ( sentinel node + axillary dissection; just sentinel node or just axillary dissection). Thank you.

**Alessandra Souza** Wed, 12/16/2020 - 14:40 [Permalink](#)

I was diagnosed with Parkinson's disease four years ago. For over two years, I relied on Levodopa and several other medications, but unfortunately, the symptoms kept getting worse. The tremors became more noticeable, and my balance and mobility started to decline quickly. Last year, out of desperation and hope, I decided to try a herbal treatment program from NaturePath Herbal Clinic. Honestly, I was skeptical at first but within a few months of starting the treatment, I began to notice real changes. My movements became smoother, I was steadier on my feet, and, incredibly, the tremors stopped completely. It's been a life-changing experience. I feel more like myself again better than I've felt in years. If you or a loved one is struggling with Parkinson's, I truly recommend looking into their natural approach. You can visit their website at [[www.naturepathherbalclinic.com](http://www.naturepathherbalclinic.com)]

**Belinda Mary** Wed, 07/23/2025 - 09:05 [Permalink](#)

## DATASET FILES

[SEER Breast Cancer Dataset .csv \(Size: 507.84 KB\)](#)

[LOGIN TO ACCESS DATASET FILES](#)

*Open Access dataset files are accessible to all logged in users. Don't have a login? Create a [free IEEE account](#). IEEE Membership is not required.*

## DOCUMENTATION

This documentation describes the dataset feature and its value.

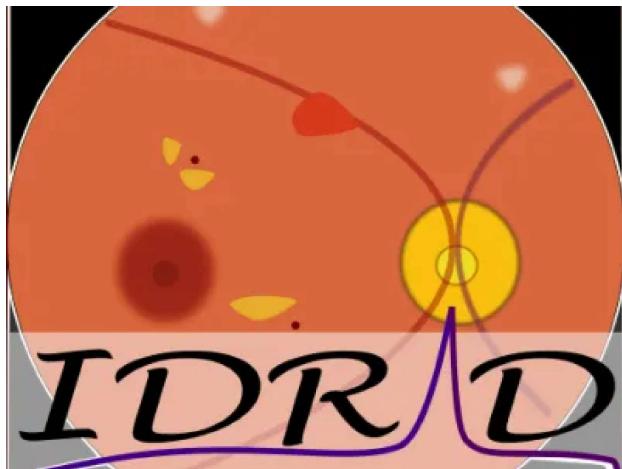
## QUESTIONS?

[✉ Send Author a Private Message](#)

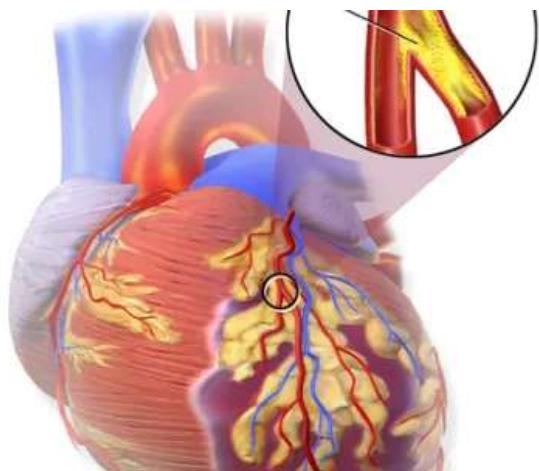
---

[⚠ Report a problem with this Dataset](#)

## More like this Dataset



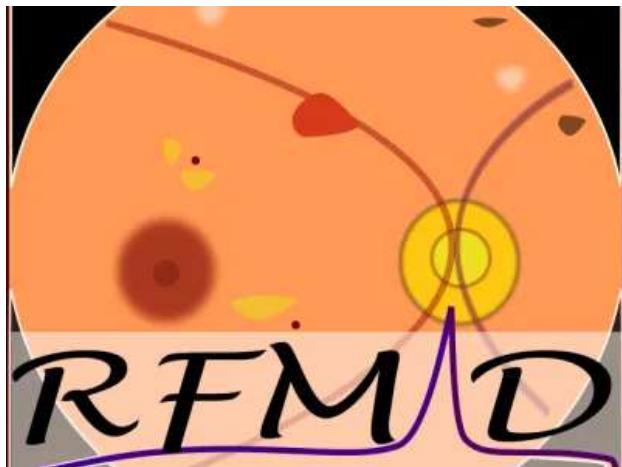
[Indian Diabetic Retinopathy Image Dataset \(IDRiD\)](#)



Heart Disease Dataset (Comprehensive)

	修改日期	类型
ue	2025/5/21 14:43	Foxit PD
gram of end...	2025/5/21 15:38	Microso
gram of the ...	2025/5/21 16:17	Microso
ver blood v...	2025/5/21 15:30	Microso
apezoidal w...	2025/5/21 16:21	Microso
angular wa...	2025/5/21 16:16	Microso

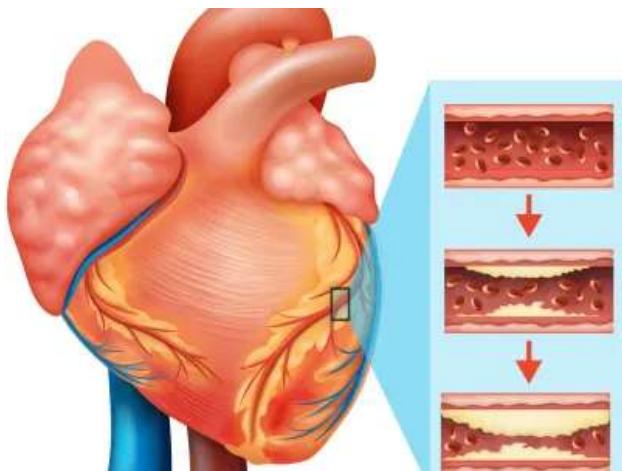
EEG data for ADHD / Control children



Retinal Fundus Multi-disease Image Dataset (RFMiD)

Dataset image

SSVEP-EEG data collection using Emotiv EPOC



### Cardiovascular Disease Dataset

Newsletter Signup:

Email Address \*

I have read and accept the IEEE Privacy Policy

Sign Up

