

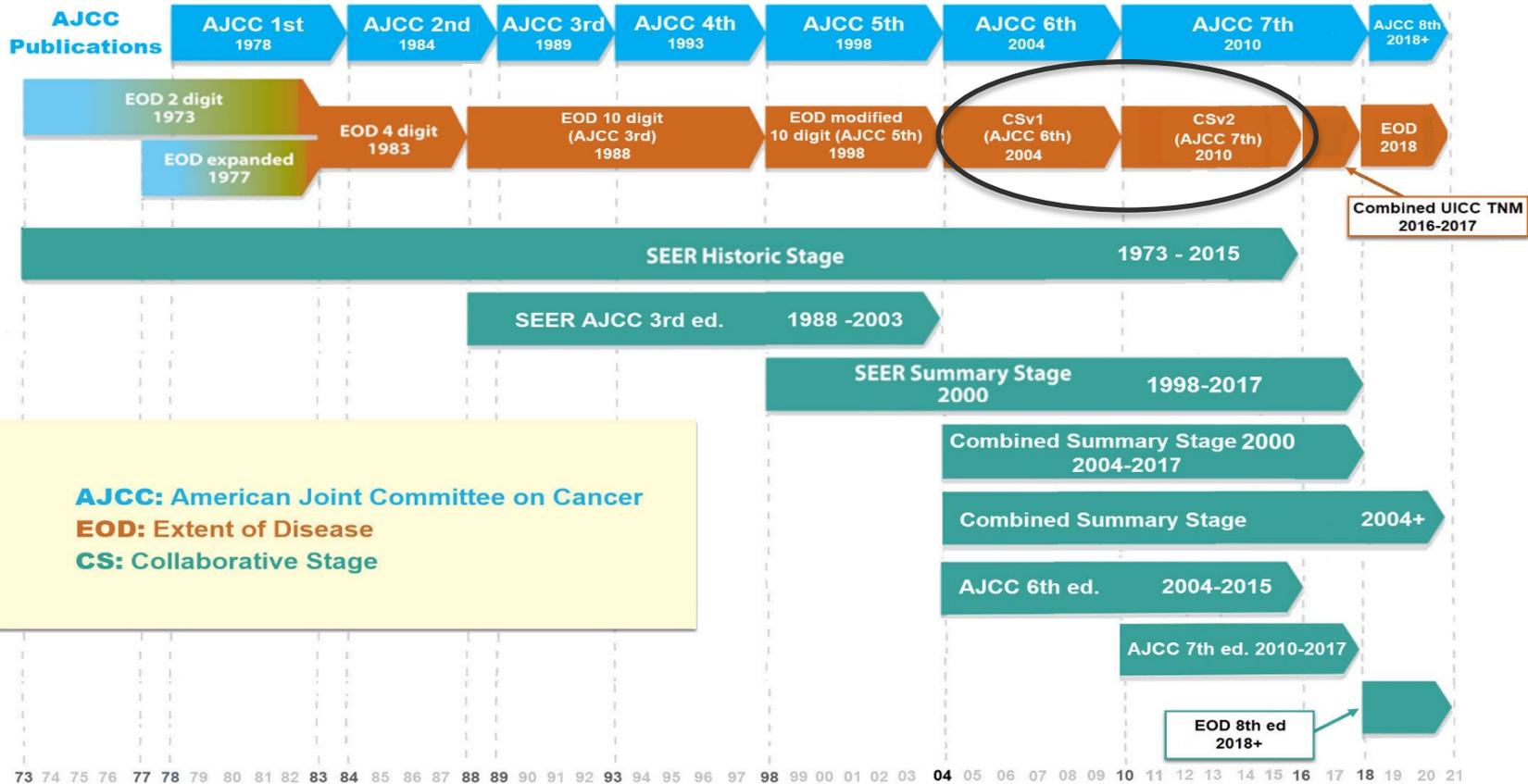
Agenda

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*

Cancer type specific variables

Background on Collaborative Staging (CS)

- A data collection system that uses a single set of data elements based on extent of disease and clinically relevant factors
 - Site specific factors (SSFs)
- It was designed to meet the needs of multiple staging systems and eliminate duplicate data collection by cancer registrars reporting to different types of registries (facility- vs population-based)
 - SEER summary stage 2000 and 1977, SEER historic stage, AJCC staging
- The CS system was introduced first in 2004 (CSv1) and updated in 2010 (CSv2) in conjunction with the release of the 7th edition of the AJCC Cancer Staging Manual
 - along with a section on prognostic factors for which collection is recommended



Background on Collaborative Staging (Cont.)

- SSFs in CS include numerous prognostic/predictive factors & biomarkers; can be used to evaluate outcome
- Much resources have been spent on the collection of CSv2 (because more factors were incorporated for collection)
- Not all of the SSFs are released on the SEER research file
 - no longer required
 - Being evaluated for completeness
 - After review was found to be of NOT as high quality as other SEER variables in the research database)

Define cancer schemas for analyzing SSFs

Define Cancer Schema vs. Site Recode

- Over the years there have been multiple standards for grouping similar cancer types by primary site and histology. Two major classifications:
 - **SEER Site Recode:** Groupings consistent with historic definitions used for collecting disease metrics in earlier time periods included in SEER data
 - **Schemas:** Groupings used in CS to dictate the appropriate staging and site-specific factors collected. Different versions of CS defined groups differently

Define Cancer Schema vs. Site recode Cont.

- How different are the cancer schemas from site recodes?
 - It depends. For example, breast cancer no change in schema vs. site recode
 - Whereas colon cancer have more changes in schema compared to site recode
 - Any analyses using CS SSFs should use CS schema to define the cohort (**not the SEER site recode variable**)

SEER Site Recode ICD-O-3/WHO 2008 Definition

Site Group	ICD-O-3 Site	ICD-O-3 Histology (Type)	Recode
Colon excluding Rectum			
Cecum	C180	excluding 9050-9055, 9140, 9590-9992	21041
Appendix	C181		21042
Ascending Colon	C182		21043
Hepatic Flexure	C183		21044
Transverse Colon	C184		21045
Splenic Flexure	C185		21046
Descending Colon	C186		21047
Sigmoid Colon	C187		21048
Large Intestine, NOS	C188-C189, C260		21049
Rectum and Rectosigmoid Junction			
Rectosigmoid Junction	C199		21051
Rectum	C209		21052

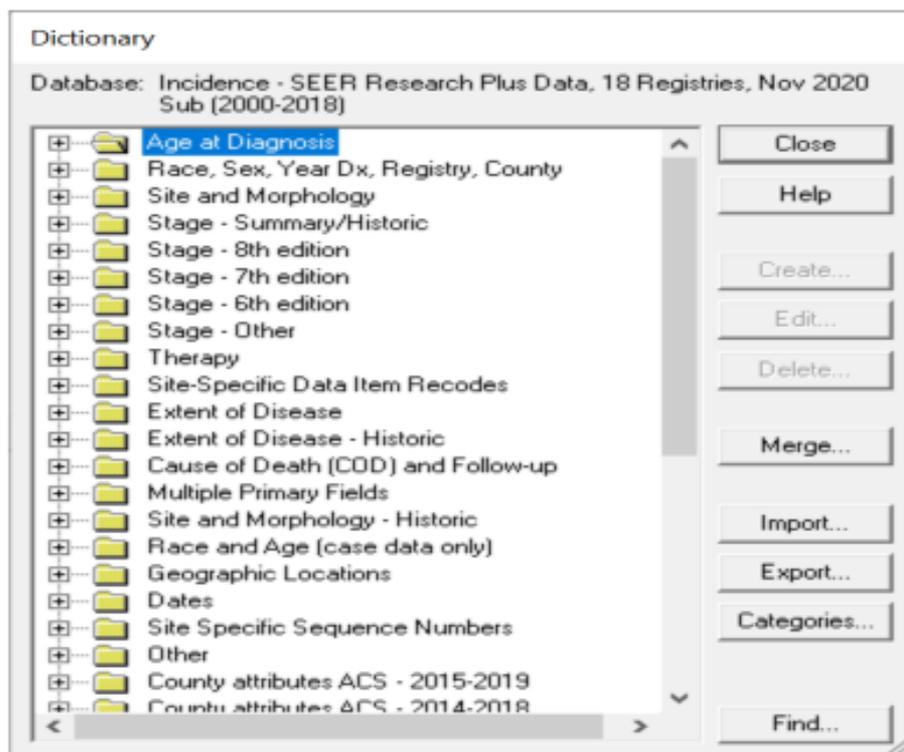
https://seer.cancer.gov/siterecode/icdo3_dwhohome/

Colon and Rectum Cancer

Site recode vs. Schema, SEER-21 2018

SEER Site Recode	Schema	Count
Colon and Rectum	Appendix	1,183
Colon and Rectum	Colon and Rectum	45,910
Colon and Rectum	Digestive Other	33
Colon and Rectum	NET Appendix	1,008
Colon and Rectum	NET Colon and Rectum	1,689
Colon and Rectum	Soft Tissues Abdomen and Thoracic (excluding Heart, Mediastinum, Pleura)	39
Colon and Rectum	GIST	56
Colon and Rectum	Soft Tissue Other (or Unusual Sites and Histology)	3

Data Dictionary in SEER*Stat



Site Recode vs Schema



Cancer Site-specific Data Items Collection Now

- The SSFs under CS system was continued to be collected until 2017 but no longer available for the latest years of diagnosed cases
- Beginning with cases diagnosed in 2018+, site-specific data items (SSDIs) are available for 2018+ cases: <https://seer.cancer.gov/seerstat/variables/seer/ssdi/>
- Some of the SSDIs are available in the latest SEER research file

Site-specific Data Items (SSDI)

Variable Name	Description	Cancer Schemas(s)
AFP Post-Orchiectomy Lab Value Recode (2010+)	Created from NAACCR Item #3805 & CS site-specific factor 12 (NAACCR #2866). Coded for cases coded as Testis in BOTH TNM 7/CS v0204+ Schema & EOD Schema Id. Formats: +	Testis
AFP Pretreatment Interpretation Recode (2010+)	Created from NAACCR Item #3809 & CS site-specific factor 1 (NAACCR #2880). Coded for cases coded as Liver in BOTH TNM 7/CS v0204+ Schema & EOD Schema Id. Formats: +	Liver
B Symptoms Recode (2010+)	Created from NAACCR Item #3812 & CS site-specific factor 2 (NAACCR #2890). Coded for cases coded as Lymphoma or Lymphoma-CLL/SLL in BOTH TNM 7/CS v0204+ Schema & EOD Schema Id.	Lymphoma, Lymphoma-CLL/SLL



EOD Schema ID Recode (2010+) in SEER*Stat



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

Site Specific data items

Available Variables

- Site-Specific Data Item Recodes
 - AFP Post-Orchiectomy Lab Value Recode (2010+)
 - AFP Pretreatment Interpretation Recode (2010+)
 - B Symptoms Recode (2010+)
 - Breslow Thickness Recode (2010+)
 - CA-125 Pretreatment Interpretation Recode (2010+)
 - CEA Pretreatment Interpretation Recode (2010+)
 - Fibrosis Score Recode (2010+)
 - Gestational Trophoblastic Prognostic Scoring Index Recode (2010+)
 - Gleason Patterns Clinical Recode (2010+)
 - Gleason Patterns Pathological Recode (2010+)
 - Gleason Score Clinical Recode (2010+)
 - Gleason Score Pathological Recode (2010+)
 - hCG Post-Orchiectomy Range Recode (2010+)
 - Invasion Beyond Capsule Recode (2010+)

Page
Row
Column
Find..

CAPS NUM

Breast cancer SSFs

Breast cancer SSFs

- A total of 24 SSFs were collected in the SEER registries
- Six of them collected since CSV1 (SSF 1-6) for cases diagnosed in 2004+; remaining SSFs collected for cases diagnosed beginning of 2010 and all the way up to 2017
- Can be grouped into 4 main categories:
 - Stage-related (N=3)
 - Prognosis/prediction (N=2)
 - Treatment/Prognosis (N=3)
 - Biomarkers/treatment (N=10)

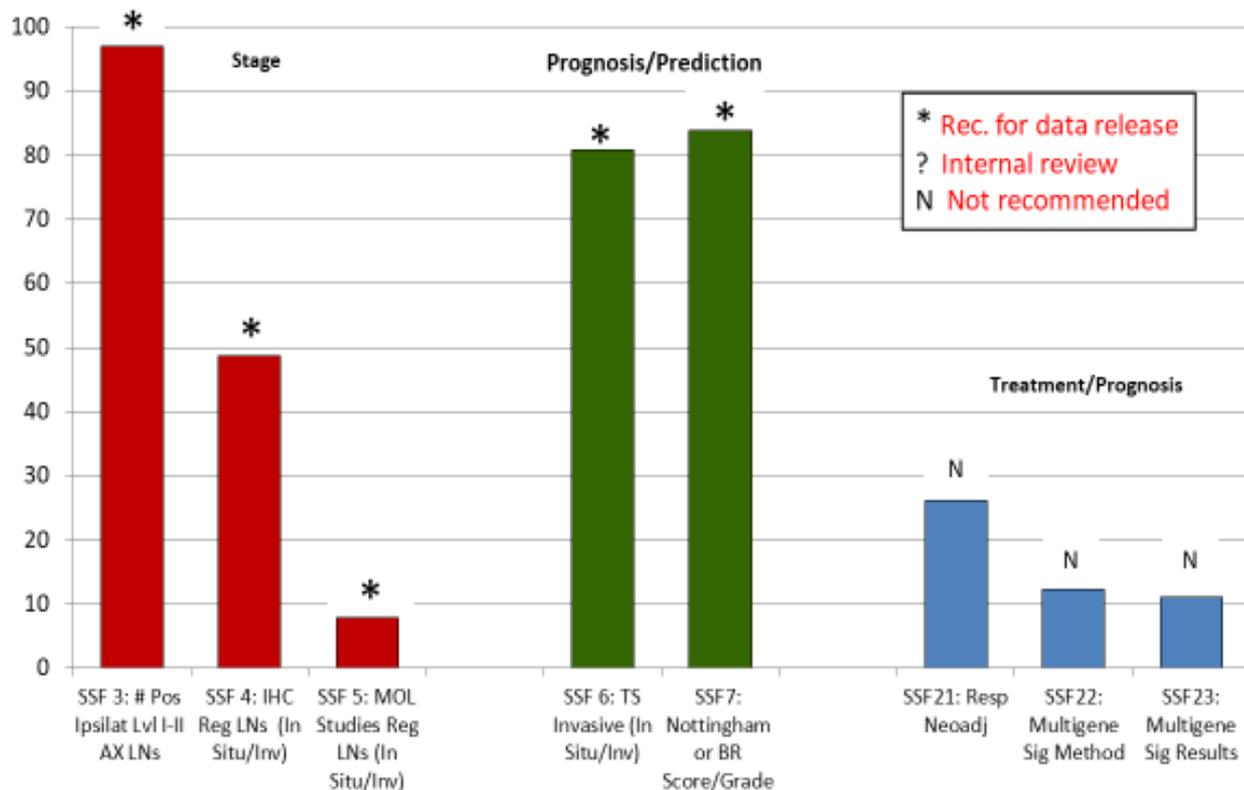
Breast Cancer SSFs

Variables	SSF Description
SSF ₁	Estrogen Receptor (ER) Assay
SSF ₂	Progesterone Receptor (PR) Assay
SSF ₃	Number of Positive Ipsilateral Axillary Lymph Nodes
SSF ₄	Immunohistochemistry (IHC) of Regional Lymph Nodes
SSF ₅	Molecular Studies of Regional Lymph Nodes
SSF ₆	Size of Tumor—Invasive Component

Note: CS version 1 began for cases diagnosed in 2004+; CS version 2 was required for 2010 +cases but may be coded for earlier years of diagnosis.

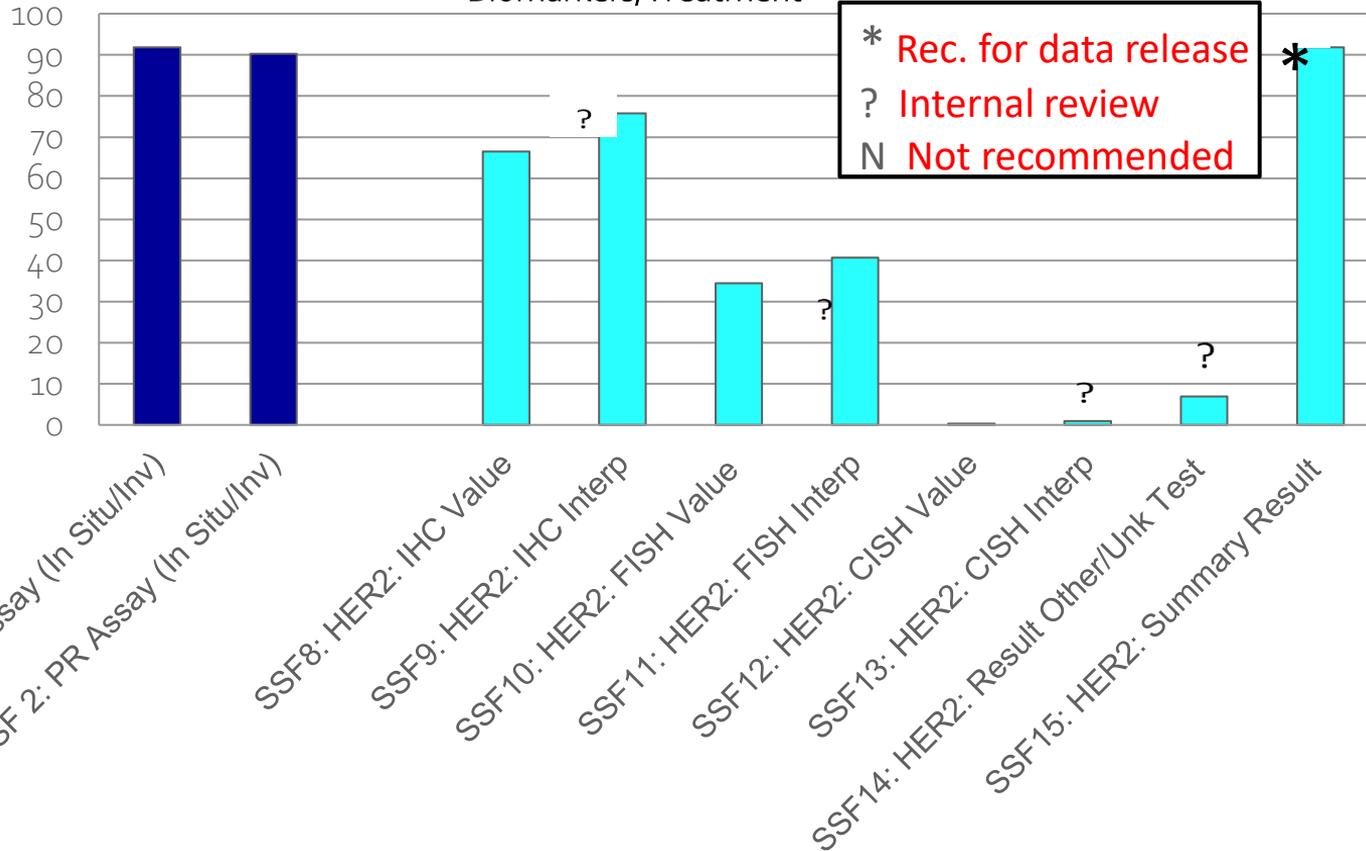
Variables	SSF Description
SSF7	Nottingham or Bloom-Richardson (BR) Score/Grade
SSF8	HER2: IHC Laboratory Value
SSF9	HER2: IHC Test Interpretation
SSF10	HER2: FISH Laboratory Value
SSF 11	HER2: FISH Test Interpretation
SSF 12	HER2: CISH Laboratory Value
SSF 13	HER2: CISH Test Interpretation
SSF 14	HER2: Result of Other or Unknown Test
SSF 15	HER2: Summary Result of Testing ^c
SSF 16	Combinations of ER, PR, and HER2 Results
SSF 17	Circulating Tumor Cells (CTC) and Method of Detection
SSF 18	Disseminated Tumor Cells (DTC) and Method of Detection
SSF 19	Assessment of Positive Ipsilateral Axillary Lymph Nodes
SSF 20	Assessment of Positive Distant Metastases
SSF 21	Response to Neoadjuvant Therapy
SSF 22	Multigene Signature Method
SSF 23	Multigene Signature Results
SSF 24	Paget Disease

% of Cases with Known Value after Filter by Site-Specific Factor Breast Cancer, SEER Combined, 2010



% of Cases with Known Value after Filter by Site-Specific Factor

Biomarkers/Treatment



CS Site specific factors

The image shows a software interface with two panels of variables. The top panel, titled "Extent of Disease", contains the following variables:

- EDD Primary Tumor (2018+)
- EDD Regional Nodes (2018+)
- EDD Mets (2018+)
- Tumor Size Summary (2016+)
- Regional nodes examined (1988+)
- Regional nodes positive (1988+)
- SEER Combined Mets at DX:bone (2010+)
- SEER Combined Mets at DX:brain (2010+)
- SEER Combined Mets at DX:Other (2010+)
- SEER Combined Mets at DX:lung (2010+)
- Mets at DX:Distant LN (2016+)
- Mets at DX:Other (2016+)
- Breast Subtype (2010+)
- ER Status Recode Breast Cancer (1990+)

The bottom panel, titled "Available Variables", contains the following variables:

- ER Status Recode Breast Cancer (1990+)
- PR Status Recode Breast Cancer (1990+)
- Derived HER2 Recode (2010+)
- Lymph-vascular Invasion (2004+ varying by schema)
- CS tumor size (2004-2015)
- CS extension (2004-2015)
- CS lymph nodes (2004-2015)
- CS mets at dx (2004-2015)
- CS Tumor Size/Ext Eval (2004-2015)
- CS Reg Node Eval (2004-2015)
- CS Mets Eval (2004-2015)
- CS site-specific factor 1 (2004-2017 varying by schema)
- CS site-specific factor 2 (2004-2017 varying by schema)
- CS site-specific factor 3 (2004-2017 varying by schema)
- CS site-specific factor 4 (2004-2017 varying by schema)

Both panels include a scroll bar and buttons for "Page", "Row", "Column", and "End..." on the right side. The interface also shows "CAPS" and "NUM" logos in the bottom right corner of each panel.

Breast cancer CS SSFs paper

Overview of Breast Cancer Collaborative Stage Data Items—Their Definitions, Quality, Usage, and Clinical Implications: A Review of SEER Data for 2004-2010

Nadia Howlader, MS¹; Vivien W. Chen, PhD²; Lynn A. G. Ries, MS^{1,3}; Michelle M. Loch, MD⁴; Richard Lee, BS⁵; Carol DeSantis, MPH⁶; Chun Chieh Lin, PhD, MBA⁶; Jennifer Ruhl, MSHCA¹; and Kathleen A. Cronin, PhD, MPH¹

Useful links for SSFs

- SSDIs that are available released in SEER research file (by cancer schema and year of diagnosis)

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

- SEER monograph

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

- Analysis of SSFs –nitty gritty

<https://surveillance.cancer.gov/reports/tech2015.01.pdf>

- SEER staging information

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

Breast cancer Molecular Subtype Analysis in SEER*Stat

Data

Statistic

Selection

Table

Output

Selected Database

Order	5
Location	ssp://seerstat.imsweb.com:2040
Database ID	2137
Link ID	2193
Database Name	Incidence - SEER Research Limited-Field Data, 21 Registries, Nov 2020 Sub (2000-2018)
Linked To	County Attributes - Time Dependent (1990-2018) Income/Rurality, 1969-2019 Counties
Linked By	State-county, Year of diagnosis
Population Variables	Race recode (W, B, AI, AP)

Change Selected Database...

Suggested Citation

Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database: Incidence - SEER Research Limited-Field Data, 21 Registries, Nov 2020 Sub (2000-2018) - Linked To County Attributes - Time Dependent (1990-2018) Income/Rurality, 1969-2019 Counties, National Cancer Institute, DCCPS, Surveillance Research Program, released April 2021, based on the November 2020 submission.

Copy

Ready

Data

Statistic

Selection

Table

Output

Statistic

Frequencies

Trends

Percentages

None

Row

Column

Parameters

P-Values: 0.05 [Edit...](#)

Trend Variable: Year of diagnosis

PC End Points

One Year

Two-Year Average

APC Calculation

Weighted Least Squares

Non-Weighted Least Squares

Show Confidence Interval

Delay Adjust Delay Factor:

Data

Statistic

Selection

Table

Output

Select Only

[View Statements](#)

- Malignant Behavior
- Known Age
- Male or Female Sex
- Cases in Research Database

Selection Statement

```
{Race, Sex, Year Dx.Year of diagnosis} = '2010','2011','2012','2013','2014','2015','2016','2017','2018'  
AND {Site and Morphology.EOD Schema ID Recode (2010+)} = 'Breast'  
AND {Race, Sex, Year Dx.Sex} = ' Female'
```

Edit...

Copy

Clear

- Select only the first Matching Record for each Person

Person Selection

- Use Person Selection

Data
Statistic
Selection
Table
Output

Display Variables

- Page
- Row
 - Breast Subtype (2010+)
- Column

Move Up Move Down Remove

Available Variables

- Age at Diagnosis
- Race, Sex, Year Dx
- Site and Morphology
- Stage - Summary/Historic
- Extent of Disease
 - Breast Subtype (2010+)
- Multiple Primary Fields
- Site and Morphology - Historic
- Race and Age (case data only)
- Site Specific Sequence Numbers
- Other
- County attributes - Time Dependent
- User-Defined

Page Row Column Find...

Data

Statistic

Selection

Table

Output

Title (Up to 5 Lines)

SEER 21, 2010-2018 Malignant Female Breast Cancer by SubType

Options

Number of Decimal Places for Trends/Percentages:

0.1

Set Default

Hide Statistics When Fewer Than

25

Cases

Include Individual Year Counts

Matrix

Settings

Graph

SEER 21, 2010-2018 Malignant Female Breast Cancer by SubType

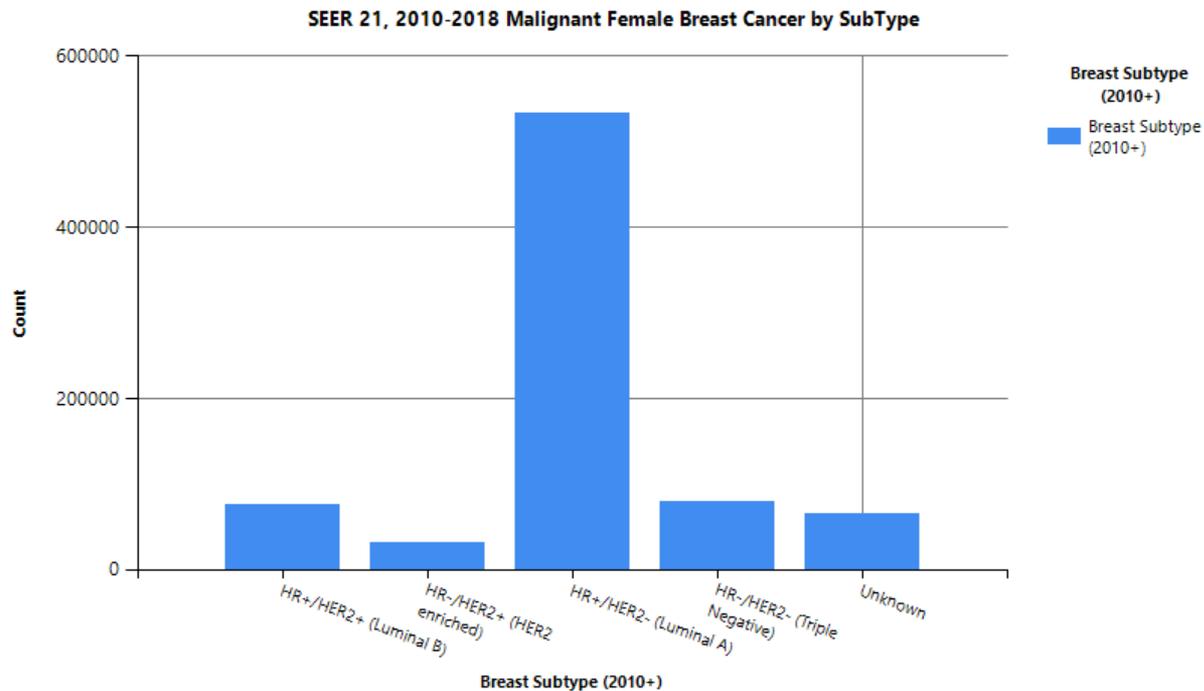
	Count	Column %	Cum %
HR+/HER2+ (Luminal B)	75,905	9.7%	9.7%
HR-/HER2+ (HER2 enriched)	32,205	4.1%	13.7%
HR+/HER2- (Luminal A)	533,629	67.9%	81.6%
HR-/HER2- (Triple Negative)	79,478	10.1%	91.7%
Unknown	65,237	8.3%	100.0%
Recode not available	0	0.0%	100.0%
Total	786,454	100.0%	100.0%

Matrix

Settings

Graph

Graph Type **Column** Statistic **Count** X-Axis **Breast Subtype (2010+)**
By-Var (Series) Chart Variables **None** Chart 0 of 0





**NATIONAL
CANCER
INSTITUTE**

www.cancer.gov

www.cancer.gov/espanol