

## Feature Mapping

Paper Specification Text <sup>1</sup>	SEER*Stat text <sup>2</sup>	match	reason if dropped, modified, or added	variable type	reason if excluded
Patient ID	Patient ID	yes		removed	irrelevancy
SEER registry		drop	confidentiality, data item not available at individual level	categorical	
Marital status at diagnosis	Marital status at diagnosis	yes		categorical	
Race/ethnicity	Race/ethnicity	yes		categorical	
NHIA Derived Hisp Origin		drop	Originally coded 0-8 for different hispanic countries. Not available in 2023 database.	categorical	
Sex	Sex	yes		categorical	
Age at diagnosis	Age recode with single ages and 90+	mod	originally capped at 130, best available fit is "Single ages and 90+"	continuous	
Year of birth		calc	calculate as (age at diagnosis - year of diagnosis)	continuous	
Sequence number	Sequence number	yes		categorical	
Month of diagnosis		drop	confidentiality, data item not available at individual level	categorical	
Year of diagnosis	Year of diagnosis	yes		continuous	
Primary site ICD-O-2 (1973+)	Primary Site	yes		categorical	
Laterality	Laterality	yes		categorical	
Histologic Type ICD-O-2		drop	entire "Site and Morphology - Historic" category. Not available in 2023 database.	categorical	
Behavior Code ICD-O-2*/		drop	entire "Site and Morphology - Historic" category. Not available in 2023 database.	categorical	
Histologic Type ICD-O-3	Histologic Type ICD-O-3	yes		categorical	
Behavior code ICD-O-3	Behavior code ICD-O-3	yes		categorical	
Grade	Grade Recode (thru 2017)	yes		categorical	
Diagnostic confirmation	Diagnostic Confirmation	yes		removed	post-dx
Type of reporting source	Type of Reporting Source	yes		categorical	
EOD 10 - size (1988+)	EOD 10 - size (1988-2003)	yes	FYI, This field is only coded for cases diagnosed 1988-2003. For 2004-2015, similar type of information was collected in CS Tumor Size (2004-2015) in the collaborative stage variables.	categorical	
EOD 10 - extension	EOD 10 - extent (1988-2003)	yes	This field is only coded for cases diagnosed 1988-2003. For 2004-2015, similar type of information was collected in CS extension (2004-2015) and CS Mets at DX (2004-2015) in the collaborative stage variables.	categorical	
EOD 10 - path extension	EOD 10 - Prostate path ext (1995-2003)	yes	This field is only coded for prostate cases diagnosed 1995-2003. For 2004+, similar type of information was collected in CS SSF 3 in the collaborative stage variables	categorical	
EOD 10 - lymph node	EOD 10 - nodes (1988-2003)	yes	This field is only coded for cases diagnosed 1988-2003. For 2004-2015, similar type of information was collected in CS Lymph Nodes (2004-2015) in the collaborative stage variables.	categorical	
EOD 10 - positive lymph nodes examined	Regional nodes positive (1988+)	yes		continuous	
EOD 10 - number of lymph nodes examined	Regional nodes examined (1988+)	yes		continuous	
EOD--old 13 digit		drop	compound information, only valid 1973-1982	removed	compound info
EOD--old 2 digit		drop	compound information, only valid 1973-1982	removed	compound info
EOD--old 4 digit		drop	compound information, only valid 1973-1982	removed	compound info
Coding system for EOD	Coding system-EOD (1973-2003)	yes		categorical	
Tumor marker 1	Tumor marker 1 (1990-2003)	yes		categorical	
Tumor marker 2	Tumor marker 2 (1990-2003)	yes		categorical	
Tumor marker 3	Tumor marker 3 (1998-2003)	yes		categorical	
CS Tumor size	CS tumor size (2004-2015)	yes		continuous	
CS Extension	CS extension (2004-2015)	yes		categorical	
CS Lymph Nodes	CS lymph nodes (2004-2015)	yes		categorical	
CS Mets at DX	CS mets at dx (2004-2015)	yes		categorical	
CS Site-Specific Factor 1	CS site-specific factor 1 (2004-2017 varying by schema)	yes		categorical	

CS Site-Specific Factor 2	CS site-specific factor 2 (2004-2017 varying by schema)	yes		categorical	
CS Site-Specific Factor 3	CS site-specific factor 3 (2004-2017 varying by schema)	yes		categorical	
CS Site-Specific Factor 4	CS site-specific factor 4 (2004-2017 varying by schema)	yes		categorical	
CS Site-Specific Factor 5	CS site-specific factor 5 (2004-2017 varying by schema)	yes		categorical	
CS Site-Specific Factor 6	CS site-specific factor 6 (2004-2017 varying by schema)	yes		categorical	
CS Site-Specific Factor 25	CS site-specific factor 25 (2004-2017 varying by schema)	yes		categorical	
Derived AJCC T	Derived AJCC T, 6th ed (2004-2015)	yes		categorical	
Derived AJCC N	Derived AJCC N, 6th ed (2004-2015)	yes		categorical	
Derived AJCC M	Derived AJCC M, 6th ed (2004-2015)	yes		categorical	
Derived AJCC Stage Group	Derived AJCC Stage Group, 6th ed (2004-2015)	yes		categorical	
Derived SS1977		drop	same as 2016's "SEER summary stage 1977". Not available in 2023 database. Only applies to data before 2000	categorical	
Derived SS2000	SEER Combined Summary Stage 2000 (2004-2017)	mod	"For 2004-2015 cases, it is copied from Derived SS2000" (in 2016 same as "SEER summary stage 2000")	categorical	
Derived AJCC - flag		drop	always 1 after 2004. Not available in 2023 database.	categorical	
CS Version Input Original	CS version input original (2004-2015)	yes		categorical	
CS Version Derived	CS version derived (2004-2015)	yes		categorical	
CS Version Input Current	CS version input current (2004-2015)	yes		categorical	
RX Summ--surg prim site	RX Summ--Surg Prim Site (1998+)	yes		removed	treatment info
RX Summ--scope reg LN sur 2003+*/	RX Summ--Scope Reg LN Sur (2003+)	yes		removed	treatment info
RX Summ--surg oth reg/dis	RX Summ--Surg Oth Reg/Dis (2003+)	yes		removed	treatment info
Number of lymph nodes	RX Summ--Reg LN Examined (1998-2002)	yes		removed	treatment info
Reason no cancer-directed surgery	Reason no cancer-directed surgery	yes		removed	treatment info
Site specific surgery (1983-1997)	Site specific surgery (1973-1997 varying detail by year and site)	yes		removed	treatment info
Scope of lymph node surgery 98-02	Scope of reg lymph nd surg (1998-2002)	yes		removed	treatment info
Surgery to other sites	Surgery of oth reg/dis sites (1998-2002)	yes		removed	treatment info
Record number	Record number recode	mod	some small changes to order of sequential number	removed	irrelevancy
Type of followup expected		drop	Not available in 2023 database	removed	irrelevancy
Age recode <1 year olds	Age recode with <1 year olds	yes		continuous	
Site recode ICD-O-3/WHO 2008	Site recode ICD-O-3/WHO 2008	yes		categorical	
Recode ICD-O-2 to 9		drop	entire "Site and Morphology - Historic" category dropped. Not available in 2023 database.	categorical	
Recode ICD-O-2 to 10		drop	entire "Site and Morphology - Historic" category dropped. Not available in 2023 database.	categorical	
ICCC site recode ICD-O-3/WHO 2008	ICCC site recode 3rd edition/IARC 2017	mod	switched from WHO 2008 definitions to IARC 2017 definitions, same type of information but different values	categorical	
ICCC site rec extended ICD-O-3/ WHO 2008	ICCC site recode extended 3rd edition/IARC 2017	mod	switched from WHO 2008 definitions to IARC 2017 definitions, same type of information but different values	categorical	
Behavior recode for analysis	Behavior recode for analysis	yes		categorical	
Broad Histology recode	Histology recode - broad groupings	yes		categorical	
Brain recode	SEER Brain and CNS Recode	mod	same type of information but different values	categorical	
CS Schema v0204*/	TNM 7/CS v0204+ Schema (thru 2017)	yes		categorical	
Race recode A	Race recode (White, Black, Other)	yes		categorical	
Race recode Y	Race recode (W, B, AI, API)	yes		categorical	
Origin Recode NHIA	Origin recode NHIA (Hispanic, Non-Hisp)	yes		categorical	
SEER historic stage A	SEER historic stage A (1973-2015)	yes		categorical	
AJCC stage 3rd edition (1988+)	AJCC stage 3rd edition (1988-2003)	yes		categorical	
SEER modified AJCC stage 3rd ed (1988+)	SEER modified AJCC stage 3rd (1988-2003)	yes		categorical	
SEER Summary Stage 1977 (1995-2000)		drop	same as 2016's "Derived SS1977". Not available in 2023 database	categorical	
SEER Summary Stage 2000 2000 (2001-2003)	Combined Summary Stage (2004+)	mod	"Combined Summary Stage field to facilitate stage analyses over time."	categorical	
First malignant primary indicator	First malignant primary indicator	yes		categorical	
State-county recode		drop	confidentiality, data item not available at individual level, replaced with "County Attributes - Time Dependent" features	categorical	
Cause of death to SEER site recode	COD to site recode	yes		removed	post-dx
COD to site rec KM	COD to site rec KM	yes		removed	post-dx

Vital status recode (study cutoff used)	Vital status recode (study cutoff used)	yes		removed	post-dx
IHS link	IHS Link	yes		categorical	
Historic SSG 2000 Stage	Summary stage 2000 (1998-2017)	yes		categorical	
AYA site recode/WHO 2008	AYA site recode 2020 Revision	mod	updated revision, same type of information but different values	categorical	
Lymphoma subtype recode/WHO 2008	Lymphoid neoplasm recode 2021 Revision	mod	updated revision, similar except doesn't incorporate grade	categorical	
SEER cause of death classification	SEER cause-specific death classification	yes		target	
SEER other cause of death classification	SEER other cause of death classification	yes		removed	post-dx
CS EXT/Size Eval	CS Tumor Size/Ext Eval (2004-2015)	yes		removed	treatment
CS Nodes Eval	CS Reg Node Eval (2004-2015)	yes		removed	treatment
CS Mets Eval	CS Mets Eval (2004-2015)	yes		removed	treatment
Primary by International Rules	Primary by international rules	yes		categorical	
ER Status Recode Breast Cancer (1990+)* /	ER Status Recode Breast Cancer (1990+)	yes		categorical	
PR Status Recode Breast Cancer (1990+)* /	PR Status Recode Breast Cancer (1990+)	yes		categorical	
CS Schema - AJCC 6th Edition	CS Schema - AJCC 6th Edition	yes		categorical	
Cs Site-specific Factor 8	CS site-specific factor 8 (2004-2017 varying by schema)	yes		categorical	
CS Site-Specific Factor 10* /	CS site-specific factor 10 (2004-2017 varying by schema)	yes		categorical	
CS Site-Specific Factor 11* /	CS site-specific factor 11 (2004-2017 varying by schema)	yes		categorical	
CS Site-Specific Factor 13* /	CS site-specific factor 13 (2004-2017 varying by schema)	yes		categorical	
CS Site-Specific Factor 15* /	CS site-specific factor 15 (2004-2017 varying by schema)	yes		categorical	
CS Site-Specific Factor 16* /	CS site-specific factor 16 (2004-2017 varying by schema)	yes		categorical	
Lymph-vascular Invasion (2004+)* /	Lymph-vascular Invasion (2004+ varying by schema)	yes		categorical	
Survival months	Survival months	yes		target	
Survival months flag	Survival months flag	yes		removed	irrelevancy
Insurance Recode (2007+)		drop	Caution is urged when using this variable due to Medicare eligibility. Not available in 2023 database.	categorical	
Derived AJCC T 7th ed	Derived AJCC T, 7th ed (2010-2015)	yes		categorical	
Derived AJCC N 7th ed	Derived AJCC N, 7th ed (2010-2015)	yes		categorical	
Derived AJCC M 7th ed	Derived AJCC M, 7th ed (2010-2015)	yes		categorical	
Derived AJCC 7 Stage Group	Derived AJCC Stage Group, 7th ed (2010-2015)	yes		categorical	
Adjusted AJCC 6th T (1988+)	Breast - Adjusted AJCC 6th T (1988-2015)	yes	only differ 126/Blank added for data after 2015	categorical	
Adjusted AJCC 6th N (1988+)	Breast - Adjusted AJCC 6th N (1988-2015)	yes	only differ 126/Blank added for data after 2015	categorical	
Adjusted AJCC 6th M (1988+)	Breast - Adjusted AJCC 6th M (1988-2015)	yes	only differ 126/Blank added for data after 2015	categorical	
Adjusted AJCC 6th Stage (1988+)	Breast - Adjusted AJCC 6th Stage (1988-2015)	yes	only differ 126/Blank added for data after 2015	categorical	
CS Site-Specific Factor 7	Adjusted CS site-specific factor 7 (2004-2017 varying by schema)	yes	no immediately apparent differences	categorical	
CS Site-specific Factor 9	CS site-specific factor 9 (2004-2017 varying by schema)	yes		categorical	
CS Site-Specific Factor 12	CS site-specific factor 12 (2004-2017 varying by schema)	yes		categorical	
Derived HER2 Recode (2010+)	Derived HER2 Recode (2010+)	yes		categorical	
Breast Subtype (2010+)	Breast Subtype (2010+)	yes		categorical	
Lymphoma - Ann Arbor Stage (1983+)	Lymphoma - Ann Arbor Stage (1983-2015)	yes		categorical	
CS mets at DX-bone (2010+)	SEER Combined Mets at DX-bone (2010+)	yes	only differ 14/Blank added	categorical	
CS mets at DX-brain (2010+)	SEER Combined Mets at DX-brain (2010+)	yes	only differ 14/Blank added	categorical	
CS mets at DX-liver (2010+)	SEER Combined Mets at DX-liver (2010+)	yes	only differ 14/Blank added	categorical	
CS mets at DX-lung (2010+)	SEER Combined Mets at DX-lung (2010+)	yes	only differ 14/Blank added	categorical	
T value - based on AJCC 3rd (1988-2003)	T value - based on AJCC 3rd (1988-2003)	yes		categorical	
N value - based on AJCC 3rd (1988-2003)	N value - based on AJCC 3rd (1988-2003)	yes		categorical	
M value - based on AJCC 3rd (1988-2003)	M value - based on AJCC 3rd (1988-2003)	yes		categorical	
Total number of in situ/malignant tumors for patient	Total number of in situ/malignant tumors for patient	yes		removed	post-dx
Total number of benign/borderline tumors for patient	Total number of benign/borderline tumors for patient	yes		removed	post-dx
	Median household income inflation adj to 2022	add	added for relevant information to replace 'State-county recode'	categorical	
	Rural-Urban Continuum Code	add	added for relevant information to replace 'State-county recode'	categorical	

<sup>1</sup><https://seer.cancer.gov/data-software/documentation/seerstat/nov2016/TextData.FileDescription.pdf>

<sup>2</sup><https://seer.cancer.gov/data-software/documentation/seerstat/nov2023/TextData.FileDescription-nov2023.pdf>