

Documentation for the Patient Entitlement and Diagnosis Summary File (PEDSF)
November 9, 2016

FILENAME 1: pedsf.cancer.txt
LRECL 1: 5,442
FILENAME 2: pedsf.5percent.txt (if requested)
LRECL 2: 2,184

SEER Cases diagnosed 1973-2013

PEDSF File created on October 4, 2016

SEER data extracted from November, 2015 SEER Submission

ABSTRACT: The PEDSF file contains one record per person for individuals in the SEER Program data base who have been matched with Medicare enrollment records. Of persons who were reported by the SEER Registries to have been diagnosed with cancer at age 65 or older, 94 percent were matched with Medicare enrollment records. For persons appearing in the PEDSF file, basic SEER Program diagnostic information is available for up to 10 diagnosed cancer occurrences. Data also include Medicare entitlement and utilization data from 1991-2015.

PROGRAMMER: Angela Meekins, Information Management Services, Inc.

CONTACT: SEER-Medicare Technical Support, Information Management Services, Inc.
Email: SEER-Medicare@imsweb.com

NOTE1: With Linkage 2014 PEDSF no longer includes SEER cases from Los Angeles and San Jose from 1988 thru 1991. LA and SJ now include data from 1992 on.

NOTE2: For all non-custom requests, patients have been selected if the cancer(s) requested appeared in the site array (siterwho1-10) and the associated year of diagnosis (yrdx1-10) fell in the year of diagnosis window requested.

NOTE3: With Linkage 2014 the PEDSF file layout has changed, census data is no longer included on the PEDSF file but will be sent out in separate files for census tract and zipcode census data. Part D information has been limited to only Plan Coverage and Dual Eligible Status information. In addition the SEER information on the PEDSF file has been rearranged to mimic the SEER research file and the site variables is now 5 digits long instead of 2.

NOTE4: If you are using old data and link-up the data for **Louisiana** patients using the "oshufcase" it's important to note that some old IDs are repeated as current IDs for different patients. It is suggested that after patients are matched on "oshufcase" they should use the current shufcase so duplicate IDs won't lead to different patients.

Documentation for the Patient Entitlement and Diagnosis Summary File (PEDSF)
November 9, 2016

NOTE5: A recent evaluations of the SEER data have uncovered problems with the quality of the PSA value and PSA interpretation variables. As data quality is a primary concern, PSA values have been removed from the SEER data files. Thereby, updated SEER-Medicare data no longer includes PSA values. SEER-Medicare Data Users are advised not to use the PSA values that were previously released. NCI is currently in the process of investigating these problems further both centrally and at the SEER registries. For additional information on the PSA variable, see <http://seer.cancer.gov/data/psa-values.html> or contact seerstat@imsweb.com

Documentation for the Patient Entitlement and Diagnosis Summary File (PEDSF)
November 9, 2016

<u>COL</u>	<u>FIELD</u>	<u>LENGTH</u>	<u>SOURCE</u>	<u>NOTES</u>
1	Patient ID (patient_id)	10	Master File	
1	SEER Cases - Registry	2	SEER	First two digits in the patient ID. 02 = Connecticut (1973+) 20 = Detroit (1973+) 21 = Hawaii (1973+) 22 = Iowa (1973+) 23 = New Mexico (1973+) 25 = Seattle (1974+) 26 = Utah (1973+) 42 = Kentucky (2000+) 43 = Louisiana (2000+) 44 = New Jersey (2000+) 87 = Georgia 88 = California NOTE: Specific California and Georgia registries will be differentiated in the repeated registry codes (starting in column 2202, variables reg1-reg10)
3	Case Number	8	Assigned by IMS	Encrypted SEER Case Number
11	Five Pct Indicator (fivepct)	1	Master File	Indicates inclusion in the 5% Medicare sample. Y = Included N = Not included
12	Match Type (mat_type)	1	Match File	CMSs matching algorithm is available at IMS. A = Alphabetic Match B = SSN Match
13	Match Flag 1 (yobflg1)	1	Match File	Year of Birth flag for SSN and Alpha Match. Y = Agreement between SEER and Medicare N = Did not agree 0 (Zero) = Cases matched on SSN U = Unknown, Missing data
14	Match Flag 2 (dobflg2)	1	Match File	Day of Birth Match for SSN and Alpha Match (See Match Flag 1 for code definition)

Documentation for the Patient Entitlement and Diagnosis Summary File (PEDSF)
November 9, 2016

<u>COL</u>	<u>FIELD</u>	<u>LENGTH</u>	<u>SOURCE</u>	<u>NOTES</u>
15	Match Flag 3 (yodflg3)	1	Match File	Month and Year of Death for Alpha Match. (See Match Flag 1 for code definition)
17	Match Flag 5 (lstflg5)	1	Match File	Last Name Flag for SSN Match. (See Match Flag 1 for code definition)
18	Match Flag 6 (fstflg6)	1	Match File	First Name Flag for SSN Match. (See Match Flag 1 for code definition)
19	Match Flag 7 (mobflg7)	1	Match File	Month of Birth Flag for SSN Match. (See Match Flag 1 for code definition)
20	Match Flag 8 (midflg8)	1	Match File	Middle Initial Flag for Alpha Match. (See Match Flag 1 for code definition)
21	Number of digits in SSN that matched (numdigit)	1	Match File	Alpha search only. 0-8 9 = SSN match U = Unknown; data missing A = digits transposed and matched
22	Date of Death Flag (dod_flg)	1	Created at IMS	Shows the level of agreement between SEER and MEDICARE on the patient's month of death. Dates of death after 12/13 were treated as not dead for comparison purposes. 0 = Not dead by 12/13. 1 = Dead, both files agree. 2 = Dead, off by 1-3 months. 3 = Dead, off by 4-6 months. 4 = Dead in MEDICARE only, 5 = Dead in SEER only. 6 = Dead but number of months could not be calculated due to the fact that month was missing for either SEER or MEDICARE.
23	Date of Birth Flag (dob_flg)	1	Created at IMS	Shows the level of agreement between SEER and MEDICARE on the patient's month of birth. 0 = Both files agree on birth date. 1 = Birth date off by 1-3 months. 2 = Birth date off by 4-6 months. 3 = Birth date off by 7-11 months. 4 = Birth date off by one year 5 = Birth date off by 13-23 months 6 = Birth date off by 2 years

Documentation for the Patient Entitlement and Diagnosis Summary File (PEDSF)
November 9, 2016

<u>COL</u>	<u>FIELD</u>	<u>LENGTH</u>	<u>SOURCE</u>	<u>NOTES</u>
				7 = Birth date off by 25-35 months 8 = Birth date off by 3 years 9 = Birth date off by 37+ months Blank = Birth date is missing
24	Linkage Flag (linkflag)	1	Created at IMS	Indicates the linkage the patient is first associated with. 1 = 2005 linkage, earliest claim year is 1991 2 = 2008 linkage, earliest claim year is 1998 3 = 2010 linkage, earliest claim year is 2000 4 = 2012 linkage, earliest claim year is 2002 5 = 2014 linkage, earliest claim year is 2004 6 = 2016 linkage, earliest claim year is 2006
25	Medicare Date of Death (med_dodm, med_dodd, med_dody)	8	EDB	MMDDYYYY or blanks for not dead. Reported to CMS by SSA – not date of last contact. Date complete through 12/31/14
33	Medicare Month of Birth (birthm)	2	EDB	MM, zeroes for invalid
37	Medicare Year of Birth (birthyr)	4	EDB	YYYY, zeroes for invalid
41	Sex (m_sex)	1	EDB	1 = Male 2 = Female
42	Race (race)	1	EDB	1 = White 2 = Black 3 = Other 4 = Asian 5 = Hispanic 6 = N. Am. Native 0 = Unknown
43	Original Reason for Entitlement (rsncd1)	1	EDB	0 = OASI (age) 1 = DIB (disability) 2 = ESRD 3 = DIB/ESRD This is the first entitlement reason located on EDB

Documentation for the Patient Entitlement and Diagnosis Summary File (PEDSF)
November 9, 2016

<u>COL</u>	<u>FIELD</u>	<u>LENGTH</u>	<u>SOURCE</u>	<u>NOTES</u>
44	Current Reason for Entitlement (cur_ent)	1	EDB	Blank = Not Enrolled 0 = OASI (age) 1 = DIB 2 = ESRD 3 = DIB/ESRD Variable is selected by the first diagnosis date at age 65 or older, or the last diagnosis if all are prior to turning age 65. If information not available at those times then the first available data is used.
45	Current Reason for Entitlement Year (cur_yr)	4	EDB	YYYY
49	Chronic Renal Disease (chr_esrd)	1	EDB	0 (zero) = No ESRD Y = Source of ESRD Variable is set to zero. If there is a start date before the diagnosis date at age 65 or older, or the last diagnosis if all are prior to turning age 65 then it is set to Y.
50	Chronic Renal Disease Year (chr_esrd_yr)	4	EDB	YYYY
54	First Chronic Renal Disease Year (first_esrd_yr)	4	EDB	First occurrence of Chronic Renal Disease regardless of age.
58	Medicare Status Code (med_stcd)	2	EDB	Blank = Not Enrolled 10 = Aged 11 = Aged with ESRD 20 = Disabled 21 = Disabled with ESRD 31 = ESRD only Variable is selected by the first diagnosis date at age 65 or older, or the last diagnosis if all are prior to turning age 65. If information not available at those times then the first available data is used.
60	Medicare Status Code Year (medst_yr)	4	EDB	YYYY
64	Valid Date of Death (vrfydth)	1	EDB	N = No Y = Yes

Documentation for the Patient Entitlement and Diagnosis Summary File (PEDSF)
November 9, 2016

<u>COL</u>	<u>FIELD</u>	<u>LENGTH</u>	<u>SOURCE</u>	<u>NOTES</u>
<u>Geographic Information</u>				
SEER program data are from the SEER record corresponding to the first diagnosis of cancer at age 65 or older. Zip codes are from Medicare enrollment file in the year of first diagnosis at age 65 or older or the last diagnosis if never 65.				
65	State (state)	2	SEER	FIPS State Code
67	County (county)	3	SEER	FIPS County Code
70	Zip Code (zip5)	5	EDB	Encrypted Zip code, 5 digits. * Special permission required for unencrypted zip code.
75	Zip Code (zip4)	4	EDB	Last four digits of zip code. Blanked out when Encrypted Zip code is given. * Special permission required for unencrypted zip code.
79	Census Tract Flag (code_sys)	1	SEER	Use in conjunction with Census Tract (tract) 0 = Not tracted 1 = 1970 Census tract definitions 2 = 1980 Census tract definitions 3 = 1990 Census tract definitions
80	Census Tract (tract1990)	6	SEER	1970/80/90 Encrypted census tract * Special permission required for unencrypted census tract.
86	Census Tract 2000 (tract2000)	6	SEER	2000 encrypted census tract * Special permission required for unencrypted census tract.
92	Census Tract 2010 (tract2010)	6	SEER	2010 encrypted census tract * Special permission required for unencrypted census tract.
98	HSA (hsa)	3	ARF	Health Service Area. Taken from the 2004 Area Resource File (ARF).
101	Urban/Rural recode (urbrur)	1	ARF	Urban/Rural Code 1 = Big Metro (Urban = 00 or 01) 2 = Metro (Urban = 02 or 03) 3 = Urban (Urban = 04 or 05) 4 = Less Urban (Urban = 06 or 07) 5 = Rural (Urban = 08 or 09) 9 = Unknown (Urban = 99)
102	Urban/Rural code	2	ARF	01-09, 99 see table at end of this

Documentation for the Patient Entitlement and Diagnosis Summary File (PEDSF)
November 9, 2016

<u>COL</u>	<u>FIELD</u>	<u>LENGTH</u>	<u>SOURCE</u>	<u>NOTES</u>
	(urban)			document
	<u>Demographic Information</u>			
	The following information comes from the SEER record corresponding to the first diagnosis at age 65 or over, or the latest diagnosis if all are prior to turning age 65.			
104	Sex (s_sex)	1	SEER	See Attachment A
105	Race Recode B (rac_recb)	2	SEER	Race Recode B from SEER 01,11 = White 01 = Caucasian, NOS 02 = Black 03 = American Indian/Alaska Native 04 = Chinese 05 = Japanese 06 = Filipino 07 = Hawaiian 08 = Other Asian or Pac. Islander 09 = Unknown 11 = Caucasian, Spanish origin or surname 12 = Other unspecified (1991+)
107	Race Recode Y (rac_recy)	1	SEER	See Attachment A
108	Race Recode A (rac_reca)	1	SEER	See Attachment A
109	ICD Code – Cause of Death (icd_code)	1	SEER	0 = Patient is alive at last follow-up 1 = Tenth ICD revision 8 = Eighth ICD Revision 9 = Ninth ICD revision
110	Cause of Death ICD -8 or 9 (cod89v)	4	SEER	0000 = Alive at last contact 7777 = State Death Certificate not available. 7797 = Death Certificate available, no COD listed.
114	Cause of Death ICD-10 (cod10v)	4	SEER	A020-Y891 0000 = Alive at last contact 7777 = State Death Certificate not available. 7797 = Death Certificate available, no COD listed.

Documentation for the Patient Entitlement and Diagnosis Summary File (PEDSF)
November 9, 2016

<u>COL</u>	<u>FIELD</u>	<u>LENGTH</u>	<u>SOURCE</u>	<u>NOTES</u>
118	Cause of Death to site recode KM (codkm)	5	SEER	This is a recode based on underlying cause of death to designate cause of death into groups similar to the incidence site recode with KS and mesothelioma. Study cutoff date has been applied, i.e. coded as alive if death occurred after study cutoff. Go to the end of this document for a listing of codes and their definition.
123	Cause of Death to site recode (codpub)	5	SEER	This recode was introduced to account for several newly valid ICD-10 codes and includes both cancer and non-cancer causes of death. Go to the end of this document for a listing of codes and their definition.
128	NHIA Derived Hispanic Origin (nhiade)	1	SEER	See Attachment A
129	SEER Month of Death (ser_dodm)	2	SEER	00 = Alive blank = Unknown month Date complete through 12/31/13
131	SEER Year of Death (ser_dody)	4	SEER	0000 = Alive 2053 = Unknown year Date complete through 12/31/13
135	Date Flag for Follow Up (deathflag)	2	SEER	0 = Alive
137	SEER Race/Ethnicity (srace)	2	SEER	See Attachment A
139	SEER Hispanic Surname (origin)	1	SEER	0 = Non-Spanish/Non-Hispanic 1 = Mexican (includes Chicano) 2 = Puerto Rican 3 = Cuban 4 = South or Central American (except Brazil) 5 = Other specified Spanish/Hispanic origin (includes European; excludes Dominican Republic) 6 = Spanish, NOS; Hispanic, NOS; Latino, NOS (There is evidence, other than surname or maiden name, that the person is Hispanic but he/she cannot be assigned to any of the

Documentation for the Patient Entitlement and Diagnosis Summary File (PEDSF)
November 9, 2016

<u>COL</u>	<u>FIELD</u>	<u>LENGTH</u>	<u>SOURCE</u>	<u>NOTES</u>
				categories 1-5). 7 = Spanish surname only (effective with diagnosis on or after 1/1/1994) (The only evidence of the person's Hispanic origin is the surname or maiden name and there is no contrary evidence that the patient is not Hispanic.) 8 = Dominican Republic (effective with diagnosis on or after 1/1/2005) 9 = Unknown whether Spanish/Hispanic or not NOTE: codes 1-4 for 1988+, code 7 for 1994+
140	Origin recode NHIA (Hispanic, Non-Hisp) (origrecb)	1	SEER	See Attachment A
141	Filler	1		
142	Vital Status Recode (stat_rec)	1	SEER	See Attachment A
143	Census Tract Certainty (cen_cert)	1	SEER	Associated with 1970/80/90 Census data 1 = Census tract based on complete and valid street address of residence 2 = Census tract based on residence ZIP+4 3 = Census tract based on residence ZIP+2 4 = Census tract based on residence ZIP code only 5 = Census tract based on ZIP code of post office box 6 = Census tract/BNA based on residence city where city has only one census tract, or based on residence ZIP code where ZIP code has only one census tract 9 = Unable to assign census tract based on available information
144	Census Tract Certainty 2000 (ctcer2k)	1	SEER	Associated with 2000 Census data 1 = Census tract based on complete and valid street address of residence 2 = Census tract based on residence

Documentation for the Patient Entitlement and Diagnosis Summary File (PEDSF)
November 9, 2016

<u>COL</u>	<u>FIELD</u>	<u>LENGTH</u>	<u>SOURCE</u>	<u>NOTES</u>
				ZIP+4 3 = Census tract based on residence ZIP+2 4 = Census tract based on residence ZIP code only 5 = Census tract based on ZIP code of post office box 6 = Census tract/BNA based on residence city where city has only one census tract, or based on residence ZIP code where ZIP code has only one census tract 9 = Unable to assign census tract based on available information
145	Census Tract Certainty 2010 (ctcer2010)	1	SEER	Associated with 2010 Census data 1 = Census tract based on complete and valid street address of residence 2 = Census tract based on residence ZIP+4 3 = Census tract based on residence ZIP+2 4 = Census tract based on residence ZIP code only 5 = Census tract based on ZIP code of post office box 6 = Census tract based on residence city where city has only one census tract, or based on residence ZIP code where ZIP code has only one census tract 9 = Unable to assign census tract based on available information
146	Census Tract Poverty Indicator (census_pov_ ind)	1	SEER	1= 0%-<5% poverty 2 = 5% to <10% poverty 3 = 10% to <20% poverty 4 = 20% to 100% poverty 9 = Unknown
147	SEER Year of Birth (yr_brth)	4	SEER	See Attachment A
151	Date of Birth Flag (dbrflag)	2	SEER	12 = A proper value is applicable but not known.
153	Number of SEER	2	Created	Number of eligible SEER records. SEER

Documentation for the Patient Entitlement and Diagnosis Summary File (PEDSF)
November 9, 2016

<u>COL</u>	<u>FIELD</u>	<u>LENGTH</u>	<u>SOURCE</u>	<u>NOTES</u>
	records (count)		at IMS	records prior to 1992 for Rural Georgia and records prior to 2000 for Greater California, New Jersey, Kentucky, and Louisiana were not included in PEDSF.
157	Diagnosis indicator (resnrec)	1	Created at IMS	0 = Last Dx; Patient always less than 65. 1 = First Dx at age 65 or later.
164	Old Encrypted Casenum (oshufcase)	8	IMS	The encrypted case number for Louisiana and Seattle in the 2008 linkage.
172	Filler	53		
<u>Repeated Part D Denominator Information</u>				
These variables (columns 225-264) are repeated once for every year from 2007 to 2014. (YY = the year of the file, ex. 07)				
225	Plan-Value Indicators (Jan. – Dec.) (planYY_01- planYY_12)	12*1	Part D Denom	<p>12 monthly 1-byte indicators Indicates for each month of the Denominator reference year, the Part D enrollment, which is based on the 1st digit or the Part D contract number.</p> <p>Recodes only indicating type of plan; no 5-digit plan numbers. 0 = Not Medicare enrolled for the month * = Enrolled in Medicare A and/or B, but no MIIR record for the month H = Managed Care Organizations other than Regional PPO R = Regional PPO S = PDP N = Not Part D Enrolled E = Employer-sponsored (starting in Jan. 2007)</p>
237	State Reported Dual Eligible Status Code (Jan. – Dec.) (dualYY_01- dualYY_12)	12*2	Part D Denom	<p>12 monthly 2-byte indicators Indicates for each month of the Denominator reference year, the dual eligibility status, if any, for the beneficiary.</p> <p>00 = Not Medicare enrolled for the month ** = Enrolled in Medicare A and/or B, but no MIIR record for the month NA = Non-Medicaid 01 = QMB only 02 = QMB and Medicaid coverage including RX</p>

Documentation for the Patient Entitlement and Diagnosis Summary File (PEDSF)
November 9, 2016

<u>COL</u>	<u>FIELD</u>	<u>LENGTH</u>	<u>SOURCE</u>	<u>NOTES</u>
				03 = SLMB only 04 = SLMB and Medicaid coverage including RX 05 = QDWI 06 = Qualifying Individuals 08 = Other Dual Eligibles (Non-QMB, SLMB, QWDI, or QI) w/Medicaid coverage including RX 09 = Other Dual Eligibles but without Medicaid coverage 99 = Unknown
261	Plan Coverage Months (ptdYY)	2	Part D Denom	Contains the total number of months of Part D plan coverage for the beneficiary. The value in this field will be within the valid range of values 00 through 12, inclusive, dependent on the number of occurrences when the Plan Indicators = H, R, S, or E.
263	Medicaid Dual Eligible Months (dualcntYY)	2	Part D Denom	Contains the total numbers of months of dual eligibility for the beneficiary. The value in this field will be within the valid range of values 00 through 12, inclusive, dependent on the number of occurrences when the Medicaid Dual Eligible Indicators not equal to 00, **, NA, or 99.

Repeated Entitlement Information

These variables (columns 590-633) are repeated once for every year from 1991 to 2015.

590	Entitlement Indicators (mon1-mon300)	12*1	EDB	One indicator for each month in the year. 0 = Not entitled 1 = Part A only 2 = Part B only 3 = Part A and B Note: mon1 = 1/1991, mon2 = 2/1991, mon3 = 3/1991mon300 = 12/2015
602	HMO Indicators (gho1-gho300)	12*1	EDB	One indicator for each month in the year. 0 = Not a member of HMO 1 = Non-Lock-in, CMS to process provider claims 2 = Non-Lock-in, GHO to process in-plan Part A & in-area Part B claims 4 = Chronic care disease management

Documentation for the Patient Entitlement and Diagnosis Summary File (PEDSF)
November 9, 2016

<u>COL</u>	<u>FIELD</u>	<u>LENGTH</u>	<u>SOURCE</u>	<u>NOTES</u>
				organizations-FFS plan 5 = ? A = Lock-in, CMS to process provider claims B = Lock-in, GH0 to process in-plan Part A & in-area Part B claims C = Lock-in, GH0 to process all Part A and Part B claims Note: gho1 = 1/1991, gho2 = 2/1991, gho3 = 3/1991 gho300 = 12/2015
614	State Buy-in Indicator (allflag1-allflag300)	12*1	EDB	One indicator for each month in the year. 0 = Not entitled 1 = Part A only 2 = Part B only 3 = Part A and B Note: allflag1 = 1/1991, allflag2 = 2/1991, allflag3 = 3/1991allflag300 = 12/2015
626	Part A Months (ptacnt1991-ptacnt2015)	2	EDB	Number of months covered for part A in each year covering 1991-2015. 00 = No coverage that year 01-12 = Number of months of coverage
628	Part B Months (ptbcnt1991-ptbcnt2015)	2	EDB	Number of months covered for part B in each year covering 1991-2015. 00 = No coverage that year 01-12 = Number of months of coverage
630	HMO Months (hmocnt1991-hmocnt2015)	2	EDB	Number of months as an HMO Member in each year covering 1991-2015. 00 = No coverage that year 01-12 = Number of months of coverage
632	State Buy-in Months (stbuyin1991-stbuyin2015)	2	EDB	Number of months with state buy-in coverage In each year covering 1991-2015. 00 = No coverage that year 01-12 = Number of months of coverage

Repeated Zip Code and Registry Information

Zip Code and Registry (based on state and county) information has been retained from the EDB file for each year from 1991 to 2015.

1734	Zip Codes (5) (zip5_1991-zip5_2015)	5	EDB	Encrypted Zip code (5 digits) is assigned as the last zip code the patient resided in that year. *Special permission required for unencrypted zip codes.
------	---	---	-----	---

Documentation for the Patient Entitlement and Diagnosis Summary File (PEDSF)
November 9, 2016

<u>COL</u>	<u>FIELD</u>	<u>LENGTH</u>	<u>SOURCE</u>	<u>NOTES</u>
1739	Zip Codes (4) (zip4_1991-zip4_2015)	4	EDB	Last 4 digits of zip code. Blanked out when the zip code is encrypted. *Special permission required for unencrypted zip codes.
1743	Registry Code (registry1991-registry2015)	2	EDB	01 = San Francisco 02 = Connecticut 20 = Detroit 21 = Hawaii 22 = Iowa 23 = New Mexico 25 = Seattle 26 = Utah 27 = Atlanta 31 = San Jose 35 = Los Angeles 37 = Rural Georgia 41 = Greater California 42 = Kentucky 43 = Louisiana 44 = New Jersey 47 = Greater Georgia NA = Has not lived in a registry yet NOTE: First registry found is retained until patient moved into a new registry.
1745	Registry Code #2 (reg2cd1991-reg2cd2015)	2	EDB	01 = San Francisco 02 = Connecticut 20 = Detroit 21 = Hawaii 22 = Iowa 23 = New Mexico 25 = Seattle 26 = Utah 27 = Atlanta 31 = San Jose 35 = Los Angeles 37 = Rural Georgia 41 = Greater California 42 = Kentucky 43 = Louisiana 44 = New Jersey 47 = Greater Georgia NA = Other

Documentation for the Patient Entitlement and Diagnosis Summary File (PEDSF)
November 9, 2016

<u>COL</u>	<u>FIELD</u>	<u>LENGTH</u>	<u>SOURCE</u>	<u>NOTES</u>
				NOTE: Registry is based off the state and county, which is the last residence the patient lived at for that year.
1747	State (state1991-state2015)	2	EDB	State code from EDB file.
1749	County (cnty1991-cnty2015)	3	EDB	County code from EDB file.

Repeated SEER Cancer Information

A patient can have up to 10 diagnoses in SEER, so the below variables are repeated for each cancer diagnosis the patient has in SEER. If the patient has more than 10 only the first 10 are retained for this file.

In parenthesis you will find the variable names. In italics you will find the SEER variable name.

2202	SEER Registry Code at Diagnosis (reg1-reg10) (<i>reg</i>)	2	SEER	01 = San Francisco (1973+) 02 = Connecticut (1973+) 20 = Detroit (1973+) 21 = Hawaii (1973+) 22 = Iowa (1973+) 23 = New Mexico (1973+) 25 = Seattle (1974+) 26 = Utah (1973+) 27 = Atlanta (1975+) 31 = San Jose (1988+) 35 = Los Angeles (1988+) 37 = Rural Georgia (1992+) 41 = Greater California (2000+) 42 = Kentucky (2000+) 43 = Louisiana (2000+) 44 = New Jersey (2000+) 47 = Greater Georgia (2000+)
2204	Marital Status at Diagnosis (marst1-marst10) (<i>mar_stat</i>)	1	SEER	See Attachment A
2205	Age at Diagnosis (agedx1-agedx10) (<i>age_dx</i>)	3	SEER	See Attachment A
2208	Sequence Number (seq1-seq10) (<i>seq_num</i>)	2	SEER	See Attachment A

Documentation for the Patient Entitlement and Diagnosis Summary File (PEDSF)
November 9, 2016

<u>COL</u>	<u>FIELD</u>	<u>LENGTH</u>	<u>SOURCE</u>	<u>NOTES</u>
2210	Month of Diagnosis (modx1-modx10) (date_mo)	2	SEER	See Attachment A
2212	Year of Diagnosis (yrdx1-yrdx10) (date_yr)	4	SEER	See Attachment A
2216	Primary Site (site1-site10) (site02v)	3	SEER	See Attachment A Note: This variable is length 4 in the SEER file. The first character (which is always 'C') was not written to the PEDSF.
2219	Laterality (lat1-lat10) (lateral)	1	SEER	See Attachment A
2220	Histology ICD-O-2 (1992-2000) (hist2_1-hist2_10) (hist02v)	4	SEER	See Attachment A
2224	Behavior code ICD-O-2 (1992-2000) (beh2_1-beh2_10) (beh02v)	1	SEER	See Attachment A
2225	Histology ICD-O-3 (hist1-hist10) (hist03v)	4	SEER	See Attachment A
2229	Behavior code ICD-O-3 (beh1-beh10) (beh03v)	1	SEER	See Attachment A
2230	Grade (grade1-grade10) (grade)	1	SEER	See Attachment A
2231	Diagnostic Confirmation (dxconf1-dxconf10) (dx_conf)	1	SEER	See Attachment A
2232	Type of Reporting Source (src1-src10) (rept_src)	1	SEER	See Attachment A

Documentation for the Patient Entitlement and Diagnosis Summary File (PEDSF)
November 9, 2016

<u>COL</u>	<u>FIELD</u>	<u>LENGTH</u>	<u>SOURCE</u>	<u>NOTES</u>
2233	EOD 10 - Tumor Size (1988-2003) (e10sz1-e10sz10) (eod10_sz)	3	SEER	See Attachment A
2236	EOD 10 – Tumor Extent (1988-2003) (e10ex1-e10ex10) (eod10_ex)	2	SEER	See Attachment A
2238	EOD 10 – Prostate pathology ext (1995-2003) (e10pe1-e10pe10) (eod10_pe)	2	SEER	See Attachment A
2240	EOD 10 - Nodes (1988-2003) (e10nd1-e10nd10) (eod10_nd)	1	SEER	See Attachment A
2241	EOD 10 - # positive Nodes (1988+) (e10pn1-e10pn10) (eod10_pn)	2	SEER	See Attachment A
2243	EOD 10 - # Nodes Examined (1988+) (e10ne1-e10ne10) (eod10_ne)	2	SEER	See Attachment A
2245	EOD – OLD 13 Digit (1973-1982) (eod13_1-eod13_10) (eod13)	13	SEER	See Attachment A
2258	EOD – OLD 2 Digit (1973-1982) (eod2_1-eod2_10) (eod2)	2	SEER	See Attachment A
2260	EOD 4 –size (1983-1987) (e4siz1-e4siz10) (eod4)	2	SEER	See Attachment A

Documentation for the Patient Entitlement and Diagnosis Summary File (PEDSF)
November 9, 2016

<u>COL</u>	<u>FIELD</u>	<u>LENGTH</u>	<u>SOURCE</u>	<u>NOTES</u>
2262	EOD 4 – extent (1983-1987) (e4ext1-e4ext10) (eod4)	1	SEER	See Attachment A
2263	EOD 4 – nodes (1983-1987) (e4nod1-e4nod10) (eod4)	1	SEER	See Attachment A
2264	Coding system – extent of disease (1973-2003) (eod_cd1-eod_cd10) (eodcode)	1	SEER	See Attachment A
2265	Tumor Marker 1 (1990-2003) (tumor1_1-tumor1_10) (tumor_1v)	1	SEER	See Attachment A
2266	Tumor Marker 2 (1990-2003) (tumor2_1-tumor2_10) (tumor_2v)	1	SEER	(Prostate data has been removed from this field) See Attachment A
2267	Tumor Marker 3 (1998-2003) (tumor3_1-tumor3_10) (tumor_3v)	1	SEER	See Attachment A
2268	CS Tumor Size (2004+) (cstum1-cstum10) (cs_size)	3	SEER	See Attachment A
2271	CS Extension (2004+) (csex1-csex10) (cs_ext)	3	SEER	See Attachment A
2274	CS Lymph Nodes (2004+) (cslym1-cslym10) (cs_node)	3	SEER	See Attachment A
2277	CS Mets at Dx (2004+) (csmet1-csmet10) (cs_mets)	2	SEER	See Attachment A

Documentation for the Patient Entitlement and Diagnosis Summary File (PEDSF)
November 9, 2016

<u>COL</u>	<u>FIELD</u>	<u>LENGTH</u>	<u>SOURCE</u>	<u>NOTES</u>
2279	CS Site-Specific Factor 1 (2004+) (cs1st1-cs1st10) (cs_ssf1)	3	SEER	(Prostate data has been removed from this field) See Attachment A
2282	CS Site-Specific Factor 2 (2004+) (cs2st1-cs2st10) (cs_ssf2)	3	SEER	(Prostate data has been removed from this field) See Attachment A
2285	CS Site-Specific Factor 3 (2004+) (cs3st1-cs3st10) (cs_ssf3)	3	SEER	See Attachment A
2288	CS Site-Specific Factor 4 (2004+) (cs4st1-cs4st10) (cs_ssf4)	3	SEER	See Attachment A
2291	CS Site-Specific Factor 5 (2004+) (cs5st1-cs5st10) (cs_ssf5)	3	SEER	See Attachment A
2294	CS Site-Specific Factor 6 (2004+) (cs6st1-cs6st10) (cs_ssf6)	3	SEER	See Attachment A
2297	CS Site-Specific Factor 25 (2004+) (cs25st1-cs25st10) (cs_ssf25)	3	SEER	See Attachment A
2300	Derived AJCC T, 6 th edition (2004+) (dajcct1-dajcct10) (d_ajcc_t)	2	SEER	See Attachment A
2302	Derived AJCC N, 6 th edition (2004+) (dajccn1-dajccn10) (d_ajcc_n)	2	SEER	See Attachment A
2304	Derived AJCC M, 6 th edition (2004+) (dajccm1-dajccm10) (d_ajcc_m)	2	SEER	See Attachment A

Documentation for the Patient Entitlement and Diagnosis Summary File (PEDSF)
November 9, 2016

<u>COL</u>	<u>FIELD</u>	<u>LENGTH</u>	<u>SOURCE</u>	<u>NOTES</u>
2306	Derived AJCC Stage Group, 6 th edition (2004+) (dajccstg1-dajccstg10) (d_ajcc_s)	2	SEER	See Attachment A
2308	Derived SS1977 (2004+) (dss77s1-dss77s10) (d_ssg77)	1	SEER	See Attachment A
2309	Derived SS2000 (2004+) (dss00s1-dss00s10) (d_ssg00)	1	SEER	See Attachment A
2310	Derived AJCC-Flag (dajcflg1-dajcflg10) (d_ajcc_f)	1	SEER	See Attachment A
2311	Derived SS1977 - Flag (dss77f1-dss77f10) (d_ssg77f)	1	SEER	1 = SS1977 derived from Collaborative Staging Manual and Coding Instructions, Version 1.0 2 = SS1977 derived from EOD (prior to 2004)
2312	Derived SS2000 – Flag (dss00f1-dss00f10) (d_ssg00f)	1	SEER	1 = SS2000 derived from Collaborative Staging Manual and Coding Instructions, Version 1.0 2 = SS2000 derived from EOD (prior to 2004)
2313	CS Version 1 st (2004+) (csvf1-csvf10) (csv_org)	6	SEER	See Attachment A
2319	CS Version Latest (2004+) (csvl1-csvl10) (csv_der)	6	SEER	See Attachment A
2325	CS version input current (2004+) (cscurrent1-cscurrent10) (csv_cur)	6	SEER	See Attachment A
2331	RX Summ – Surgery of Primary Site (1998+) (sxprif1-sxprif10) (surgprim)	2	SEER	See Attachment A

Documentation for the Patient Entitlement and Diagnosis Summary File (PEDSF)
November 9, 2016

<u>COL</u>	<u>FIELD</u>	<u>LENGTH</u>	<u>SOURCE</u>	<u>NOTES</u>
2333	RX –Summ Scope Reg LN Sur (2003+) (sxscof1-sxscof10) (scope)	1	SEER	See Attachment A
2334	RX Summ--Surg Oth Reg/Dis (2003+) (sxsitf1-sxsitf10) (surgoth)	1	SEER	See Attachment A
2335	Number of regional lymph nd exam (1998-2002) (numnd1-numnd10) (surgnode)	2	SEER	See Attachment A
2337	First course of reconstruction (1998-2002) (recstr1-recstr10) (reconst)	1	SEER	0 = No reconstruction/restoration 1 = Reconstruction, NOS (unknown if flap) 2 = Implant; reconstruction WITHOUT flap 3 = Reconstruction WITH flap, NOS 4 = Latissimus dorsi flap 5 = Abdominus recti flap 6 = Flap, NOS plus implant 7 = Latissimus dorsi flap plus implant 8 = Abdominus recti plus implant 9 = Unknown; not stated; death certificate only
2338	Reason no cancer- directed surgery (nosrg1-nosrg10) (no_surg)	1	SEER	See Attachment A
2339	Radiation (rad1-rad10) (radiatn)	1	SEER	See Attachment A
2340	Radiation to Brain and /or CNS (radbrn1-radbrn10) (rad_brn)	1	SEER	See Attachment A
2341	Radiation sequence with surgery (radsurg1-radsurg10) (rad_surg)	1	SEER	See Attachment A
2342	Site Specific Surgery (1983 – 1997) (sssurg1-sssurg10) (ss_surg)	2	SEER	See Attachment A

Documentation for the Patient Entitlement and Diagnosis Summary File (PEDSF)
November 9, 2016

<u>COL</u>	<u>FIELD</u>	<u>LENGTH</u>	<u>SOURCE</u>	<u>NOTES</u>
2344	Scope of regional lymph node surgery (1998-2002) (sxscop1-sxscop10) (scope02)	1	SEER	See Attachment A
2345	Surgery of other reg/dist sites (1998-2002) (sxsite1-sxsite10) (srgoth02)	1	SEER	See Attachment A
2346	Record Number (recnum1-recnum10)	2	SEER	See Attachment A
2348	Age-site edit override (ositage1-ositage10) (o_sitage)	1	SEER	1 = Reviewed and confirmed that age/site/histology combination is correct as coded 2 = Reviewed and confirmed that case was diagnosed in utero 3 = Reviewed and confirmed that conditions 1 and 2 apply
2349	Sequence number-dx conf override (oseqcon1-oseqcon10) (o_seqcon)	1	SEER	1 = Reviewed: Multiple primaries of special sites in which at least one diagnosis has not been microscopically confirmed have been reviewed.
2350	Site-type-lat-seq override (oseqlat1-oseqlat10) (o_seqlat)	1	SEER	1 = Reviewed: Multiple primaries of the same histology (3 digit) in the same primary site group have been reviewed.
2351	Surgery-diagnostic conf override (osurcon1-osurcon10) (o_surcon)	1	SEER	1 = Reviewed: A patient who had (cancer directed) surgery, but the tissue removed was not sufficient for microscopic confirmation.
2352	Site-type edit override (osittyp1-osittyp10) (o_sittyp)	1	SEER	1 = Reviewed: The coding of an unusual combination of primary site and histologic type has been reviewed.
2353	Histology edit override (hbenign1-hbenign10) (h_benign)	1	SEER	1 = Unusual Histology behavior combination reviewed. 2 = Unusual Diagnostic Confirmation/Behavior combination reviewed. 3 = Both scenarios above reviewed

Documentation for the Patient Entitlement and Diagnosis Summary File (PEDSF)
November 9, 2016

<u>COL</u>	<u>FIELD</u>	<u>LENGTH</u>	<u>SOURCE</u>	<u>NOTES</u>
2354	Report source sequence override (orptsrc1-orptsrc10) (o_rptsrc)	1	SEER	1 = Reviewed: A second or subsequent primary with a reporting source of death certificate only has been reviewed and is indeed an independent primary.
2355	Seq-ill-defined site override (odfsite1-odfsite10) (o_dfsite)	1	SEER	1 = Reviewed: A second or subsequent primary reported with an ill-defined primary site (C76.0 C76.8, C80.9) has been reviewed and is an independent primary.
2356	Leuk-Lymph dx confirmation override (oleukdx1-oleukdx10) (o_leukdx)	1	SEER	1 = Reviewed: A patient was diagnosed with leukemia or lymphoma and the diagnosis was not microscopically confirmed.
2357	Site-behavior override (ositbeh1-ositbeh10) (o_sitbeh)	1	SEER	1 = Reviewed: A patient has an in situ cancer of a nonspecific site and no further information about the primary site is available. Note: The IF39 edit does not allow in situ cases of nonspecific sites, such as gastrointestinal tract, NOS; uterus, NOS; female genital tract, NOS; male genital organs, NOS; and others. This over-ride indicates that the conflict has been reviewed.
2358	Site-EOD-dx date override (oeoddt1-oeoddt10) (o_eoddt)	1	SEER	1 = Reviewed: A patient had "localized" disease with a non-specific site and no further information about the primary site is available. Note: The IF40 edit does not allow "localized" disease with non-specific sites, such as mouth, NOS; colon, NOS (except histology 8220); bone, NOS; female genital system, NOS; male genital organs, NOS; and others. This over-ride indicates that the conflict has been reviewed.
2359	Site-laterality-EOD override (ositeod1-ositeod10) (o_siteod)	1	SEER	1 = Reviewed: A patient had laterality coded non-specifically and EOD coded specifically. Note: The IF41 edit for paired organs does not allow EOD to be specified as in situ, localized, or regional by direct extension if laterality is coded as "bilateral, side unknown" or "laterality unknown." This over-ride indicates that the conflict has been reviewed.

Documentation for the Patient Entitlement and Diagnosis Summary File (PEDSF)
November 9, 2016

<u>COL</u>	<u>FIELD</u>	<u>LENGTH</u>	<u>SOURCE</u>	<u>NOTES</u>
2360	Site-laterality-morph override (ositmor1-ositmor10) (o_sitmor)	1	SEER	1 = Reviewed: A patient had behavior code of "in situ" and laterality is not stated as right: origin of primary; left: origin of primary; or only one side involved, right or left origin not specified. Note: The IF42 edit does not allow behavior code of "in situ" with non-specific laterality codes. This over-ride indicates that the conflict has been reviewed.
2361	Type of follow-up expected (typefu1-typefu10) (typefup)	1	SEER	See Attachment A
2362	Age at Diagnosis recode (ager1-ager10) (age_rec)	2	SEER	See Attachment A
2364	Site Recode ICD-O-3/ WHO 2008 (siterwho1- siterwho10) (siterwho)	5	SEER	See Attachment A *This variable is used to select patients for data requests.
2369	Filler	4		
2373	Recode ICD-O-2 to 9 (icdot09_1- icdot09_10) (icdot09v)	4	SEER	See Attachment A
2377	Recode ICD-O-2 to 10 (icdot10_1- icdot10_10) (icdot10v)	4	SEER	See Attachment A
2381	ICCC site recode ICD-O-3/WHO 2008 (ICCC3WHO1- ICCC3WHO10) (ICCC3WHO)	3	SEER	See Attachment A
2384	ICCC site recode extended ICD-O-3/ WHO 2008 (ICC3XWHO1- ICC3XWHO10) (ICCC3XWHO)	3	SEER	See Attachment A

Documentation for the Patient Entitlement and Diagnosis Summary File (PEDSF)
November 9, 2016

<u>COL</u>	<u>FIELD</u>	<u>LENGTH</u>	<u>SOURCE</u>	<u>NOTES</u>
2387	Behavior recode for analysis (behtrend1-behtrend10) (<i>behanal</i>)	1	SEER	See Attachment A
2388	Histology recode - broad groupings (histrec1-histrec10) (<i>histrec</i>)	2	SEER	See Attachment A
2390	Histology recode – Brain groupings (hisrcb1-hisrcb10) (<i>brainrec</i>)	2	SEER	See Attachment A
2392	CS Schema – v 0204 (cs04sch1-cs04sch10) (<i>cs0204schema</i>)	3	SEER	See Attachment A
2395	SEER Historic Stage A (hstst1-hstst10) (<i>hst_stga</i>)	1	SEER	See Attachment A
2396	AJCC Stage 3 rd Edition (1988-2003) (ajccstg1-ajccstg10) (<i>ajcc_stg</i>)	2	SEER	See Attachment A
2398	SEER Modified AJCC Stage 3 rd edition (1988-2003) (aj3sr1-aj3sr10) (<i>aj_3seer</i>)	2	SEER	See Attachment A
2400	SEER Summary Stage 1977 (1995-2000) (sss77v1-sss77v10) (<i>ssg77</i>)	1	SEER	See Attachment A
2401	SEER summary stage 2000 (2001-2003) (sssm2Z1-sssm2Z10) (<i>ssg2000</i>)	1	SEER	See Attachment A
2402	First Malignant Primary Indicator (frstprm1-frstprm10) (<i>firstprm</i>)	1	SEER	See Attachment A
2403	FIPS State Code (statecd1-statecd10) (<i>stcounty</i>)	2	SEER	See Attachment A

Documentation for the Patient Entitlement and Diagnosis Summary File (PEDSF)
November 9, 2016

<u>COL</u>	<u>FIELD</u>	<u>LENGTH</u>	<u>SOURCE</u>	<u>NOTES</u>
2405	FIPS County Code (cnty1-cnty10) (stcounty)	3	SEER	See Attachment A
2408	IHS Code (ihscd1-ihscd10)	1	SEER	See Attachment A
2409	Summary Stage 2000 (1998+) (summ2k1-summ2k10) (hist_ssg_2000)	1	SEER	See Attachment A
2410	AYA site recode/WHO 2008 (ayawho1-ayawho10) (aya_recode)	2	SEER	See Attachment A
2412	Lymphoma subtype recode/WHO 2008 (lymwho1-lymwho10) (lymphoma_recode)	2	SEER	See Attachment A
2414	SEER cause-specific death classification (vsrtdx1-vsrt dx10) (dth_class)	1	SEER	See Attachment A
2415	SEER other cause of death classification (odthclass1-odthclass10) (o_dth_class)	1	SEER	See Attachment A
2416	CS Tumor Size/Ext Eval (2004+) (csts1-csts10) (exteval)	1	SEER	See Attachment A
2417	CS Reg Nodes Eval (2004+) (csrg1-csrg10) (nodeeval)	1	SEER	See Attachment A
2418	CS Mets Eval (2004+) (csmt1-csmt10) (metseval)	1	SEER	See Attachment A
2419	Primary by International Rules (intprim1-intprim10) (intprim)	1	SEER	See Attachment A

Documentation for the Patient Entitlement and Diagnosis Summary File (PEDSF)
November 9, 2016

<u>COL</u>	<u>FIELD</u>	<u>LENGTH</u>	<u>SOURCE</u>	<u>NOTES</u>
2420	ER Status Recode Breast Cancer (1990+) (erstat1-erstat10) (erstatus)	1	SEER	See Attachment A
2421	PR Status Recode Breast Cancer (1990+) (prstat1-prstat10) (prstatus)	1	SEER	See Attachment A
2422	CS Schema (cssch1-cssch10) (csschema)	2	SEER	See Attachment A
2424	CS Site-Specific Factor 8 (2004+) (cs8st1-cs8st10) (cs_ssf8)	3	SEER	See Attachment A
2427	CS Site-Specific Factor 10 (2004+) (cs10st1-cs10st10) (cs_ssf10)	3	SEER	See Attachment A
2430	CS Site-Specific Factor 11 (2004+) (cs11st1-cs11st10) (cs_ssf11)	3	SEER	See Attachment A
2433	CS Site-Specific Factor 13 (2004+) (cs13st1-cs13st10) (cs_ssf13)	3	SEER	See Attachment A
2436	CS Site-Specific Factor 15 (2004+) (cs15st1-cs15st10) (cs_ssf15)	3	SEER	See Attachment A
2439	CS Site-Specific Factor 16 (2004+) (cs16st1-cs16st10) (cs_ssf16)	3	SEER	See Attachment A
2442	Lymph-vascular Invasion (2004+) (vasinv1-vasinv10) (vasinv)	1	SEER	See Attachment A
2443	Survival Months (srvm1-srvm10) (srv_time_mon)	4	SEER	See Attachment A

Documentation for the Patient Entitlement and Diagnosis Summary File (PEDSF)
November 9, 2016

<u>COL</u>	<u>FIELD</u>	<u>LENGTH</u>	<u>SOURCE</u>	<u>NOTES</u>
2447	Survival Months Flag (srvmflag1- srvmflag10) (<i>srv_time_mon_flag</i>)	1	SEER	See Attachment A
2448	Filler	5		
2453	Insurance Recode (2007+) (insrecpb1- insrecpb10) (<i>insrec_pub</i>)	1	SEER	See Attachment A
2454	Derived AJCC T 7 th ed (dajcc7t1-dajcc7t10) (<i>dajcc7t</i>)	3	SEER	See Attachment A
2457	Derived AJCC N 7 th ed (dajcc7n1-dajcc7n10) (<i>dajcc7n</i>)	3	SEER	See Attachment A
2460	Derived AJCC M 7 th ed (dajcc7m1-dajcc7m10) (<i>dajcc7m</i>)	3	SEER	See Attachment A
2463	Derived AJCC 7 Stage Group (dajcc7_01-dajcc7_10) (<i>dajcc7stg</i>)	3	SEER	See Attachment A
2466	Breast Adjusted AJCC 6 th T (1988+) (adjajc6t1-adjajc6t10) (<i>adjtm_6value</i>)	2	SEER	See Attachment A
2468	Breast Adjusted AJCC 6 th N (1988+) (adjajc6n1-adjajc6n10) (<i>adjnm_6value</i>)	2	SEER	See Attachment A
2470	Breast Adjusted AJCC 6 th M (1988+) (adjajc6m1- adjajc6m10) (<i>adjm_6value</i>)	2	SEER	See Attachment A
2472	Breast Adjusted AJCC 6 th Stage (1988+) (adjajc6_01- adjajc6_10) (<i>adjajccstg</i>)	2	SEER	See Attachment A

Documentation for the Patient Entitlement and Diagnosis Summary File (PEDSF)
November 9, 2016

<u>COL</u>	<u>FIELD</u>	<u>LENGTH</u>	<u>SOURCE</u>	<u>NOTES</u>
2474	CS Site-Specific Factor 7 (cs7st1-cs7st10) (<i>cs7site</i>)	3	SEER	See Attachment A
2477	CS Site-Specific Factor 9 (cs9st1-cs9st10) (<i>cs9site</i>)	3	SEER	See Attachment A
2480	CS Site-Specific Factor 12 (cs12st1-cs12st10) (<i>cs12site</i>)	3	SEER	See Attachment A
2483	Derived HER2 Recode (2010+) (her2rec1-her2rec10) (<i>her2</i>)	1	SEER	See Attachment A
2484	Breast Subtype (2010+) (brstsub1-brstsub10) (<i>brst_sub</i>)	1	SEER	See Attachment A
2485	Lymphoma – Ann Arbor Stage (1983+) (annarbor1-annarbor10) (<i>annarbor</i>)	1	SEER	See Attachment A
2486	CS Mets at DX-Bone (2010+) (csmetsdxb_pub1-csmetsdxb_pub10) (<i>csmetsdxb_pub</i>)	1	SEER	See Attachment A
2487	CS Mets at DX-Brain (2010+) (csmetsdxbr_pub1-csmetsdxbr_pub10) (<i>csmetsdxbr_pub</i>)	1	SEER	See Attachment A
2488	CS Mets at DX-Liver (2010+) (csmetsdxliv_pub1-csmetsdxliv_pub10) (<i>csmetsdxliv_pub</i>)	1	SEER	See Attachment A
2489	CS Mets at DX-Lung (2010+) (csmetsdxlung_pub1-csmetsdxlung_pub10) (<i>csmetsdxlung_pub</i>)	1	SEER	See Attachment A

Documentation for the Patient Entitlement and Diagnosis Summary File (PEDSF)
November 9, 2016

<u>COL</u>	<u>FIELD</u>	<u>LENGTH</u>	<u>SOURCE</u>	<u>NOTES</u>
2490	T value-based on AJCC 3 rd (1988-2003) (t_value1-t_value10) (t_value)	2	SEER	See Attachment A
2492	N value-based on AJCC 3 rd (1988-2003) (n_value1-n_value10) (n_value)	2	SEER	See Attachment A
2494	M value-based on AJCC 3 rd (1988-2003) (m_value1-m_value10) (m_value)	2	SEER	See Attachment A
2496	Diagnosis Date Flag (ddxflag1-ddxflag10)	2	SEER	12 = Date is unknown Missing = A valid date is provided, or the date was not expected to have been transmitted
2498	Therapy date flag (dthflag1-dthflag10)	2	SEER	10 = Unknown if therapy was administered 11 = Therapy was not administered 12 = Therapy was administered, but the date is unknown Missing = A valid date is provided, or was not expected to have been transmitted
2500	Month Therapy Started (monrx1-monrx10) (monthrx)	2	SEER	Month of Therapy 00 = No Therapy 01-12 = Valid month blank = Unknown
2502	Year Therapy Started (yearrx1-yearrx10) (year_rx)	4	SEER	Year of Therapy. If year is blank, refer to Therapy Date Flag (dthflag) for more information.
2506	Other Therapy (other_tx1-other_tx10) (othr_rx)	1	SEER	0 = No other cancer-directed therapy 1 = Other cancer-directed therapy 2 = Other experimental cancer-directed therapy 3 = Double blind study, code not yet broken 6 = Unproven therapy (including Laetrile, Krebiozen, etc) 7 = Refused therapy 1-3 above 8 = Recommended, unknown if administered 9 = Unknown
2507	ICD-O coding scheme (icdo1-icdo10) (icdoover)	1	SEER	2 = Originally coded in ICD-O-2, (1973-2000) 3 = Originally coded in ICD-O-3, (2001+)

Documentation for the Patient Entitlement and Diagnosis Summary File (PEDSF)
November 9, 2016

<u>COL</u>	<u>FIELD</u>	<u>LENGTH</u>	<u>SOURCE</u>	<u>NOTES</u>
2508	NAPIIA Derived API Race (napiia1-napiia10) (napiia)	2	SEER	<p>The NAACCR Asian Pacific Islander Identification Algorithm version 1 (NAPIIA v1) uses a combination of NAACCR variables to classify cases directly or indirectly as Asian Pacific Islander for analytic purposes. This version of the algorithm is focused on coding cases with a race code of Asian NOS (race code 96) to a more specific Asian race category, using the birthplace and name fields (first, last, and maiden names). Birthplace can be used to indirectly assign a specific race to one of eight Asian race groups (Chinese, Japanese, Vietnamese, Korean, Asian Indian, Filipino, Thai, and Cambodian). Names can be used to indirectly assign a specific race to one of seven Asian groups (Chinese, Japanese, Vietnamese, Korean, Asian Indian, Filipino, and Hmong). Future versions of NAPIIA will incorporate Pacific Islanders and will potentially incorporate name lists for Thai, Cambodian, and Laotians.</p> <p>Codes are the same as for Race/Ethnicity. Blanks in field mean the algorithm was not run.</p>
2510	Primary Payer at DX (payer_dx1-payer_dx10) (payerdx)	2	SEER	<p>Primary Payer at Diagnosis identifies the patients primary insurance carrier or method of payment at the time of initial diagnosis and/or treatment.</p> <p>01 = Not insured 02 = Not insured, self pay 10 = Insurance, NOS 20 = Private Insurance: Managed care, HMO, or PPO 21 = Private Insurance: Fee-for-Service 31 = Medicaid 35 = Medicaid - Administered through a Managed Care plan 60 = Medicare/Medicare, NOS 61 = Medicare with supplement, NOS 62 = Medicare - Administered through a Managed Care plan 63 = Medicare with private supplement 64 = Medicare with Medicaid eligibility</p>

Documentation for the Patient Entitlement and Diagnosis Summary File (PEDSF)
November 9, 2016

<u>COL</u>	<u>FIELD</u>	<u>LENGTH</u>	<u>SOURCE</u>	<u>NOTES</u>
				65 = TRICARE 66 = Military 67 = Veterans Affairs 68 = Indian/Public Health Service 99 = Insurance status unknown
2512	OncoType DX recurrence score (onco_score1-onco_score10) (<i>oncotype_score</i>)	3	SEER	000-100 *Special permission required to receive this variable
2515	OncoType DX risk group (onco_rg1-onco_rg10) (<i>oncotype_rg</i>)	1	SEER	1 = Low 2 = Intermediate 3 = High *Special permission required to receive this variable
2516	OncoType DX reason no score (onco_rns1-onco_rns10) (<i>oncotype_rns</i>)	1	SEER	1 = Cancelled Test – Cancelled by physician or patients before result was returned 2 = Failed Test – Inadequate tissue sample 3 = Outside Analysis Criteria 4 = Non-Unique – Patient had multiple tumors *Special permission required to receive this variable
2517	OncoType DX delivered year (onco_year1-onco_year10) (<i>oncotype_year</i>)	4	SEER	2004-2015 *Special permission required to receive this variable
2521	OncoType DX delivered month (onco_month1-onco_month10) (<i>oncotype_month</i>)	2	SEER	01-12 *Special permission required to receive this variable
2523	OncoType DX months since diagnosis (onco_time1-onco_time10) (<i>oncotype_time</i>)	3	SEER	000-996 = Valid months 997 = Unknown 998 = Test before DX Unknown occurs when month of diagnosis is unknown and year of diagnosis is not greater than test year. *Special permission required to receive this variable
5442	Filler	1		

Documentation for the Patient Entitlement and Diagnosis Summary File (PEDSF)
November 9, 2016

Documentation for the Patient Entitlement and Diagnosis Summary File (PEDSF)
November 9, 2016

Codes for the variables:

-Cause of Death to site recode KM (codkm)

-Cause of Death to site recode (codpub)

00000 = Alive
20010 = Lip
20020 = Tongue
20030 = Salivary Gland
20040 = Floor of Mouth
20050 = Gum and Other Mouth
20060 = Nasopharynx
20070 = Tonsil
20080 = Oropharynx
20090 = Hypopharynx
20100 = Other Oral Cavity and Pharynx
21010 = Esophagus
21020 = Stomach
21030 = Small Intestine
21040 = Colon excluding Rectum
21050 = Rectum and Rectosigmoid Junction
21060 = Anus, Anal Canal and Anorectum
21071 = Liver
21072 = Intrahepatic Bile Duct
21080 = Gallbladder
21090 = Other Biliary
21100 = Pancreas
21110 = Retroperitoneum
21120 = Peritoneum, Omentum and Mesentery
21130 = Other Digestive Organs
22010 = Nose, Nasal Cavity and Middle Ear
22020 = Larynx
22030 = Lung and Bronchus
22050 = Pleura
22060 = Trachea, Mediastinum and Other Respiratory Organs
23000 = Bones and Joints
24000 = Soft Tissue including Heart
25010 = Melanoma of the Skin
25020 = Other Non-Epithelial Skin
26000 = Breast
27010 = Cervix Uteri
27020 = Corpus Uteri
27030 = Uterus, NOS
27040 = Ovary
27050 = Vagina
27060 = Vulva
27070 = Other Female Genital Organs

Documentation for the Patient Entitlement and Diagnosis Summary File (PEDSF)
November 9, 2016

28010 = Prostate
28020 = Testis
28030 = Penis
28040 = Other Male Genital Organs
29010 = Urinary Bladder
29020 = Kidney and Renal Pelvis
29030 = Ureter
29040 = Other Urinary Organs
30000 = Eye and Orbit
31010 = Brain and Other Nervous System
32010 = Thyroid
32020 = Other Endocrine including Thymus
33010 = Hodgkin Lymphoma
33040 = Non-Hodgkin Lymphoma
34000 = Myeloma
35011 = Acute Lymphocytic Leukemia
35012 = Chronic Lymphocytic Leukemia
35013 = Other Lymphocytic Leukemia
35021 = Acute Myeloid Leukemia
35031 = Acute Monocytic Leukemia
35022 = Chronic Myeloid Leukemia
35023 = Other Myeloid/Monocytic Leukemia
35041 = Other Acute Leukemia
35043 = Aleukemic, Subleukemic and NOS
36010 = Mesothelioma (ICD-10 only) (not in CODPUB)
36020 = Kaposi Sarcoma (ICD-10 only) (not in CODPUB)
37000 = Miscellaneous Malignant Cancer
38000 = In situ, benign or unknown behavior neoplasm
50000 = Tuberculosis
50010 = Syphilis
50030 = Septicemia
50040 = Other Infectious and Parasitic Diseases including HIV
50050 = Diabetes Mellitus
50051 = Alzheimers (ICD-9 and 10 only)
50060 = Diseases of Heart
50070 = Hypertension without Heart Disease
50080 = Cerebrovascular Diseases
50090 = Atherosclerosis
50100 = Aortic Aneurysm and Dissection
50110 = Other Diseases of Arteries, Arterioles, Capillaries
50120 = Pneumonia and Influenza
50130 = Chronic Obstructive Pulmonary Disease and Allied Cond
50140 = Stomach and Duodenal Ulcers
50150 = Chronic Liver Disease and Cirrhosis
50160 = Nephritis, Nephrotic Syndrome and Nephrosis
50170 = Complications of Pregnancy, Childbirth, Puerperium

Documentation for the Patient Entitlement and Diagnosis Summary File (PEDSF)
November 9, 2016

50180 = Congenital Anomalies
50190 = Certain Conditions Originating in Perinatal Period
50200 = Symptoms, Signs and Ill-Defined Conditions
50210 = Accidents and Adverse Effects
50220 = Suicide and Self-Inflicted Injury
50230 = Homicide and Legal Intervention
50300 = Other Cause of Death
41000 = State DC not available or state DC available but no COD
99999 = Unknown/missing/invalid COD

Note:

Rural/Urban Continuum as Defined in the 2004 ARF file
(urban/rural code)

The **2003 Rural/Urban Continuum Codes** are from Economic Research Service (ERS), Department of Agriculture. The codes form a classification scheme that distinguishes metropolitan (metro) counties by the population size of their metro area and nonmetropolitan (nonmetro) counties by degree of urbanization and adjacency to a metro area or nonmetro areas. All U.S. counties and county equivalents are grouped according to the official metro status announced by the Office of Management and Budget (OMB) in June 2003, when the population and worker commuting criteria used to identify metro counties were applied to results of the 2000 Census.

Metro counties are distinguished by population size of the Metropolitan Statistical Area of which they are part. Nonmetro counties are classified according to the aggregate size of their urban population. Within the three urban size categories, nonmetro counties are further identified by whether or not they have some functional adjacency to a metro area or areas. A nonmetro county is defined as adjacent if it physically adjoins one or more metro areas, and has at least 2 percent of its employed labor force commuting to central metro counties. Nonmetro counties that do not meet these criteria are classed as nonadjacent.

In concept, the 2003 version of the Rural-Urban Continuum Codes is comparable with that of earlier decades. However, OMB made major changes in its metro area delineation procedures for the 2000 Census, and the Census Bureau changed the way in which rural and urban are measured. Therefore, the new Rural-Urban Continuum Codes are not fully comparable with those of earlier years. OMB's changes added some additional metro areas by no longer requiring that a metro area must have at least 100,000 population if its urbanized area has no place of at least 50,000 people. More importantly, simplifying the worker commuting criteria that determine outlying metro counties had the effect of both adding numerous new outlying counties to metro status while deleting a smaller number that were previously metro.

The Census Bureau made a radical shift in determining rural-urban boundaries by changing and liberalizing the procedures for delineating urbanized areas of 50,000 or more people, and abandoning place boundaries in measuring urban or rural population. The procedures used in defining Urbanized Areas were extended down to clusters of 2,500 or more people, based solely on population density per square mile.

In earlier versions of the Rural-Urban Continuum Codes, metro areas with 1 million population or more were subdivided between central counties (Code 0) and fringe counties (Code 1). The Code 1 group has become much less meaningful in the last two censuses as more and more counties of large metro areas have been rated as central counties by OMB procedures. In 2000, only 1.6 percent of the population of large metro areas was in fringe counties. Therefore, this distinction has been dropped. Codes 0 and 1 have been combined, and the new code 1 represents all counties in metro areas of 1 million or more population.

Documentation for the Patient Entitlement and Diagnosis Summary File (PEDSF)
April 24, 2015

The 2003 Rural/Urban Continuum Codes are defined as follows:

CODE METROPOLITAN COUNTIES (1-3)

- 01 Counties of metro areas of 1 million population or more
- 02 Counties in metro areas of 250,000 - 1,000,000 population
- 03 Counties in metro areas of fewer than 250,000 population

CODE NON METROPOLITAN COUNTIES (4-9)

- 04 Urban population of 20,000 or more, adjacent to a metro area
- 05 Urban population of 20,000 or more, not adjacent to a metro area
- 06 Urban population of 2,500-19,999, adjacent to a metro area
- 07 Urban population of 2,500-19,999, not adjacent to a metro area
- 08 Completely rural or less than 2,500 urban population, adjacent to a metro area
- 09 Completely rural or less than 2,500 urban population, not adjacent to a metro area
- 99 Missing Value