

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

FILENAME: pedsf.5percent.txt

LRECL: 2,184

SEER Cases diagnosed 1973-2013

PEDSF File created on October 4, 2016

SEER data extracted from November, 2015 SEER Submission

ABSTRACT: The 5% PEDSF file contains one record per person for individuals in the SEER Program data base who have been matched with Medicare enrollment records. These persons are limited to those in the 5% sample (defined by CMS). 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

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

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 | <p>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.</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> |
|------------|--|---------------|---------------|--|
| 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. |

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)
November 9, 2016

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