

List of variables requested from the California Cancer Registry for “Exploring disparities in cancer incidence, treatment and survival using de-identified California Cancer Registry data”

ADMINISTRATIVE

PATIENT_ID: Unique statewide identification number assigned by the CCR to identify each patient.

TUMOR_ID: Unique statewide identification number assigned by the CCR to identify each tumor.

RESREG: Number of region assigned by the CCR to the residence where the patient was diagnosed.

TYPERP: Best source of information to prepare this case abstract.

YEARDX: Year of Diagnosis

CLASS: Distinguishes cases that are usually included in a hospital’s treatment and survival statistics (00-22) from those that are not (30-49 and 99). Can also be used to exclude path only and death certificate only cases.

PATIENT DEMOGRAPHICS

AGE: Age, in complete years, when the patient was first diagnosed with this tumor.

AGECAT: Age group (five-year) where AGE falls for statistical purposes.

BIRTHPL: Place of birth for the patient- state or country.

MARSTAT: Marital status when the patient was first diagnosed with this tumor.

RACE1: Race/ethnicity of the patient

RACE2: Race/ethnicity of the patient. This field was added in 2000 to allow for the identification of multiple races.

RACE3: Race/ethnicity of the patient. This field was added in 2000 to allow for the identification of multiple races.

RACE4: Race/ethnicity of the patient. This field was added in 2000 to allow for the identification of multiple races.

RACE5: Race/ethnicity of the patient. This field was added in 2000 to allow for the identification of multiple races.

RACE08: Race/ethnicity of the patient including Hispanic ethnicity. Hispanics may be of any race.

RACE99: Race/ethnicity of the patient including Hispanic ethnicity. Hispanics may be of any race.

RELIGION: Patient’s religion at time of diagnosis.

SEX : Sex of the patient.

SPANISH: Used to denote persons of Spanish origin. Persons of Spanish origin may be of any race.

YOSTSCL: Yost’s index of socioeconomic status (SES) level based on a principal components analysis.

TUMOR DIAGNOSTICS

BEHAVIOR: SEER behavior recode for analysis: designating the malignancy or behavior of the tumor.

CS_EXT: Part of the Collaborative Staging (CS) data item set that identifies the primary tumor growth within the organ of origin or its extension into neighboring organs. Values are site-specific.

CS_LN: Part of the Collaborative Staging (CS) data item set that identifies the regional lymph nodes involved. This data item is collapsible into AJCC N-code. Values are site-specific.

CSS_TUM_SIZE: Part of the Collaborative Staging (CS) data item set that, for most sites, records the largest dimension or diameter of the primary tumor in millimeters. For many sites, the CS algorithm uses this data item to derive the DERIVED_AJCC_T. Values are site-specific.

DIFFERN2: Sixth digit of ICD-O-3; designates the grade or differentiation of the tumor.

DIREXTTU: Two-digit extension code that represents the growth of the primary tumor within the organ of origin, its extension into neighboring organs, or its metastasis to distant structures. It is a hierarchical code in which the most extensive disease is coded. For cases diagnosed 2004 and forward, these values are recorded in CS_TUM_SIZE.

DXCONF: Indicates whether at any time during the patient's medical history there was microscopic confirmation of this cancer.

HISTO_T2: The first four digits of the ICD-O-2 morphology code, indicates the histology/cell type of this tumor.

HISTO_T3: The first four digits of the ICD-O-3 morphology code, indicates the histology/cell type of this tumor. Coded directly for cases diagnosed 2001 and forward. Cases coded prior to 2001 were converted to ICDO-3.

LATERAL: For some specific primary sites, the side of the body in which the tumor originated.

MARKER1: Field to record estrogen receptor for breast cancer 1990-2003, acid phosphatase for prostate cancer 1998-2003, alphafetoprotein for testicular cancer 1998-2003, carcinoembryonic antigen (CEA) for colorectal cancer 2000-2003, and carbohydrate antigen (CA-125) for ovarian cancer 2000-2003. It was not required to record CEA or CA-125 for this data item. For cases diagnosed 2004 and forward, these values were recorded in CS_SITE_SPEC_F1.

CS_SITE_SPEC_F1: Part of the Collaborative Staging (CS) data item set. This is 1 of 6 variables used to code additional site-specific information needed to derive TNM or AJCC stage, or to code prognostic factors that have an effect on stage or survival. Values are site-specific.

MARKER2: Field to record progesterone receptor for breast cancer 1990-2003, prostate specific antigen (PSA) for prostate cancer 1998-2003, and HCG for testicular cancer 1998-2003. For cases diagnosed 2004 and forward, these values were recorded in CS_SITE_SPEC_F2. Cases diagnosed 2004 forward should be blank. If a hospital is collecting tumor markers for other cancer sites, they may use this field:

CS_SITE_SPEC_F2: Part of the Collaborative Staging (CS) data item set. This is 1 of 6 variables used to code additional site-specific information needed to derive TNM or AJCC stage, or to code prognostic factors that have an effect on stage or survival. Values are site-specific.

CS_SITE_SPEC_F3: Part of the Collaborative Staging (CS) data item set. This is 1 of 6 variables used to code additional site-specific information needed to derive TNM or AJCC stage, or to code prognostic factors that have an effect on stage or survival. Values are site-specific.

MARKERCA: Breast cancer tumor marker for California- 1: Her2/neu (also known as c-erbB2 or ERBB2) since 1999. For more information on values see Volume I V.6.4

SEER: SEER site recode- site group used for statistical reporting according to SEER's standards. Separates leukemias and lymphomas from site-specific cancers.

SEQNOCEN: Indicates the sequence of all reportable neoplasms during the patient's lifetime as determined by the central registry. When two or more tumors are diagnosed simultaneously, the tumor with the worse prognosis is assigned the lowest sequence number:

SITE_02: Location where this tumor originated in as much detail as is known and for which a code is provided in ICD-O-2 for cases 1988-2000 and ICDO-3 for cases 2001 forward.

TSIZETU: Extent of disease variable- size of tumor in millimeters. For cases diagnosed 2004 and forward, values are recorded in CS_TUM_SIZE.

STAGE AT DIAGNOSIS

LNSUMTU: Extent of disease variable that specifies furthest lymph node involvement. For cases diagnosed after 2004, values are recorded in CS_LN.

NNODES: Number of regional lymph nodes identified in the pathology report during surgical procedure- this variable is only valid for cases diagnosed prior to Jan. 2004. Information in NNODES from 1988-2003 was incorporated into the variable SCOPE.

PNODETU: Number of regional lymph nodes with evidence of involvement (positive).

STAGE_AJCC: Conversion of EOD codes to 3rd edition AJCC staging manual.

STAGE_AJCCSEER: Conversion of EOD codes to 3rd edition AJCC staging manual, modified by SEER.

STAGE_SEER: Combined AJCC stage variable, modified by SEER. Extent of Disease for 1994-2003 diagnoses or collaborative stage for 2004+ diagnoses converted to AJCC stage as modified by SEER.

SUMSTAGE: Summary stage at time of diagnosis.

XNODETU: Number of regional lymph nodes examined by the pathologist.

SURGICAL TREATMENT

SURGPRIM: Most extensive type of surgery performed during the first course of treatment for the tumor.

SURGOTH: Surgical removal of tissue other than the primary tumor or organ of origin (i.e., regional nodes, distant sites, or distant nodes).

NOSURG: Reason why the first course of treatment did not include definitive surgery. Reason for No Surgery only applies to surgery of the primary site.

NON-SURGICAL TREATMENT

CHEMOSUM: Identifies the type of chemotherapy given as first course of treatment
HORMSUM: Records whether systemic hormonal agents were given as first course of treatment at any facility, or the reason why they were not given.
IMMUSUM: Records whether systemic immunotherapy was given as first course of treatment at any facility, or the reason why it was not given.
RADSUM: Summary of radiation therapy given as first course of treatment.
RADSEQ: Indicates the sequence of radiation therapy with surgery (pre-op, post-op, etc.) during the first course of treatment.
NORAD: Reason why the first course of treatment did not include radiation.
TRANSSUM: Identifies systemic therapeutic procedures given as first course of treatment, or the reason why they procedures were not used.
VITSTAT: Patient's vital status as of the date of data extraction.

FOLLOW-UP

YEARFU: Date (year only) of last known information about this patient, or date of death if patient deceased.
CDEATH: Underlying cause of death, based on causes of death information from the death certificate.
SURV_TIME: Survival time, in months.

GEOGRAPHIC INFORMATION

ZIPCODE3: First three digits of residence zip code
RUCA: The Rural-Urban Commuting Area (RUCA) codes published by the U.S. Department of Agriculture for each census tract.
RURALMSSA90: Designation of rural/urban for a Medical Service Study Area (MSSA) by the California Health Manpower Policy Commission, based on the 1990 U.S. Census.
RURALMSSA00: Designation of rural/urban for a Medical Service Study Area (MSSA) by the California Health Manpower Policy Commission, based on the 2000 U.S. Census.
QUINYOST: Socioeconomic Status (SES) quintile of YOSTSCL score- based on Census variables, at the block group level.
YOSTSCL: Index of socioeconomic status (ses) level based on a principal components analysis (continuous variable).

Note: Further information about CCR variables, including codes for each variable and completeness tables, can be obtained in the California Cancer Registry Data Dictionary available on the CCR website.

(http://www.ccrca.org/Data_and_Statistics/Cancer_Data_for_Research.shtml)