AACR
American Association
for Cancer Research
FINDING CURES TOGETHER℠

PROJECTGENIE
Genomics Evidence Neoplasia Information Exchange

GENIE BPC Analytic Data Guide

BrCa v1.0-public

Prepared by the Statistical Coordinating Center at Memorial Sloan Kettering Cancer Center

Contents

OVERVIEW

The GENIE Biopharma Collaborative (BPC) Core Team has created analytic datasets that include data from the AACR Project GENIE Tier 1 registry, institutions' tumor registries and electronic health records using the PRISSMM phenomic data model. Derived variables based on these data elements are also provided. This analytic data guide serves as a resource for all data elements included in the analytic datasets containing demographic and phenomic information, which can be combined with the genomic data that have been submitted to GENIE.

Release Notes: This document accompanies the **BrCa v1.0-public** data release.

The genomic data associated with this release are GENIE **v9.1-consortium**.

Abbreviations

AACR American Association for Cancer Research
BrCa Breast Cancer
BPC Biopharma Collaborative
CPT Cancer Panel Test
DFCI Dana-Farber Cancer Institute, Boston, MA, United States
EHR Electronic Health Record
GENIE Genomics Evidence Neoplasia Information Exchange
HIPAA The Health Insurance Portability and Accountability Act of 1996
MSK Memorial Sloan Kettering Cancer Center, New York, NY, United States
NAACCR North American Association of Central Cancer Registries
NDI National Death Index
NGS Next Generation Sequencing
PRISSMM Pathology, **R**adiology, **I**maging, **S**igns, **S**ymptoms, Tumor **M**arkers, **M**edical Oncologist
QA Quality Assurance
UHN University Health Network Princess Margaret, Toronto, Ontario, Canada
VICC Vanderbilt-Ingram Cancer Center, Nashville, TN, United States

PRISSMM Overview and Note

PRISSMM is a system for extracting clinical data from longitudinal EHRs and for ascertaining high salience outcomes of cancer treatment in multi-site projects that span different health systems. PRISSMM includes a system for estimating "real world endpoints" from the text data contained in the EHR. PRISSMM is a flexible convenient consensus-based standard in the same way that RECIST criteria have become a standard for radiology endpoints in clinical trials. PRISSMM is the phenomic data that characterizes treatment exposures, and outcomes based on the results of pathology (P), imaging of local (R) and distant sites (I), signs (S), symptoms (S), tumor markers (M) and Medical Oncology assessments (M). The system specifies data provenance, timing for review, approaches to handling treatment gaps and discontinuity and endpoint specification that is applicable to longitudinal EHR data for patients with cancer.

PRISSMM training materials and this data guide are shared with the GENIE-BPC release. A license to the complete set of PRISSMM tools including databases used to extract data and training materials can be obtained by e-mailing PRISSMM@mskcc.org. The PRISSMM system was developed at DFCI and licensing is shared between MSK and DFCI. The materials are shared at no cost with academic medical centers and for a fee that is used to support development activities with for profit entities.

BPC Projects

GENIE BPC sponsors six projects in phase I that involve augmenting the genomic data in Project GENIE with PRISSMM phenomic data and cancer-related outcomes. Each project encompasses a single cancer site. The six projects are: bladder cancer, breast cancer, colon/rectal cancer, non-small cell lung cancer, pancreas cancer, and prostate cancer. A description of the GENIE BPC Project and its associated data quality assurance processes have been published (Lavery et al, 2022).

Eligibility

The BPC Project-specific eligibility criteria for the BrCa project are as follows:

- Eligible OncoTree Diagnoses:
  - Breast Angiosarcoma (BA)
  - Invasive Breast Carcinoma (BRCA)
  - Breast Invasive Cancer, NOS (BRCANOS)
  - Breast Invasive Carcinoma, NOS (BRCNOS)
  - Breast Invasive Ductal Carcinoma (IDC)
  - Breast Invasive Lobular Carcinoma (ILC)
  - Metaplastic Breast Cancer (MBC)
  - Breast Mixed Ductal and Lobular Carcinoma (MDLC)
  - Mixed Type Metaplastic Breast Cancer (MMBC)
  - Malignant Phyllodes Tumor of the Breast (MPT)
  - Metaplastic Squamous Cell Carcinoma (MSCC)
- Stage I-IV at diagnosis
- Genomic sequencing report at DFCI, MSK, or VICC between January 1, 2013 and December 31, 2018
- Aged 18-56 at the time of genomic sequencing
- Minimum of two years of possible follow-up after sequencing

Cases meeting specified eligibility criteria are randomly selected from the AACR Project GENIE Cancer Registry.

Note that a patient can be selected for more than one project (e.g., if a patient is selected for the non-small cell lung cancer project, it is possible for that patient to also be selected for the breast cancer project if they met eligibility criteria for both projects).

These selection criteria may impact the generalizability of results since genomic sequencing is not always performed at diagnosis and therefore may lead to several forms of bias. Investigators who wish to test specific hypotheses should work with a statistician to perform analyses that account for these biases. Failure to do so may result in incorrect inferences (Brown et al, 2022).

Data Privacy

Compliance with data privacy required redaction of the name and duration of investigational drugs as well as date intervals that could lead to identification of a patient as >89 years of age at any time point.

{genieBPC} R Package

The {genieBPC} R package can be used to import the data files directly into R and create analytic cohorts.

Lavery JA, Brown S, Curry MA, Martin A, Sjoberg DD, Whiting K. A data processing pipeline for the AACR project GENIE biopharma collaborative data with the {genieBPC} R package. Bioinformatics. 2023 Jan 1;39(1):btac796. doi: 10.1093/bioinformatics/btac796. PMID: 36519837; PMCID: PMC9822536.

References

Acebedo A, Bedard PL, Brown S, Ceca E, Fiandalo M, Fuchs H, Guo X, Hoppe JN, Kehl KL, Kundra R, Lavery JA, LeNoue-Newton ML, Lepisto E, Mastrogiacomo B, Micheel CM, Nayan C, Newcomb A, Nichols C, Panageas KS, Piening B, Pillai S, Postle A, Potter R, Riely GJ, Rizvi H, Rudolph J, Schrag D, Sweeney SM, Sweet-Cordero EA, Turski M, Wingord E, Wu TJ, Yu T, Yu C. Collaborating across sectors in service of open science, precision oncology, and patients: an overview of the AACR Project GENIE (Genomics Evidence Neoplasia Information Exchange) Biopharma Collaborative (BPC). ESMO Real World Data and Digital Oncology 7:100097, 2025.

Kehl KL, Lavery JA, Brown S, Fuchs H, Riely G, Schrag D, Newcomb A, Nichols C, Micheel CM, Bedard PL, Sweeney SM, Fiandalo M, Panageas KS, AACR Project GENIE BPC Core Team. Biomarker Inference and the Timing of Next-Generation Sequencing in a Multi-Institutional, Cross-Cancer Clinicogenomic Data Set. JCO Precis Oncol. 2024 Mar;8:e2300489. doi: 10.1200/PO.23.00489. PMID: 38484212; PMCID: PMC10954072.

Kehl KL, Uno H, Gusev A, Groha S, Brown S, Lavery JA, Deborah Schrag D, Katherine S. Panageas KS. Elucidating analytic bias due to informative cohort entry in cancer clinico-genomic datasets. Cancer Epidemiol Biomarkers Prev. 2023 Mar 6;32(3):344-352. doi: 10.1158/1055-9965.EPI-22-0875. PMID: 36626408; PMCID: PMC9992002.

Lavery JA, Lepisto EM, Brown S, Rizvi H, McCarthy C, LeNoue-Newton M, Yu C, Lee J, Guo X, Yu T, Rudolph J, Sweeney S, AACR Project GENIE Consortium, Park BH, Warner JL, Bedard PL, Riely G, Schrag D, Panageas KS. A Scalable Quality Assurance Process for Curating Oncology Electronic Health Records: The Project GENIE Biopharma Collaborative Approach. JCO Clin Cancer Inform. 2022 Feb;6:e2100105. doi: 10.1200/CCI.21.00105. PMID: 35192403; PMCID: PMC8863125.

Brown S, Lavery JA, Shen R, Martin AS, Kehl KL, Sweeney SM, Lepisto EM, Rizvi H, McCarthy CG, Schultz N, Warner JL, Park BH, Bedard PL, Riely GJ, Schrag D, Panageas KS; AACR Project GENIE Consortium. Implications of Selection Bias Due to Delayed Study Entry in Clinical Genomic Studies. JAMA Oncol. 2022 Feb 1;8(2):287-291. doi: 10.1001/jamaoncol.2021.5153. PMID: 34734967; PMCID: PMC9190030.

Kehl KL, Schrag D, Hassett MJ, Uno H. Assessment of Temporal Selection Bias in Genomic Testing in a Cohort of Patients With Cancer. JAMA Netw Open. 2020 Jun 1;3(6):e206976. doi: 10.1001/jamanetworkopen.2020.6976. PMID: 32511717; PMCID: PMC7280950.

Format

There are 9 analytic datasets included in the BrCa v1.0-public data release. This data guide describes the variables included each dataset.

Dataset	File Name on Synapse
1. Patient Characteristics Dataset	patient_level_dataset.csv
2. BPC Project Cancer Diagnosis Dataset	cancer_level_dataset_index.csv
3. Non-BPC Project Cancer Diagnosis Dataset	cancer_level_dataset_non_index.csv
4. Cancer-Directed Regimen Dataset	regimen_cancer_level_dataset.csv
5. PRISSMM Pathology Dataset	pathology_report_level_dataset.csv
6. PRISSMM Imaging Dataset	imaging_level_dataset.csv
7. PRISSMM Medical Oncologist Assessment Dataset	med_onc_note_level_dataset.csv
8. PRISSMM Tumor Marker Dataset	tm_level_dataset.csv
9. Cancer Panel Test Dataset	cancer_panel_test_level_dataset.csv

We describe each variable included in these datasets using the following format:

Field name
[Variable name]
Value (character/numeric/date/date-time)
Description of the field
Data Standard (where applicable)

Variables have been color coded to help users understand their provenance.
Variables shown in orange indicate variables obtained from the AACR Project GENIE Tier 1 data.
Variables shown in green indicate variables obtained directly from the institution's tumor registry.
Variables shown in blue indicate variables obtained from curation of the EHR.
Variables shown in purple were derived by the Statistical Coordinating Center.

Field names shaded in gray indicate that another variable is preferred. For example, stage at diagnosis is captured from the tumor registry and also curated from the EHR, but a derived variable combining the two sources is also provided. In this case, the tumor registry and curated stage at diagnosis are shown with gray shading and the composite derived variable is preferred for analysis.

The following table provides further information about the types of variables.

Type of Variable	Description
AACR Project GENIE Tier 1 Data	Variables denoted in orange represent data that is obtained directly from the AACR Project GENIE Tier 1 Data. These variables correspond to the values stored for each sample in the Tier 1 Data. Users should refer to the AACR Project GENIE website for further detail.
Tumor registry at each institution	Variables denoted in green are obtained from each hospital's tumor registry system. These are NAACR-defined variables and thus are highly standardized across institutions. Not all patients have their data captured by the tumor registry. The provenance of the data is abstraction by Certified Tumor Registrars at each institution. The NAACCR data standards are noted for each registry-specific field and users should refer to the NAACCR website for further detail.
Curated	Variables denoted in blue have been curated according to the PRISSMM phenomic data curation model to extract unstructured data from the EHR.
Derived	Variables denoted in purple are calculated based on AACR Project GENIE Tier 1, tumor registry and/or curated variables. These variables are provided to facilitate ease of use as well as consistency across different analyses.

PATIENT CHARACTERISTICS DATASET

The Patient Characteristics dataset is structured as one record per patient. It can be linked to all other datasets using the variables [cohort] and [record_id].

BPC Project Cohort

[cohort]

Value (Character)

- BrCa

Description

- Indicates the BPC Project Cancer type
- Primary key for merging across all datasets

Record ID

[record_id]

Value (Character)

- GENIE-[INSTITUTION]-XXXXXX

Description

- De-identified, unique patient ID
- Conforms to the following the convention: GENIE-[INSTITUTION]-XXXX. The first component is the string, “GENIE”; the second component is the institution’s abbreviation; the third component is a unique ID for the patient.
- Primary key for the AACR Project GENIE genomic datasets [PATIENT_ID].

Institution

[institution]

Value (Character)

- DFCI = Dana Farber Cancer Institute
- MSK = Memorial Sloan Kettering Cancer Center
- VICC = Vanderbilt Ingram Cancer Center

Description

- Indicates the patient’s internal institution of cancer care
- Corresponds to variable [center] in AACR Project GENIE data

Patient has any Redacted Time Interval Data to Comply with Health Insurance Portability and Accountability Act (HIPAA)

[redacted]

Value (Character)

- Yes
- No

Description

- Indicates whether any portion of the patient’s data (across all datasets) was redacted to comply with HIPAA.
- Currently, any patients with [redacted] = "Yes" are not included in the data release.

Year of Birth

[birth_year]

Value (Numeric)

- YYYY

Description

- Patient's year of birth

Ethnicity: Spanish/Hispanic Origin

[naaccr_ethnicity_code]

Value (Character)

- Non-Spanish/non-Hispanic
- Spanish/Hispanic
- Unknown
- Not Collected

Description

- Ethnicity of patient, independent of patient’s race
- Institutions not collecting Spanish/Hispanic origin have set this field to “Unknown whether Spanish or not”

Data Standard: NAACCR #190

Race (Primary)

[naaccr_race_code_primary]

Value (Character)

- Asian
- Black
- Native American
- White
- Pacific Islander
- Unknown
- Other
- Not Applicable
- Not collected

Description

- First race specified, independent of ethnicity
- For institutions collecting more than one race category, this race variable indicates the primary race for the patient.
- Institutions not collecting race set this field to “Unknown”

Data Standard: NAACCR #160

Race (Secondary)

[naaccr_race_code_secondary]

Value (Character)

- Asian
- Black
- Native American
- White
- Pacific Islander
- Unknown
- Other
- Not Applicable
- Not collected

Description

- Second race specified, independent of ethnicity
- Institutions not collecting secondary race set this field to “No further race documented”

Data Standard: NAACCR #161

Race (Tertiary)

[naaccr_race_code_tertiary]

Value (Character)

- Asian
- Black
- Native American
- White
- Pacific Islander
- Unknown
- Other
- Not Applicable
- Not collected

Description

- Third race specified, independent of ethnicity
- Institutions not collecting tertiary race set this field to “No further race documented”

Data Standard: NAACCR #162

Sex

[naaccr_sex_code]

Value (Character)

- Male
- Female
- Other
- Transsexual
- Not Collected
- Unknown

Description

- Patient’s sex at time of diagnosis of index cancer

Data Standard: NAACCR #220

Time (Days) from Date of Birth to Date of Last Oncology Visit to Internal Institution

[last_oncvisit_int]

Value (Numeric)

Description

- Interval in days from date of birth to most recent date that the patient had an in-person or tele-visit visit with an oncology provider at the internal institution of care; these visits may include medical oncology, surgical oncology, radiation oncology, palliative care, social work, lab draws, imaging scans, emergency room, or hospital visits.
- Completeness of data is subject to availability of visit information at each participating institution. Consequently, the preferred variable for determining the date of last contact is [dob_lastalive_int].

Time (Days) from Date of Birth to Date Last Known Alive

[last_alive_int]

Value (Numeric)

Description

- Interval in days from date of birth to most recent date that the patient was known to be alive
- Requires documentation of a phone call or email exchange with the patient or a family member.
- Completeness of data is subject to availability of visit information at each participating institution. Consequently, the preferred variable for determining the date of last contact is [dob_lastalive_int].

Time (Days) from Date of Birth to Date of Last Known Non-Oncology Visit to Internal Institution

[last_anyvisit_int]

Value (Numeric)

Description

- Interval in days from date of birth to most recent date that the patient had an in-person or tele-visit visit to the internal institution for non-oncology related care, including visits with primary

care or cardiology.
- Completeness of data is subject to availability of visit information at each participating institution. Consequently, the preferred variable for determining the date of last contact is [dob_lastalive_int].

Time (Days) from Date of Birth to Enrollment in Hospice Care

[enroll_hospice_int]

Value (Numeric)

Description

- Interval in days from date of birth to the date that the patient was enrolled in hospice care.
- Completeness of data is subject to availability of visit information at each participating institution. Consequently, the preferred variable for determining the date of last contact is [dob_lastalive_int].

Time (Days, Months, Years) from Date of Birth to Last Known Alive Date

[dob_lastalive_int], [dob_lastalive_int_mos], [dob_lastalive_int_yrs]

Value (Numeric)

Description

- Interval in days [dob_lastalive_int], months [dob_lastalive_int_mos], and years [dob_lastalive_int_yrs] from date of birth to last known alive date.
- This variable is recommended as the censoring date in survival analyses.
- Based on the composite of the most recent date that the patient: 1. received oncology care at the internal institution, 2. received any care at the internal institution, 3. documentation from any source that the patient is alive, and 4. the patient is enrolled in hospice.

Age (Years) at Last Known Alive Date

[age_last_fu_yrs]

Value (Numeric)

Description

- Age of the patient at the time last known to be alive.
- Based on the composite of the most recent date that the patient: 1. received oncology care at the internal institution, 2. received any care at the internal institution, 3. documentation from any source that the patient is alive, and 4. the patient is enrolled in hospice.

Time (Days) from Date of Birth to Death

[hybrid_death_int]

Value (Numeric)

Description

- Interval in days from date of birth to date of death
- Populated only if patient is known to be dead at time of curation

Age (Years) at Death Date

[age_death_yrs]

Value (Numeric)

Description

- Age at death, in years
- Populated only if patient is known to be dead at the time of curation

Source of Death Information

[hybrid_death_source]

Value (Character)

- Curated
- EHR
- NDI
- Tumor Registry
- Other

Description

- Indicates source of death information

Number of Cancers, Any Type

[n_cancers]

Value (Numeric)

Description

- Number of invasive and non-invasive/in situ cancer diagnoses ever experienced by the patient
- Based on the count of records in Cancer Diagnosis dataset for each record ID
- Basal cell and squamous cell skin cancers are not included

Number of BPC Project Cancers (Index Cancers)

[n_cancers_index]

Value (Numeric)

Description

- Number of BPC Project Cancers that were identified for a patient
- A BPC Project Cancer is defined as the cancer that met eligibility criteria, underwent genomic sequencing and was submitted to AACR Project GENIE.
- Each patient has at least one BPC Project Cancer. Patients may have multiple BPC Project Cancers.
- PRISSMM data elements are curated for BPC Project Cancers.
- The terms “BPC Project Cancer” and “index cancer” are used interchangeably.
- Further details regarding the definition of BPC Project and non-BPC Project Cancers can be found in Appendix 1: BPC Project and Non-BPC Project Cancers

Number of Cancer-Directed Regimens Curated

[n_regimens_pt]

Value (Numeric)

Description

- The total number of cancer-directed regimens, including anti-neoplastic, immunotherapy, and hormone therapy that the patient has ever received for any cancer diagnosis
- Based on the count of records in the Cancer-Directed Regimen dataset for each record ID
- This number includes cancer-directed regimens given for non-BPC Project Cancers.

Number of Imaging Reports

[n_imaging_reports_pt]

Value (Numeric)

Description

- The total number of imaging scans curated starting in the month/year of the BPC Project Cancer diagnosis
- Based on the count of records in the Imaging dataset for each record ID
- Imaging scans include: CTs, MRIs, PET, PET/CTs, Bone Scans, and Nuclear Medicine scans
- This number includes scans performed for non-BPC Project Cancers
- For patients without a diagnosis of cancer in the breast, mammograms were not reviewed. For patients with a diagnosis of cancer in the breast, mammograms are reviewed starting at the month/year of the BPC Project Cancer diagnosis.

Number of CT Scans

[n_scans_ct_pt]

Value (Numeric)

Description

- The number of CT imaging scans curated starting in the month/year of the BPC Project Cancer diagnosis
- Based on the count of CT scans in the Imaging dataset for each record ID
- This number includes CT scans performed for non-BPC Project Cancers

Number of MRIs

[n_scans_mri_pt]

Value (Numeric)

Description

- The number of MRI imaging scans curated starting in the month/year of the BPC Project Cancer diagnosis
- Based on the count of MRIs in the Imaging dataset for each record ID
- This number includes MRIs performed for non-BPC Project Cancers

Number of PET or PET-CT Scans

[n_scans_pet_ct_pt]

Value (Numeric)

Description

- The number of PET or PET-CT imaging scans curated starting in the month/year of the BPC Project Cancer diagnosis

- Based on the count of PET and PET-CT scans in the Imaging dataset for each record ID
- This number includes PET or PET-CT scans performed for non-BPC Project Cancers

Number of Mammograms (Breast Cancer Only)

[n_scans_mammog_pt]

Value (Numeric)

Description

- The number of mammograms curated starting in the month/year of the BPC Project Cancer diagnosis for patients with any (BPC Project or non-BPC Project) diagnosis of breast cancer
- Based on the count of mammograms in the Imaging dataset for each record ID
- This number includes mammograms performed for non-BPC Project Cancers of the breast.
- For patients without a diagnosis of cancer in the breast, mammograms were not reviewed

Number of Bone Scans

[n_scans_bone_pt]

Value (Numeric)

Description

- The number of bone scans curated starting in the month/year of the BPC Project Cancer diagnosis
- Based on the count of bone scans in the Imaging dataset for each record ID
- This number includes bone scans performed for non-BPC Project Cancers

Number of Other Scans

[n_scans_other_pt]

Value (Numeric)

Description

- The number of other imaging scans, including other nuclear medicine scans, curated starting in the month/year of the BPC Project Cancer diagnosis
- Based on the count of other scans in the Imaging dataset for each record ID
- This number includes scans performed for non-BPC Project Cancers

Number of Pathology Reports

[n_path_reports_pt]

Value (Numeric)

Description

- The number of pathology reports starting in the month/year of the BPC Project Cancer diagnosis
- Based on the count of records in the Pathology dataset for each record ID
- This number includes pathology reports for non-BPC Project Cancers

Number of Medical Oncologist Assessments Curated

[n_md_notes_pt]

Value (Numeric)

Description

- The number of medical oncologist assessments curated starting in the month/year of the BPC Project Cancer diagnosis
- Based on the count of records in the Medical Oncologist Assessment dataset for each record ID
- One medical oncologist assessment per month was curated. Curation instructions are provided in Appendix 3.

Number of Tumor Marker Results

[n_tm_pt]

Value (Numeric)

Description

- The number of tumor marker results that were curated
- The tumor markers curated include: AFP, BhCG, CA125, CA15-3, CA19-9, CA2729, Calcitonin, CEA, Chromogranin A, LDH, PSA, NSE, Testosterone and Thyroglobulin
- Based on the count of records in the Tumor Marker dataset for each record ID

Number of Tumor Marker AFP Results

[n_tm_afp_pt]

Value (Numeric)

Description

- The number of AFP tumor marker results that were curated
- Based on the count of AFP results in the Tumor Marker dataset for each record ID

Number of Tumor Marker BhCG Results

[n_tm_bh_cg_pt]

Value (Numeric)

Description

- The number of BhCG tumor marker results that were curated
- Based on the count of BhCG results in the Tumor Marker dataset for each record ID

Number of Tumor Marker CA125 Results

[n_tm_ca125_pt]

Value (Numeric)

Description

- The number of CA125 tumor marker results that were curated
- Based on the count of CA125 results in the Tumor Marker dataset for each record ID

Number of Tumor Marker CA15-3 Results

[n_tm_ca15_3_pt]

Value (Numeric)

Description

- The number of CA15-3 tumor marker results that were curated
- Based on the count of CA15-3 results in the Tumor Marker dataset for each record ID

Number of Tumor Marker CA19-9 Results

[n_tm_ca19_9_pt]

Value (Numeric)

Description

- The number of CA19-9 tumor marker results that were curated
- Based on the count of CA19-9 results in the Tumor Marker dataset for each record ID

Number of Tumor Marker CA2729 Results

[n_tm_ca2729_pt]

Value (Numeric)

Description

- The number of CA2729 tumor marker results that were curated
- Based on the count of CA2729 results in the Tumor Marker dataset for each record ID

Number of Tumor Marker Calcitonin Results

[n_tm_calcitonin_pt]

Value (Numeric)

Description

- The number of Calcitonin tumor marker results that were curated
- Based on the count of Calcitonin results in the Tumor Marker dataset for each record ID

Number of Tumor Marker CEA Results

[n_tm_cea_pt]

Value (Numeric)

Description

- The number of CEA tumor marker results that were curated
- Based on the count of CEA results in the Tumor Marker dataset for each record ID

Number of Tumor Marker Chromogranin A Results

[n_tm_chromogranin_a_pt]

Value (Numeric)

Description

- The number of Chromogranin tumor marker results that were curated
- Based on the count of Chromogranin results in the Tumor Marker dataset for each record ID

Number of Tumor Marker LDH Results

[n_tm_ldh_pt]

Value (Numeric)

Description

- The number of LDH tumor marker results that were curated
- Based on the count of LDH results in the Tumor Marker dataset for each record ID

Number of Tumor Marker PSA Results

[n_tm_psa_pt]

Value (Numeric)

Description

- The number of PSA tumor marker results that were curated
- Based on the count of PSA results in the Tumor Marker dataset for each record ID

Number of Tumor Marker NSE Results

[n_tm_nse_pt]

Value (Numeric)

Description

- The number of NSE tumor marker results that were curated
- Based on the count of NSE results in the Tumor Marker dataset for each record ID

Number of Tumor Marker Testosterone Results

[n_tm_testosterone_pt]

Value (Numeric)

Description

- The number of Testosterone tumor marker results that were curated
- Based on the count of Testosterone results in the Tumor Marker dataset for each record ID

Number of Tumor Marker Thyroglobulin Results

[n_tm_thyroglobulin_pt]

Value (Numeric)

Description

- The number of Thyroglobulin tumor marker results that were curated
- Based on the count of Thyroglobulin results in the Tumor Marker dataset for each record ID

Number of Eligible Cancer Panel Tests Curated

[n_cpt_pt]

Value (Numeric)

Description

- The number of cancer panel tests that met the eligibility criteria and were curated
- Based on the count of records in the Cancer Panel Test dataset
- Not all sequenced specimens included in the AACR Project GENIE repository appear in this dataset due to eligibility requirements.

Release Version Number

[release_version]

Value (Character)

- 1.0-public

Description

- Indicates the version number of the data release

CANCER-DIAGNOSIS DATASETS

1. The BPC Project Cancer Diagnosis dataset contains one record per BPC Project Cancer diagnosis, per patient. A BPC Project Cancer is the cancer that met the eligibility criteria for the project and was selected for PRISSMM phenomic data curation.

This dataset can be linked to the following datasets:

- *Cancer-Directed Regimen, Cancer Panel Test, and Cancer-Directed Radiation Therapy datasets using the variables [cohort], [record_id], and [ca_seq].*
- *Patient Characteristics, PRISSMM Pathology, PRISSMM Imaging, and PRISSMM Medical Oncologist Assessment datasets using the variables [cohort] and [record_id].*

2. The Non-BPC Project Cancer Diagnosis dataset contains one record per non-BPC Project Cancer diagnosis, per patient. This dataset includes two types of cancer diagnoses: 1) non-BPC Project invasive cancer and in situ/non-invasive cancer diagnoses, and 2) other tumors.

This dataset can be linked to the following datasets:

- *Cancer-Directed Regimen and Cancer-Directed Radiation Therapy datasets using the variables [cohort], [record_id], and [ca_seq].*
- *Patient Characteristics, PRISSMM Pathology, PRISSMM Imaging, PRISSMM Medical Oncologist Assessment, and PRISSMM Tumor Marker datasets using the variables [cohort] and [record_id].*
- *Cannot be linked to the Cancer Panel Test dataset because non-BPC Project Cancer diagnoses were not genomically sequenced (Appendix 1).*

Further details regarding the definition of BPC Project and non-BPC Project Cancers can be found in Appendix 1.

Field names shaded in gray indicate that an alternative variable is preferred for analysis. The recommended variable is noted in the description.

Note that NAACCR variables are not available for patients from UHN.

BPC Project Cohort

[cohort]

Value (Character)

- BrCa

Description

- Indicates the BPC Project Cancer type
- Primary key for merging across all datasets

Record ID

[record_id]

Value (Character)

- GENIE-[INSTITUTION]-XXXXXX

Description

- De-identified, unique patient ID
- Conforms to the following the convention: GENIE-[INSTITUTION]-XXXX. The first component is the string, “GENIE”; the second component is the institution’s abbreviation; the third component is a unique ID for the patient.
- Primary key for the AACR Project GENIE genomic datasets [PATIENT_ID].

Institution

[institution]

Value (Character)

- DFCI = Dana Farber Cancer Institute
- MSK = Memorial Sloan Kettering Cancer Center
- VICC = Vanderbilt Ingram Cancer Center

Description

- Indicates the patient’s internal institution of cancer care
- Corresponds to variable [center] in AACR Project GENIE data

Diagnosis Eligible for Tumor Registry Curation

[tr_eligible]

Value (Numeric)

- 1 = Eligible for full curation
- 0 = Ineligible for full curation

Description

- Indicates whether the diagnosis is eligible for curation from tumor registry. A cancer diagnosis is eligible for curation if the behavior code [naaccr_behavior_cd] is 2 (non-invasive/in situ) or 3 (invasive/malignant). Cancer diagnoses with behavior codes 0 or 1 are not eligible for full annotation in the BPC project, but the tumor registry data are available.

Cancer Sequence Identifier

[ca_seq]

Value (Numeric)

- 0 = first and only cancer
- 1 = first of two or more primaries
- 2 = second of two or more primaries
- …10 = tenth of ten or more primaries

Description

- Sequential order of cancer diagnoses based on the date of diagnosis
- Primary key for the Cancer Diagnosis, Cancer-Directed Regimen and Cancer Panel Test datasets

BPC Project (Index) Cancer Indicator

[redcap_ca_index]

Value (Character)

- Yes
- No

Description

- The BPC Project Cancer is defined as the cancer that met eligibility criteria, underwent genomic sequencing and was submitted to AACR Project GENIE.
  - Each patient has at least one BPC Project Cancer. Patients may have multiple BPC Project Cancers, though this is rare.
  - PRISSMM data elements are curated for BPC Project Cancers.
  - The terms “BPC Project Cancer” and “index cancer” are used interchangeably.
  - Further details regarding the definition of BPC Project and non-BPC Project Cancers can be found in Appendix 1. BPC Project and Non-BPC Project Cancers.

Time (Days, Months, Years) from Date of Birth to Cancer Diagnosis

[dob_ca_dx_days], [dob_ca_dx_mos], [dob_ca_dx_yrs]

Value (Numeric)

Description

- Interval in days [dob_ca_dx_days], months [dob_ca_dx_mos], or years [dob_ca_dx_yrs] between date of birth and date of cancer diagnosis

Source of Cancer Diagnosis Date

[ca_dx_how]

Value (Character)

- Pathology
- Imaging
- Physical Exam
- Other

Description

- Source of the curated diagnosis date
- If [ca_dx_how] = "Pathology", the cancer diagnosis date is based upon a review of the pathology report indicating the first histologic confirmation of cancer.
- Populated only if diagnosis is eligible for curation ([tr_eligible] = 1)

Time (Days, Months, Years) from Date of Birth to Next Curated Cancer Diagnosis

[dob_next_ca_days], [dob_next_ca_mos], [dob_next_ca_yrs]

Value (Numeric)

Description

- Time in days [dob_next_ca_days], months [dob_next_ca_mos], or years [dob_next_ca_yrs] from date of birth to the next curated cancer diagnosis
- Populated only if patient has a subsequent cancer diagnosis

Time (Days, Months, Years) from First BPC Project Cancer Diagnosis

[first_index_ca_days], [first_index_ca_mos], [first_index_ca_yrs]

Value (Numeric)

Description

- For patients with multiple cancer diagnoses, the time in days [first_index_ca_days], months [first_index_ca_mos], or years [first_index_ca_yrs] from first BPC Project Cancer diagnosis to other cancer diagnosis
- Time is negative if the comparative cancer occurred prior to the BPC project cancer; time is positive if the comparative cancer occurred after the BPC Project Cancer

- Populated only if cancer diagnosis is not the first BPC Project Cancer. This field is blank for the first BPC Project Cancer

Tumor Registry Time (Days) from Date of Birth to First Contact at Institution

[naaccr_first_contact_int]

Value (Numeric)

Description

- Interval in days from date of birth to first contact at institution based on tumor registry
- Populated only if cancer diagnosis is obtained from the tumor registry. This field is blank for curated cancer diagnoses.

Data Standard: NAACCR #580

Age (Years) at Diagnosis

[age_dx]

Value (Numeric)

Description

- Patient age at diagnosis (years)

Primary Cancer Site

[ca_d_site]

Value (Character)

- ICD-O-3 topography code
- F10 = Cerebral Spinal Fluid
- F20 = Peritoneal Fluid/Ascites
- F30 = Pleural Fluid
- F40 = Urine
- F50 = Pericardial Fluid

Description

- ICD-O-3 code for the primary cancer site

Data Standard: ICD-O-3 Topography Codes

Cancer Type

[ca_type]

Value (Character)

- Adrenocortical Carcinoma
- Anal Cancer
- Appendix Cancer
- Bile Duct Cancer
- Bladder Cancer
- Brain Cancer
- Breast Cancer
- Breast Sarcoma
- NET or Carcinoid
- Cervical Cancer
- Corpus Uteri Carcinoma and Carcinosarcoma
- Corpus Uteri Sarcoma
- Colon Cancer
- Colon/Rectum Cancer
- Esophagus Cancer
- Ewing Sarcoma
- Fallopian Tube Cancer
- Gallbladder Cancer
- Germ Cell Tumor
- GIST
- Head and Neck Cancer
- Mesothelioma
- Ill Defined/Cancer of Unknown Primary
- Liver Cancer
- Lung Cancer, NOS
- Melanoma
- Merkel Cell
- Neuroblastoma
- Non Small Cell Lung Cancer
- Osteosarcoma
- Ovarian Cancer
- Pancreatic Cancer
- Parathyroid Cancer
- Penis Cancer
- Peritoneum Cancer
- Placenta Cancer
- Prostate Cancer
- Rectum and Rectosigmoid Cancer
- Renal Kidney Cancer
- Renal Pelvis Cancer

- Retinoblastoma
- Rhabdomyosarcoma
- Scrotum Cancer
- Small Cell Lung Cancer
- Small Intestine Cancer
- Soft tissue sarcoma
- Stomach Cancer
- Testis Cancer
- Thymus Cancer
- Thyroid Cancer
- Uterus Cancer
- Vagina Cancer
- Vulva Cancer
- Wilms Tumor
- Other

Description

- Cancer type was characterized based on information that includes ICD-O-3 topography and morphology codes.
- Mappable to AJCC Collaborative Stage v2.05, AJCC staging v7 and v8, SEER, and NCI.
- Populated only if diagnosis is not a hematopoietic or lymphoid neoplasm or pre-malignancy ([ca_heme_malig] = “No”)

Brain Cancer Type

[ca_type_brain]

Value (Character)

- Astrocytic Tumors
- Diffuse astrocytic and oligodendroglial tumors
- Choriod plexus tumors
- Craniopharyngioma (Grade I)
- Embryonal tumors
- Ependymal Tumors
- Germ cell tumors
- Medulloblastomas
- Meningeal Tumors
- Mesenchymal non meningothelial tumors
- Mixed Gliomas
- Neuronal and mixed neuronal-gilial tumors
- Oligodendroglial Tumors
- Other astrocytic tumors
- Other gliomas
- Pineal Parenchymal Tumors
- Tumors of cranial and paraspinal nerves

- Tumors of the pineal region
- Tumors of the sellar region
- Other

Description

- Populated only if:
    - Cancer type is recorded as brain cancer ([ca_type] = "Brain Cancer")
    - Non-BPC Project Cancer

Data Standard: World Health Organization Central Nervous System Tumor Classification System (2018)

Cancer Type, Other

[ca_type_oth]

Value (Character)

- Free-text

Description

- Populated only if:
    - Cancer type is recorded as "other" ([ca_type] = "Other")
    - Non-BPC Project Cancer

Hematopoietic or Lymphoid Neoplasm or Pre-Malignancy

[ca_heme_malig]

Value (Character)

- Yes
- No

Description

- Indicates whether cancer diagnosis is a hematologic malignancy
- Populated only if:
    - Cancer diagnosis is curated. This field is blank for cancer diagnoses obtained from the tumor registry.
    - Non-BPC Project Cancer

Type of Hematopoietic or Lymphoid Neoplasm

[ca_heme_type]

Value (Character)

- Leukemia NOS
- Non-Hodgkin's Lymphoma
- Langerhans Cell Histiocytosis
- MDS Myelodysplastic Syndrome
- MGUS Monoclonal gammopathy of undetermined significance
- Multiple Myeloma
- Plasmacytoma
- Other hematopoietic or lymphoid neoplasm
- Unspecified

Description

- Populated only if:
  - Hematopoietic or lymphoid neoplasm or pre-malignancy ([ca_heme_malig] = “Yes”).
  - Non-BPC Project Cancer

Data Standard: National Cancer Institute Cancer Types: Hematologic/Blood

Tumor Registry Histology

[naaccr_histology_cd]

Value (Numeric)

- ICD-O-3 morphology code

Description

- Histology code from tumor registry (ICD-O-3)
- Populated only if cancer diagnosis is obtained from the tumor registry. This field is blank for curated cancer diagnoses.

Data Standard: NAACCR #552

Histology

[ca_histology]

Value (Character)

- Mapping from ICD-O-3 histology and behavior codes

Description

- Histology description based on [naaccr_histology_cd] and the histology descriptions provided on this spreadsheet (https://seer.cancer.gov/icd-o-3/)

Histology Category

[ca_hist_adeno_squamous]

Value (Character)

- Adenocarcinoma
- Carcinoma
- Squamous cell
- Sarcoma
- Small cell carcinoma
- Other histologies/mixed tumor

Description

- Broad histology group based on [naaccr_histology_cd], if available. If unavailable, may be derived using information in the pathology report dataset.
- Populated only if BPC Project Cancer

Breast Cancer Histology

[ca_hist_brca]

Value (Character)

- Invasive lobular carcinoma
- Invasive ductal carcinoma
- Other histology

Description

- Breast cancer specific histology group based on first pathology report on or after diagnosis.
- Based on a classification hierarchy: 1. invasive lobular carcinoma; 2. invasive ductal carcinoma; 3. other histology.
- If any specimen on pathology report had documented lobular histology, regardless of any other histology documented, then “Invasive lobular carcinoma.” Otherwise, if any specimen had documented ductal histology, then “Invasive ductal carcinoma.” Otherwise, histology categorized as “Other histology.”
- Populated for the first BPC Project breast cancer only.

Tumor Registry ICD-O-3 Behavior Code

[naaccr_behavior_cd]

Value (Numeric)

- 0 = Benign
- 1 = Borderline
- 2 = In situ and/or carcinoma in situ
- 3 = Invasive

Description

- Behavior code from tumor registry (ICD-O-3)
- Populated only if cancer diagnosis is obtained from the tumor registry. This field is blank for curated cancer diagnoses.

Data Standard: NAACCR #523

Tumor Registry Laterality Code

[naaccr_laterality_cd]

Value (Character)

- 0 = Not a paired site
- 1 = Right: origin of primary
- 2 = Left: origin of primary
- 3 = Only one side involved, right or left origin unspecified
- 4 = Bilateral involvement at time of diagnosis, lateral origin unknown for a single primary; or both ovaries involved simultaneously, single histology; bilateral retinoblastomas; bilateral Wilms' tumors
- 5 = Paired site: midline tumor
- 9 = Paired site, but no information concerning laterality
- Not paired

Description

- Laterality code from the tumor registry
- Populated only if cancer diagnosis is obtained from the tumor registry. This field is blank for curated cancer diagnoses.

Data Standard: NAACCR #410

Grade or Differentiation of Tumor

[ca_grade]

Value (Character)

- I
- II
- III
- IV
- Low grade
- High grade
- B-cell
- T-cell

Description

- The grade or degree of differentiation of the tumor

Data Standard: NAACCR #440

Tumor Registry Best Group Stage

[best_ajcc_stage_cd]

Value (Character)

- Free-text

Description

- Best stage group calculated by each institution’s tumor registry software (METRIQ Tumor Registry Algorithm)
- Populated only if:
    - Cancer diagnosis is obtained from the tumor registry. This field is blank for curated cancer diagnoses.
    - BPC Project Cancer diagnosis
- Preferred derived variable [stage_dx] incorporates both tumor registry and curated stage.

Curated Stage IV at Diagnosis

[ca_stage_iv]

Value (Character)

- No
- Yes
- Not Applicable
- Unknown

Description

- Indicates whether cancer was diagnosed as stage IV
- Populated only if:
  - Cancer diagnosis is curated. This field is blank for cancer diagnoses obtained from the tumor registry, unless the tumor registry Best Group Stage is not available ([best_ajcc_stage_cd] ≠ 88, 99)
  - Diagnosis is not a hematopoietic or lymphoid neoplasm or pre-malignancy ([ca_heme_malig] = “No”).
- Preferred derived variable [stage_dx] incorporates both tumor registry and curated stage.

Curated Group Stage at Diagnosis

[ca_stage]

Value (Character)

- 0
- 0A
- 0is
- I
- IA
- IB
- IC
- II
- IIA
- IIB
- IIC
- III
- IIIA
- IIIB
- IIIC
- IV
- IVA
- IVB
- IVC
- Not Applicable
- Unknown

Description

- Curated stage group at diagnosis
- Populated only if:
  - Cancer diagnosis is curated. This field is blank for cancer diagnoses obtained from the tumor registry.
  - Cancer diagnosis was not diagnosed at stage IV ([ca_stage_iv] = “No”).
- Preferred derived variable [stage_dx] incorporates both tumor registry and curated stage.

Derived Stage at Diagnosis

[stage_dx]

Value (Character)

- Stage 0
- Stage I
- Stage II
- Stage III
- Stage I-III NOS
- Stage IV

Description

- Cancer stage at diagnosis
- Derived from a combination of tumor registry and curated stage variables based on the following hierarchy:
  - Tumor registry best group stage [best_ajcc_stage_cd] based on the AJCC version documented [naaccr_tnm_edition_num]
  - If tumor registry best group stage was unavailable, curated group stage [ca_stage_iv], [ca_stage] based on AJCC version 7 was used
  - If tumor registry and curated group stage were both unavailable:
    - If patient received neoadjuvant chemotherapy or radiation therapy before pathological stage diagnosis, then clinical staging was used, if available. If unavailable, then pathologic staging was used
    - If patient did not receive neoadjuvant chemotherapy or radiation therapy before pathological stage diagnosis, then pathologic staging was used, if available. If unavailable, then clinical staging was used
- Active surveillance (watchful waiting) at initial diagnosis is often used as a disease management strategy for patients with low-risk, localized prostate cancer. These patients may be followed for months or years prior to prostatectomy and pathological staging, and many will never undergo prostatectomy or pathological staging. For staging at initial diagnosis, active surveillance is considered a first course treatment and the overall stage for these patients is assigned based on clinical stage.
- Populated for BPC Project Cancers only

Derived Stage IV at Diagnosis

[stage_dx_iv]

Value (Character)

- Stage 0
- Stage I-III
- Stage IV

Description

- Grouped cancer stage at diagnosis
- Derived from tumor registry best group stage [best_ajcc_stage_cd], if available. If unavailable, curated group stage was used [ca_stage_iv].

Tumor Registry TNM Pathologic Stage

[naaccr_tnm_path_desc]

Value (Character)

- Free-text

Description

- TNM pathology from tumor registry
  - T describes the size of the tumor and any spread of cancer into nearby tissue
  - N describes spread of cancer to nearby lymph nodes
  - M describes metastasis
- Populated only if cancer diagnosis is obtained from the tumor registry. This field is blank for curated cancer diagnoses.

Data Standard: NAACCR #920

Tumor Registry Pathologic T Stage

[naaccr_path_t_cd]

Value (Character)

- Free-text

Description

- Pathologic T stage from tumor registry
- T describes the size of the tumor and any spread of cancer into nearby tissue
- Populated only if cancer diagnosis is obtained from the tumor registry. This field is blank for curated cancer diagnoses.

Data Standard: NAACCR #880

Curated Pathologic T Stage

[ca_path_t_stage]

Value (Character)

- TX
- T0
- Tis
- T1
- T2
- T3
- T4
- Not Applicable
- Unknown

Description

- T stage describes the size of the tumor and any spread of cancer into nearby tissue
- Populated only if:
    - Cancer diagnosis is curated. This field is blank for cancer diagnoses obtained from the tumor registry.
    - Cancer was not diagnosed at stage IV ([ca_stage_iv] = “No”).

Data Standard: NAACCR #880

Curated Pathologic Tis Stage Detail

[ca_path_tis_det]

Value (Character)

- Tis LCIS
- Tis DCIS
- Tis Paget's

Description

- Curated pathologic Tis stage detail
- T stage describes the size of the tumor and any spread of cancer into nearby tissue
- Populated only if:
    - Cancer diagnosis is curated. This field is blank for cancer diagnoses obtained from the tumor registry.
    - Pathologic T stage is Tis ([ca_path_t_stage] = “Tis”).

Curated Pathologic T1 Stage Detail

[ca_path_t1_det]

Value (Character)

- T1mic
- T1a
- T1a1
- T1a2
- T1b
- T1b1
- T1b2
- T1c
- T1d
- Not Applicable
- Unknown

Description

- Curated pathologic T1 stage detail
- T stage describes the size of the tumor and any spread of cancer into nearby tissue
- Populated only if:
  - Cancer diagnosis is curated. This field is blank for cancer diagnoses obtained from the tumor registry.
  - Pathologic T stage is T1 ([ca_path_t_stage] = “T1”).

Data Standard: NAACCR #880

Curated Pathologic T2 Stage Detail

[ca_path_t2_det]

Value (Character)

- T2a
- T2a1
- T2a2
- T2b
- T2c
- T2d
- Not Applicable
- Unknown

Description

- Curated pathologic T2 stage detail
- T stage describes the size of the tumor and any spread of cancer into nearby tissue
- Populated only if:
  - Cancer diagnosis is curated. This field is blank for cancer diagnoses obtained from the tumor registry.
  - Pathologic T stage is T2 ([ca_path_t_stage] = “T2”).

Data Standard: NAACCR #880

Curated Pathologic T3 Stage Detail

[ca_path_t3_det]

Value (Character)

- T3a
- T3b
- T3c
- T3d
- Not Applicable
- Unknown

Description

- Curated pathologic T3 stage detail
- T stage describes the size of the tumor and any spread of cancer into nearby tissue
- Populated only if:
  - Cancer diagnosis is curated. This field is blank for cancer diagnoses obtained from the tumor registry.
  - Pathologic T stage is T3 ([ca_path_t_stage] = “T3”).

Data Standard: NAACCR #880

Curated Pathologic T4 Stage Detail

[ca_path_t4_det]

Value (Character)

- T4a
- T4b
- T4c
- T4d
- T4e
- Not Applicable
- Unknown

Description

- Curated pathologic T4 stage detail
- T stage describes the size of the tumor and any spread of cancer into nearby tissue
- Populated only if:
  - Cancer diagnosis is curated. This field is blank for cancer diagnoses obtained from the tumor registry.
  - Pathologic T stage is T4 ([ca_path_t_stage] = “T4”).

Data Standard: NAACCR #880

Tumor Registry Pathologic N Stage

[naaccr_path_n_cd]

Value (Character)

- Free-text

Description

- Pathologic N stage from tumor registry
- N describes spread of cancer to nearby lymph nodes
- Populated only if cancer diagnosis is obtained from the tumor registry. This field is blank for curated cancer diagnoses.

Data Standard: NAACCR #890

Curated Pathologic N Stage

[ca_path_n_stage]

Value (Character)

- NX
- N0
- N1
- N2
- N3
- N4
- Not Applicable
- Unknown

Description

- Curated pathologic N stage
- N describes spread of cancer to nearby lymph nodes
- Populated only if:
  - Cancer diagnosis is curated. This field is blank for cancer diagnoses obtained from the tumor registry.
  - Cancers was not diagnosed at stage IV ([ca_stage_iv] = “No”).

Data Standard: NAACCR #890

Curated Pathologic N0 Stage Detail

[ca_path_n0_det]

Value (Character)

- N0(i-)
- N0(i+)
- N0(mol-)
- N0(mol+)
- Not Applicable
- Unknown

Description

- Curated pathologic N0 stage detail
- N describes spread of cancer to nearby lymph nodes
- Populated only if:
  - Cancer diagnosis is curated. This field is blank for cancer diagnoses obtained from the tumor registry.
  - Pathologic N stage is N0 ([ca_path_n_stage] = “N0”).

Curated Pathologic N1 Stage Detail

[ca_path_n1_det]

Value (Character)

- N1mi
- N1a
- N1b
- N1c
- Not Applicable
- Unknown

Description

- Curated pathologic N1 stage detail
- N describes spread of cancer to nearby lymph nodes
- Populated only if:
  - Cancer diagnosis is curated. This field is blank for cancer diagnoses obtained from the tumor registry.
  - Pathologic N stage is N1 ([ca_path_n_stage] = “N1”).

Data Standard: NAACCR #890

Curated Pathologic N2 Stage Detail

[ca_path_n2_det]

Value (Character)

- N2a
- N2b
- N2c
- Not Applicable
- Unknown

Description

- Curated pathologic N2 stage detail
- N describes spread of cancer to nearby lymph nodes
- Populated only if:
    - Cancer diagnosis is curated. This field is blank for cancer diagnoses obtained from the tumor registry.
    - Pathologic N stage is N2 ([ca_path_n_stage] = “N2”).

Data Standard: NAACCR #890

Curated Pathologic N3 Stage Detail

[ca_path_n3_det]

Value (Character)

- N3a
- N3b
- N3c
- Not Applicable
- Unknown

Description

- Curated pathologic N3 stage detail
- N describes spread of cancer to nearby lymph nodes
- Populated only if:
    - Cancer diagnosis is curated. This field is blank for cancer diagnoses obtained from the tumor registry.
    - Pathologic N stage is N3 ([ca_path_n_stage] = “N3”).

Data Standard: NAACCR #890

Tumor Registry Pathologic M Stage

[naaccr_path_m_cd]

Value (Character)

- Free-text

Description

- Pathologic M stage from tumor registry
- M describes metastasis
- Populated only if cancer diagnosis is obtained from the tumor registry. This field is blank for curated cancer diagnoses.

Data Standard: NAACCR #900

Tumor Registry Pathologic Group Stage

[naaccr_path_stage_cd]

Value (Character)

- Free-text

Description

- Pathologic group stage from tumor registry
- Populated only if cancer diagnosis is obtained from the tumor registry. This field is blank for curated cancer diagnoses.

Data Standard: NAACCR #910

Curated Pathologic Group Stage

[ca_path_group_stage]

Value (Character)

- 0
- 0A
- 0is
- I
- IA
- IA1
- IA2
- IB

- IB1
- IB2
- IC
- IS
- II
- IIA
- IIA1
- IIA2
- IIB
- IIC
- III
- IIIA
- IIIB
- IIIC
- IIIC1
- IIIC2
- IV
- IVA
- IVA1
- IVA2
- IVB
- IVC
- Occult
- Not Applicable
- Unknown

Description

- Populated only if:
  - Cancer diagnosis is curated. This field is blank for cancer diagnoses obtained from the tumor registry.
  - Cancer was not diagnosed at stage IV ([ca_stage_iv] = “No”).

Data Standard: NAACCR #910

Tumor Registry Clinical T Stage

[naaccr_clin_t_cd]

Value (Character)

- Free-text

Description

- Clinical T stage from tumor registry
- T describes the size of the tumor and any spread of cancer into nearby tissue

- Populated only if cancer diagnosis is obtained from the tumor registry. This field is blank for curated cancer diagnoses.

Data Standard: NAACCR #940

Curated Clinical T Stage

[ca_clin_t_stage]

Value (Character)

- TX
- T0
- Tis
- T1
- T2
- T3
- T4
- Not Applicable
- Unknown

Description

- Curated clinical T stage
- T stage describes the size of the tumor and any spread of cancer into nearby tissue
- Populated only if:
  - Cancer diagnosis is curated. This field is blank for cancer diagnoses obtained from the tumor registry.
  - If cancer was not diagnosed at stage IV ([ca_stage_iv] = “No”).

Data Standard: NAACCR #940

Curated Clinical Tis Stage Detail

[ca_clin_tis_det]

Value (Character)

- Tis LCIS
- Tis DCIS
- Tis Paget's

Description

- Curated clinical Tis stage detail
- T stage describes the size of the tumor and any spread of cancer into nearby tissue
- Populated only if:

  - Cancer diagnosis is curated. This field is blank for cancer diagnoses obtained from the tumor registry.
  - Curated clinical T stage is Tis ([ca_clin_t_stage] = “Tis”).

Curated Clinical T1 Stage Detail

[ca_clin_t1_det]

Value (Character)

- T1mic
- T1a
- T1a1
- T1a2
- T1b
- T1b1
- T1b2
- T1c
- T1d
- Not Applicable
- Unknown

Description

- Curated clinical T1 stage detail
- T stage describes the size of the tumor and any spread of cancer into nearby tissue
- Populated only if:
  - Cancer diagnosis is curated. This field is blank for cancer diagnoses obtained from the tumor registry.
  - Curated clinical T stage is T1 ([ca_clin_t_stage] = “T1”).

Data Standard: NAACCR #940

Curated Clinical T2 Stage Detail

[ca_clin_t2_det]

Value (Character)

- T2a
- T2a1
- T2a2
- T2b
- T2c
- T2d

- Not Applicable
- Unknown

Description

- Curated clinical T2 stage detail
- T stage describes the size of the tumor and any spread of cancer into nearby tissue
- Populated only if:
    - Cancer diagnosis is curated. This field is blank for cancer diagnoses obtained from the tumor registry.
    - Curated clinical T stage is T2 ([ca_clin_t_stage] = “T2”).

Data Standard: NAACCR #940

Curated Clinical T3 Stage Detail

[ca_clin_t3_det]

Value (Character)

- T3a
- T3b
- T3c
- T3d
- T3e
- Not Applicable
- Unknown

Description

- Curated clinical T3 stage detail
- T stage describes the size of the tumor and any spread of cancer into nearby tissue
- Populated only if:
    - Cancer diagnosis is curated. This field is blank for cancer diagnoses obtained from the tumor registry.
    - Curated clinical T stage is T3 ([ca_clin_t_stage] = “T3”).

Data Standard: NAACCR #940

Curated Clinical T4 Stage Detail

[ca_clin_t4_det]

Value (Character)

- T4a
- T4b

- T4c
- T4d
- T4e
- Not Applicable
- Unknown

Description

- Curated clinical T4 stage detail
- T stage describes the size of the tumor and any spread of cancer into nearby tissue
- Populated only if:
    - Cancer diagnosis is curated. This field is blank for cancer diagnoses obtained from the tumor registry.
    - Curated clinical T stage is T4 ([ca_clin_t_stage] = “T4”).

Data Standard: NAACCR #940

Tumor Registry Clinical N Stage

[naaccr_clin_n_cd]

Value (Character)

- Free-text

Description

- Clinical N stage from tumor registry
- N describes spread of cancer to nearby lymph nodes
- Populated only if cancer diagnosis is obtained from the tumor registry. This field is blank for curated cancer diagnoses.

Data Standard: NAACCR #950

Curated Clinical N Stage

[ca_clin_n_stage]

Value (Character)

- NX
- N0
- N1
- N2
- N3
- N4

- Not Applicable
- Unknown

Description

- Curated clinical N stage
- N describes spread of cancer to nearby lymph nodes
- Populated only if:
    - Cancer diagnosis is curated. This field is blank for cancer diagnoses obtained from the tumor registry.
    - Cancer was not diagnosed at stage IV ([ca_stage_iv] = “No”).

Data Standard: NAACCR #950

Curated Clinical N1 Stage Detail

[ca_clin_n1_det]

Value (Character)

- N1a
- N1b
- N1c
- Not Applicable
- Unknown

Description

- Curated clinical N1 stage detail
- N describes spread of cancer to nearby lymph nodes
- Populated only if:
    - Cancer diagnosis is curated. This field is blank for cancer diagnoses obtained from the tumor registry.
    - Curated clinical N stage is N1 ([ca_clin_n_stage] = “N1”).

Data Standard: NAACCR #950

Curated Clinical N2 Stage Detail

[ca_clin_n2_det]

Value (Character)

- N2a
- N2b
- N2c

- Not Applicable
- Unknown

Description

- Curated clinical N2 stage detail
- N describes spread of cancer to nearby lymph nodes
- Populated only if:
  - Cancer diagnosis is curated. This field is blank for cancer diagnoses obtained from the tumor registry.
  - Curated clinical N stage is N2 ([ca_clin_n_stage] = “N2”).

Data Standard: NAACCR #950

Curated Clinical N3 Stage Detail

[ca_clinical_n3_det]

Value (Character)

- N3a
- N3b
- N3c
- Not Applicable
- Unknown

Description

- Curated clinical N3 stage detail
- N describes spread of cancer to nearby lymph nodes
- Populated only if:
  - Cancer diagnosis is curated. This field is blank for cancer diagnoses obtained from the tumor registry.
  - Curated clinical N stage is N3 ([ca_clin_n_stage] = “N3”).

Data Standard: NAACCR #950

Tumor Registry Clinical M Stage

[naaccr_clin_m_cd]

Value (Character)

- Free-text

Description

- Clinical M stage from tumor registry
- M describes metastasis
- Populated only if cancer diagnosis is obtained from the tumor registry. This field is blank for curated cancer diagnoses.

Data Standard: NAACCR #960

Tumor Registry Clinical Group Stage

[naaccr_clin_stage_cd]

Value (Character)

- Free-text

Description

- Clinical group stage from tumor registry
- Populated only if cancer diagnosis is obtained from the tumor registry. This field is blank for curated cancer diagnoses.

Data Standard: NAACCR #970

Tumor Registry General Summary Stage

[naaccr_seer_sum_stage]

Value (Numeric)

- 0 = In situ
- 1 = Localized
- 2= Regional, direct extension only
- 3 = Regional, regional lymph nodes only
- 4 = Regional, direct extension and regional lymph nodes
- 5 = Regional, NOS
- 7 = Distant
- 8 = Not applicable
- 9 = Unstaged

Description

- General summary stage from tumor registry
- Populated only if cancer diagnosis is obtained from the tumor registry. This field is blank for curated cancer diagnoses.

Data Standard: NAACCR #759

Neoadjuvant Chemotherapy or Radiation Therapy Before Pathologic Stage Diagnosis

[ca_tx_pre_path_stage]

Value (Character)

- Yes
- No
- Not Applicable
- Unknown

Description

- Indicates whether patient received neoadjuvant chemotherapy or radiation therapy before pathologic stage diagnosis
- Populated only if cancer was not diagnosed at stage IV ([ca_stage_iv] = “No”)

Data Standard: NAACCR #920

Curated General Summary Stage

[ca_gen_sum_stage_2]

Value (Character)

- Unstaged
- In situ
- Localized
- Regional direct extension only
- Regional lymph nodes only
- Regional direct extension and regional lymph nodes
- Regional NOS
- Distant
- Not applicable

Description

- Summary stage includes all information through completion of surgery in the first course of treatment or within 4 months of diagnosis in the absence of disease progression, whichever is longer.
- Populated only if:
  - Cancer is non-BPC Project Cancer
  - Cancer diagnosis is curated. This field is blank for cancer diagnoses obtained from the tumor registry.
  - Cancer type is Neuroblastoma, Brain Cancer, Ewing Sarcoma, Retinoblastoma,

Rhabdomyosarcoma, or Wilms Tumor ([ca_type] = “Neuroblastoma”, “Brain Cancer”, “Ewing Sarcoma”, “Retinoblastoma”, “Rhabdomyosarcoma”, or “Wilms Tumor”).

Data Standard: NAACCR #759

Tumor Registry TNM Edition Number

[naaccr_tnm_edition_num]

Value (Character)

- Free-text

Description

- TNM edition number corresponding to stage variables from tumor registry
- Populated only if cancer diagnosis is obtained from the tumor registry. This field is blank for curated cancer diagnoses.

Data Standard: NAACCR #1060

Presence of Distant Metastasis at the Time of Cancer Diagnosis (Stage IV Diagnoses)

[ca_dmets_yn]

Value (Character)

- Yes
- No - patient is stage IV with no distant metastases
- Unknown or Not mentioned

Description

- Indicates whether stage IV patient had distant metastases at time of cancer diagnosis
- Populated only if cancer diagnosed at stage IV ([best_ajcc_stage_cd] = “4”, “4A” “4B” or [ca_stage_iv] = “Yes”)

Sites of Distant Metastases at Cancer Diagnosis (Stage IV Diagnoses)

[ca_first_dmets1]-[ca_first_dmets10]

Value (Character)

- ICD-O-3 topography code

Description

- Site of distant metastases at diagnosis
- Up to 10 sites of distant metastasis at diagnosis are recorded
- Populated only if distant metastases are present at time of diagnosis ([ca_dmets_yn] = “Yes”)

Breast Cancer: ER Summary Status

[ca_bca_er]

Value (Character)

- Positive/elevated
- Negative/normal
- Borderline
- Not applicable
- Test ordered/results not interpretable
- Test ordered/results not in chart
- Test not done
- Unknown or no information

Description

- Breast cancer estrogen receptor (ER) summary status
- Populated only if breast cancer diagnosis

Data Standard: NAACCR #2880

Breast Cancer: PR Summary Status

[ca_bca_pr]

Value (Character)

- Positive/elevated
- Negative/normal
- Borderline
- Not applicable
- Test ordered/results not interpretable
- Test ordered/results not in chart
- Test not done
- Unknown or no information

Description

- Breast cancer progesterone receptor (PR) summary status
- Populated only if breast cancer diagnosis

Data Standard: NAACCR #2890

Breast Cancer: HER2 ImmunoHistoChemistry (IHC) Lab Value

[ca_bca_her2ihc_val]

Value (Character)

- Score 0
- Score of 1+
- Score of 2+
- Score of 3+
- Not applicable
- Test ordered/results not in chart
- Test not done
- Unknown or no information

Description

- Breast cancer human epidermal growth factor receptor 2 (HER2) lab value
- Populated only if breast cancer diagnosis

Data Standard: NAACCR #2862

Breast Cancer: HER2 ImmunoHistoChemistry (IHC) Lab Interpretation

[ca_bca_her2ihc_intp]

Value (Character)

- Positive/elevated
- Negative/normal
- Borderline; equivocal; indeterminate; undetermined whether positive or negative
- Not applicable
- Test ordered/results not in chart
- Test not done
- Unknown or no information

Description

- Breast cancer human epidermal growth factor receptor 2 (HER2) immunohistochemistry lab interpretation
- Populated only if breast cancer diagnosis

Data Standard: NAACCR #2863

Breast Cancer: HER2 Summary Result

[ca_bca_her_summ]

Value (Character)

- Positive/elevated/amplified
- Negative/normal within normal limits
- Borderline/equivocal/indeterminant
- Not applicable
- Test ordered results not in chart
- Tests not done
- Unknown or no information

Description

- Breast cancer human epidermal growth factor receptor 2 (HER2) summary result
- Populated only if breast cancer diagnosis

Data Standard: NAACCR #2869

Breast Cancer: Oncotype DX Recurrence Score-Invasive

[ca_bca_oncotypedx]

Value (Character)

- 0-100
- Stated as less than 11
- Stated as equal to or greater than 11
- Not applicable; in situ case
- Test ordered, results not in chart
- Not documented in medical record

Description

- Breast cancer Oncotype DX recurrence score-invasive
- Populated only if breast cancer diagnosis

Data Standard: NAACCR #3904

Breast Cancer: Multigene Signature Method

[ca_bca_mgene]

Value (Character)

- Mammaprint
- PAM50 (Prosigna)
- Breast Cancer Index
- EndoPredict
- Test performed type of test unknown
- Multiple tests any tests in codes 1-4
- Test ordered results not in chart
- Not applicable

Description

- Breast cancer multigene signature method
- Multiple tests any tests in codes 1-4 refers to any tests in codes Mammaprint, PAM50 (Prosigna), Breast Cancer Index, EndoPredict
- Populated only if breast cancer diagnosis

Data Standard: NAACCR #3894

Breast Cancer: Multigene Signature Results

[ca_bca_mgeneresult]

Value (Character)

- 0-99
- Score 100
- Low risk
- Moderate risk
- High risk
- Test ordered results no results
- Not applicable
- Not documented in medical record

Description

- Breast cancer multigene signature results
- Populated only if breast cancer diagnosis

Data Standard: NAACCR #3895

Breast Cancer: HER2 In Situ Hybridization (ISH) Summary

[ca_bca_herish]

Value (Character)

- Positive [amplified]
- Negative not amplified
- Equivocal
- Test ordered, results not in chart
- Not documented in medical record

Description

- Breast cancer human epidermal growth factor receptor 2 (HER2) in situ hybridization lab interpretation
- Populated only if breast cancer diagnosis

Data Standard: NAACCR #3854

Breast Cancer Subtype

[bca_subtype]

Value (Character)

- Triple Negative
- HR-, HER2+
- HR+, HER2-
- HR+, HER2+

Description

- Breast cancer subtype characterized by HER2 and hormone receptor status
- HER2 status is positive if HER2 summary status is positive ([ca_bca_her_summ] = "Positive/elevated/amplified").
- Hormone receptor status is positive if at least one of ER or PR is positive ([ca_bca_er] = “Positive/elevated” or [ca_bca_pr] = “Positive/elevated). Hormone receptor status is negative if both ER and PR are negative.
- Populated only if breast cancer diagnosis (excluding breast sarcoma, [ca_type] = “Breast Sarcoma”)

Colorectal Cancer: Tumor Deposits

[ca_crc_td]

Value (Character)

- 0-99
- 100 or more
- Tumor Deposits identified, number unknown
- Not applicable: Information not collected for this case
- Not documented in medical record; Cannot be determined by the pathologist; Pathology report

does not mention tumor deposits; No surgical resection done; Tumor Deposits not assessed or unknown if assessed

Description

- Count of colorectal cancer tumor deposits
- Populated only if colon/rectum cancer diagnosis

Data Standard: NAACCR #3934

Colorectal Cancer: Tumor Deposits

[crc_td]

Value (Character)

- 0 = surgical resection of the primary site is performed, the pathology report is available for review, and tumor deposits are not mentioned
- 1-99
- 100 or more
- Tumor Deposits identified, number unknown
- Not applicable: Information not collected for this case
- Not documented in medical record; Cannot be determined by the pathologist; Pathology report does not mention tumor deposits; No surgical resection done; Tumor Deposits not assessed or unknown if assessed

Description

- Count of colorectal cancer tumor deposits
- Populated only if colon/rectum cancer diagnosis

Data Standard: NAACCR #3934

Colorectal Cancer: Circumferential Radial Margin

[ca_crc_crm]

Value (Character)

- 0-99
- Margins cannot be assessed
- Described as at least 1 mm
- Described as at least 2 mm
- Described as at least 3 mm
- Described as greater than 3 mm
- No resection of primary site
- 100 mm or greater

- Margins clear distance from tumor not stated
- Not applicable
- Not documented in medical record

Description

- Colorectal cancer circumferential radial margin
- Populated only if colon/rectum cancer diagnosis

Data Standard: NAACCR #2930

Colorectal Cancer: Perineural Invasion

[ca_crc_peri_inv]

Value (Character)

- None
- Perineural invasion present
- Not applicable
- No histologic examination
- Unknown not documented

Description

- Colorectal cancer perineural invasion status
- Populated only if colon/rectum cancer diagnosis

Data Standard: NAACCR #2862

Prostate Cancer: Gleason Patterns Clinical

[ca_pros_clin_gpattern2]

Value (Character)

- Primary pattern 1 secondary pattern 1
- Primary pattern 1 secondary pattern 2
- Primary pattern 1 secondary pattern 3
- Primary pattern 1 secondary pattern 4
- Primary pattern 1 secondary pattern 5
- Primary pattern 1 secondary pattern unknown
- Primary pattern 2 secondary pattern 1
- Primary pattern 2 secondary pattern 2
- Primary pattern 2 secondary pattern 3
- Primary pattern 2 secondary pattern 4
- Primary pattern 2 secondary pattern 5

- Primary pattern 2 secondary pattern unknown
- Primary pattern 3 secondary pattern 1
- Primary pattern 3 secondary pattern 2
- Primary pattern 3 secondary pattern 3
- Primary pattern 3 secondary pattern 4
- Primary pattern 3 secondary pattern 5
- Primary pattern 3 secondary pattern unknown
- Primary pattern 4 secondary pattern 1
- Primary pattern 4 secondary pattern 2
- Primary pattern 4 secondary pattern 3
- Primary pattern 4 secondary pattern 4
- Primary pattern 4 secondary pattern 5
- Primary pattern 4 secondary pattern unknown
- Primary pattern 5 secondary pattern 1
- Primary pattern 5 secondary pattern 2
- Primary pattern 5 secondary pattern 3
- Primary pattern 5 secondary pattern 4
- Primary pattern 5 secondary pattern 5
- Primary pattern 5 secondary pattern unknown
- Primary pattern unknown, secondary pattern unknown
- No needle core biopsy/TURP performed
- Not applicable information not collected for this case
- Not documented in medical record

Description

- Populated only if prostate cancer diagnosis

Data Standard: NAACCR #2861

Prostate Cancer: Gleason Score Clinical

[ca_pros_clin_gscore]

Value (Character)

- Gleason score 2
- Gleason score 3
- Gleason score 4
- Gleason score 5
- Gleason score 6
- Gleason score 7
- Gleason score 8
- Gleason score 9
- Gleason score 10
- No needle core biopsy/TURP performed

- Not applicable
- Not documented in medical record

Description

- Populated only if prostate cancer diagnosis
- For patients with multiple biopsies performed prior to prostatectomy, this variable reflects the highest Gleason score during the staging period

Data Standard: NAACCR #2862

Prostate Cancer: Gleason Patterns Pathological

[ca_pros_path_gpattern2]

Value (Character)

- Primary pattern 1 secondary pattern 1
- Primary pattern 1 secondary pattern 2
- Primary pattern 1 secondary pattern 3
- Primary pattern 1 secondary pattern 4
- Primary pattern 1 secondary pattern 5
- Primary pattern 1 secondary pattern unknown
- Primary pattern 2 secondary pattern 1
- Primary pattern 2 secondary pattern 2
- Primary pattern 2 secondary pattern 3
- Primary pattern 2 secondary pattern 4
- Primary pattern 2 secondary pattern 5
- Primary pattern 2 secondary pattern unknown
- Primary pattern 3 secondary pattern 1
- Primary pattern 3 secondary pattern 2
- Primary pattern 3 secondary pattern 3
- Primary pattern 3 secondary pattern 4
- Primary pattern 3 secondary pattern 5
- Primary pattern 3 secondary pattern unknown
- Primary pattern 4 secondary pattern 1
- Primary pattern 4 secondary pattern 2
- Primary pattern 4 secondary pattern 3
- Primary pattern 4 secondary pattern 4
- Primary pattern 4 secondary pattern 5
- Primary pattern 4 secondary pattern unknown
- Primary pattern 5 secondary pattern 1
- Primary pattern 5 secondary pattern 2
- Primary pattern 5 secondary pattern 3
- Primary pattern 5 secondary pattern 4
- Primary pattern 5 secondary pattern 5
- Primary pattern 5 secondary pattern unknown
- Primary pattern unknown, secondary pattern unknown

- No prostatectomy performed
- Not applicable information not collected for this case
- Not documented in medical record

Description

- Populated only if prostate cancer diagnosis

Data Standard: NAACCR #2863

Prostate Cancer: Gleason Score Pathological

[ca_pros_path_gscore]

Value (Character)

- Gleason score 2
- Gleason score 3
- Gleason score 4
- Gleason score 5
- Gleason score 6
- Gleason score 7
- Gleason score 8
- Gleason score 9
- Gleason score 10
- No prostatectomy done
- Not applicable
- Not documented in medical record

Description

- Populated only if prostate cancer diagnosis
- Reflects the Gleason score reported from a prostatectomy pathology report or clinician report if the pathology report is unavailable

Data Standard: NAACCR #2864

Prostate Cancer: Gleason Tertiary Pattern

[ca_pros_path_gtert]

Value (Character)

- Tertiary pattern 1
- Tertiary pattern 2
- Tertiary pattern 3
- Tertiary pattern 4

- Tertiary pattern 5
- No prostatectomy/autopsy performed
- Not applicable
- Not documented in medical record

Description

- Populated only if prostate cancer diagnosis

Data Standard: NAACCR #2865

Prostate Cancer: Number of Cores Positive

[ca_pros_num_corespos]

Value (Character)

- 00
- 1-99
- 100 or more cores
- Biopsy cores examined, number unknown
- No needle core biopsy performed
- Not applicable information not collected for this case
- Not documented in medical record

Description

- Populated only if prostate cancer diagnosis

Data Standard: NAACCR #2866

Prostate Cancer: Number of Cores Examined

[ca_pros_num_cores]

Value (Character)

- 1-99
- 100 or more cores
- Biopsy cores examined, number unknown
- No needle core biopsy performed
- Not applicable information not collected for this case
- Not documented in medical record

Description

- Populated only if prostate cancer diagnosis

Data Standard: NAACCR #2867

Cigarette Use at Time of Diagnosis

[ca_cigarette]

Value (Character)

- Never used
- Current user
- Former user (quit < 1 year)
- Former user (quit >1 year)
- Former user (unknown time)
- Unknown

Description

- Smoking status of patient at time of diagnosis
- Populated only if lung or bladder cancer diagnosis (BPC Project and non-BPC Project Cancer diagnoses).

Data Standard: National Program of Cancer Registries (Reference: Capture of tobacco use among population-based registries: Findings from 10 National Program of Cancer Registries states)

Non-Small Cell Lung Cancer: Separate Tumor Nodules

[ca_lung_sep_tumor]

Value (Character)

- No separate tumor nodules
- Separate tumor nodules of same histologic type in ipsilateral lung same lobe
- Separate tumor nodules of same histologic type in ipsilateral lung, different lobe
- Separate tumor nodules of same histologic type in ipsilateral lung same AND different lobes
- Separate tumor nodules of same histologic type in ipsilateral lung unknown if same or different lobe(s)
- Not applicable
- Unknown if separate tumor nodules

Description

- Separate tumor nodules are defined as intrapulmonary metastasis identified in the same lobe or same lung (ipsilateral) originating from a single lung primary at the time of diagnosis. Biopsy of tumors may or may not be performed.
- Populated only if NSCLC diagnosis (BPC Project and non-BPC Project Cancer diagnoses).

Data Standard: NAACCR #2880

Non-Small Cell Lung Cancer: Pleural/Elastic Layer Invasion (PL)

[ca_lung_pl_el_inv]

Value (Character)

- PL 0; No evidence of visceral pleural invasion (PL)
- PL 1; Invasion beyond the visceral elastic pleura, but limited to the pulmonary pleura
- PL 2; Invasion to the surface of the pulmonary pleura
- PL 3; Tumor extends to the parietal pleura
- Invasion of pleura
- No histologic examination of pleura to assess pleural layer invasion
- Unknown if PL present; PL/elastic layer cannot be assessed; Not documented in patient record

Description

- Pleural/Elastic Layer Invasion (PL) by Hematoxylin and Eosin stain (H & E) or Electric Strain
- Populated only if NSCLC diagnosis (BPC Project and non-BPC Project Cancer diagnoses).

Data Standard: NAACCR #2890

Number of Regimens Associated with the Cancer Diagnosis

[ca_n_regimens]

Value (Numeric)

Description

- Count of cancer-directed regimens that were associated with the cancer diagnosis

Distant Metastasis Post Diagnosis

[dmets_post_dx]

Value (Numeric)

- 0 = No distant metastasis
- 1 = Distant metastasis

Description

- Indicates the presence of distant metastasis post-diagnosis according to a radiology or pathology report
- Distant metastasis classified based on Appendix Table 4

- Populated only among patients without distant metastasis at diagnosis and if cancer diagnosis is the patient's first BPC Project Cancer

Time (Days, Months, Years) from Diagnosis of Stage I-III to Distant Metastasis

[dx_to_dmets_days], [dx_to_dmets_mos], [dx_to_dmets_yrs]

Value (Numeric)

Description

- Interval in days [dx_to_dmets_days]; months [dx_to_dmets_mos]; or years [dx_to_dmets_yrs] from diagnosis of stage I-III BPC Project Cancer to date of first indication of distant metastasis
- Populated only among patients stage I-III at diagnosis with distant metastasis post diagnosis ([dmets_post_dx] = 1) and if cancer diagnosis is the patient's first BPC Project Cancer
- Note that this variable cannot be used as a time-to-event endpoint as it does not account for death or censoring.

Distant Metastasis: Abdomen

[dist_mets_abdomen]

Value (Numeric)

- 0 = No distant metastasis, abdomen
- 1 = Distant metastasis, abdomen

Description

- Indicates the presence of distant metastasis in the abdomen post-diagnosis according to a radiology or pathology report (Appendix Table 4)
- For stage I-III patients, indicates the presence of distant metastasis in the abdomen post-diagnosis
- For stage IV patients, indicates the presence of distant metastasis in the abdomen at the time of diagnosis or post-diagnosis
- Populated only if cancer diagnosis is the patient's first BPC Project Cancer

Time (Days, Months, Years) from Diagnosis to Distant Metastasis in Abdomen

[dx_to_dist_mets_abdomen_days], [dx_to_dist_mets_abdomen_mos], [dx_to_dist_mets_abdomen_yrs]

Value (Numeric)

Description

- Interval in days [dx_to_dmets_abdomen_days]; months [dx_to_dmets_abdomen_mos]; or years [dx_to_dmets_abdomen_yrs] from diagnosis to date of first distant metastasis in the abdomen
- For stage IV patients with distant metastasis in the abdomen at diagnosis, time in days, months, and years will be 0.
- Populated only among patients with distant metastasis in the abdomen ([dmets_abdomen] = 1) and if cancer diagnosis is the patient's first BPC Project Cancer
- Note that this variable cannot be used as a time-to-event endpoint as it does not account for death or censoring.

Distant Metastasis: Adrenal

[dist_mets_adrenal]

Value (Numeric)

- 0 = No distant metastasis, adrenal gland
- 1 = Distant metastasis, adrenal gland

Description

- Indicates the presence of distant adrenal metastasis post-diagnosis according to a radiology or pathology report (Appendix Table 4)
- For stage I-III patients, indicates the presence of distant adrenal metastasis post-diagnosis
- For stage IV patients, indicates the presence of distant adrenal metastasis at the time of diagnosis or post-diagnosis
- Populated only if cancer diagnosis is the patient's first BPC Project Cancer

Time (Days, Months, Years) from Diagnosis to Adrenal Distant Metastasis

[dx_to_dist_mets_adrenal_days], [dx_to_dist_mets_adrenal_mos], [dx_to_dist_mets_adrenal_yrs]

Value (Numeric)

Description

- Interval in days [dx_to_dist_mets_adrenal_days]; months [dx_to_dist_mets_adrenal_mos]; or years [dx_to_dist_mets_adrenal_yrs] from diagnosis to first distant adrenal metastasis
- For stage IV patients with distant adrenal metastasis at diagnosis, time in days, months, and years will be 0
- Populated only among patients with distant adrenal metastasis ([dist_mets_adrenal] = 1) and if cancer diagnosis is the patient's first BPC Project Cancer
- Note that this variable cannot be used as a time-to-event endpoint as it does not account for death or censoring

Distant Metastasis: Bone

[dist_mets_bone]

Value (Numeric)

- 0 = No distant metastasis, bone
- 1 = Distant metastasis, bone

Description

- Indicates the presence of distant metastasis in the bone post-diagnosis according to a radiology or pathology report (Appendix Table 4)
- For stage I-III patients, indicates the presence of distant metastasis in the bone post-diagnosis
- For stage IV patients, indicates the presence of distant metastasis in the bone at the time of diagnosis or post-diagnosis
- Populated only if cancer diagnosis is the patient's first BPC Project Cancer

Time (Days, Months, Years) from Diagnosis to Distant Metastasis in Bone

[dx_to_dist_mets_bone_days], [dx_to_dist_mets_bone_mos], [dx_to_dist_mets_bone_yrs]

Value (Numeric)

Description

- Interval in days [dx_to_dist_mets_bone_days]; months [dx_to_dist_mets_bone_mos]; or years [dx_to_dist_mets_bone_yrs] from diagnosis to first distant metastasis in the bone
- For stage IV patients with distant metastasis in the bone at diagnosis, time in days, months, and years will be 0
- Populated only among patients with distant metastasis in the bone ([dist_mets_bone] = 1) and if cancer diagnosis is the patient's first BPC Project Cancer
- Note that this variable cannot be used as a time-to-event endpoint as it does not account for death or censoring

Distant Metastasis: Bone Marrow

[dist_mets_bone_marrow]

Value (Numeric)

- 0 = No distant metastasis, bone marrow
- 1 = Distant metastasis, bone marrow

Description

- Indicates the presence of distant metastasis in the bone marrow post-diagnosis according to a radiology or pathology report (Appendix Table 4)
- For stage I-III patients, indicates the presence of distant metastasis in the bone marrow post-diagnosis
- For stage IV patients, indicates the presence of distant metastasis in the bone marrow at the time of diagnosis or post-diagnosis
- Populated only if cancer diagnosis is the patient's first BPC Project Cancer

Time (Days, Months, Years) from Diagnosis to Distant Metastasis in Bone Marrow

[dx_to_dist_mets_bone_marrow_days], [dx_to_dist_mets_bone_marrow_mos], [dx_to_dist_mets_bone_marrow_yrs]

Value (Numeric)

Description

- Interval in days [dx_to_dist_mets_bone_marrow_days]; months [dx_to_dist_mets_bone_marrow_mos]; or years [dx_to_dist_mets_bone_marrow_yrs] from diagnosis to first distant metastasis in the bone marrow
- For stage IV patients with distant metastasis in the bone marrow at diagnosis, time in days, months, and years will be 0
- Populated only among patients with distant metastasis in the bone marrow ([dist_mets_bone_marrow] = 1) and if cancer diagnosis is the patient's first BPC Project Cancer
- Note that this variable cannot be used as a time-to-event endpoint as it does not account for death or censoring

Distant Metastasis: Brain/Central Nervous System

[dist_mets_brain_cns]

Value (Numeric)

- 0 = No distant metastasis, brain/central nervous system
- 1 = Distant metastasis, brain/central nervous system

Description

- Indicates the presence of distant metastasis in the brain/central nervous system post-diagnosis according to a radiology or pathology report (Appendix Table 4)
- For stage I-III patients, indicates the presence of distant metastasis in the brain/central nervous system post-diagnosis
- For stage IV patients, indicates the presence of distant metastasis in the brain/central nervous system at the time of diagnosis or post-diagnosis
- Populated only if cancer diagnosis is the patient's first BPC Project Cancer

Time (Days, Months, Years) from Diagnosis to Distant Metastasis in Brain/Central Nervous System

[dx_to_dist_mets_brain_cns_days], [dx_to_dist_mets_brain_cns_mos], [dx_to_dist_mets_brain_cns_yrs]

Value (Numeric)

Description

- Interval in days [dx_to_dist_mets_brain_cns_days]; months [dx_to_dist_mets_brain_cns_mos]; or years [dx_to_dist_mets_brain_cns_yrs] from diagnosis to first distant metastasis in the brain/central nervous system
- For stage IV patients with distant metastasis in the brain/central nervous system at diagnosis, time in days, months, and years will be 0
- Populated only among patients with distant metastasis in the brain/central nervous system ([dist_mets_brain_cns] = 1) and if cancer diagnosis is the patient's first BPC Project Cancer
- Note that this variable cannot be used as a time-to-event endpoint as it does not account for death or censoring

Distant Metastasis: Breast

[dist_mets_breast]

Value (Numeric)

- 0 = No distant metastasis, breast
- 1 = Distant metastasis, breast

Description

- Indicates the presence of distant metastasis in the breast post-diagnosis according to a radiology or pathology report (Appendix Table 4)
- For stage I-III patients, indicates the presence of distant metastasis in the breast post-diagnosis
- For stage IV patients, indicates the presence of distant metastasis in the breast at the time of diagnosis or post-diagnosis
- Populated only if cancer diagnosis is the patient's first BPC Project Cancer

Time (Days, Months, Years) from Diagnosis to Distant Metastasis in Breast

[dx_to_dist_mets_breast_days], [dx_to_dist_mets_breast_mos], [dx_to_dist_mets_breast_yrs]

Value (Numeric)

Description

- Interval in days [dx_to_dist_mets_breast_days]; months [dx_to_dist_mets_breast_mos]; or years [dx_to_dist_breast_yrs] from diagnosis to first distant metastasis in the breast
- For stage IV patients with distant metastasis in the breast at diagnosis, time in days, months, and years will be 0
- Populated only among patients with distant metastasis in the breast ([dist_mets_breast] = 1) and if cancer diagnosis is the patient's first BPC Project Cancer
- Note that this variable cannot be used as a time-to-event endpoint as it does not account for death or censoring

Distant Metastasis: Head and Neck

[dist_mets_head_and_neck]

Value (Numeric)

- 0 = No distant metastasis, head and neck
- 1 = Distant metastasis, head and neck

Description

- Indicates the presence of distant metastasis in the head and neck post-diagnosis according to a radiology or pathology report (Appendix Table 4)
- For stage I-III patients, indicates the presence of distant metastasis in the head and neck post-diagnosis
- For stage IV patients, indicates the presence of distant metastasis in the head and neck at the time of diagnosis or post-diagnosis
- Populated only if cancer diagnosis is the patient's first BPC Project Cancer

Time (Days, Months, Years) from Diagnosis to Distant Metastasis in Head and Neck

[dx_to_dist_mets_head_and_neck_days], [dx_to_dist_mets_head_and_neck_mos], [dx_to_dist_mets_head_and_neck_yrs]

Value (Numeric)

Description

- Interval in days [dx_to_dist_mets_head_and_neck_days]; months [dx_to_dist_mets_head_and_neck_mos]; or years [dx_to_dist_head_and_neck_yrs] from diagnosis to first distant metastasis in the head and neck
- For stage IV patients with distant metastasis in the head and neck at diagnosis, time in days, months, and years will be 0
- Populated only among patients with distant metastasis in the head and neck

([dist_mets_head_and_neck] = 1) and if cancer diagnosis is the patient's first BPC Project Cancer
- Note that this variable cannot be used as a time-to-event endpoint as it does not account for death or censoring

Distant Metastasis: Liver

[dist_mets_liver]

Value (Numeric)

- 0 = No distant metastasis, liver
- 1 = Distant metastasis, liver

Description

- Indicates the presence of distant metastasis in the liver post-diagnosis according to a radiology or pathology report (Appendix Table 4)
- For stage I-III patients, indicates the presence of distant metastasis in the liver post-diagnosis
- For stage IV patients, indicates the presence of distant metastasis in the liver at the time of diagnosis or post-diagnosis
- Populated only if cancer diagnosis is the patient's first BPC Project Cancer

Time (Days, Months, Years) from Diagnosis to Distant Metastasis in Liver

[dx_to_dist_mets_liver_days], [dx_to_dist_mets_liver_mos], [dx_to_dist_mets_liver_yrs]

Value (Numeric)

Description

- Interval in days [dx_to_dist_mets_liver_days]; months [dx_to_dist_mets_liver_mos]; or years [dx_to_dist_liver_yrs] from diagnosis to first distant metastasis in the liver
- For stage IV patients with distant metastasis in the liver at diagnosis, time in days, months, and years will be 0
- Populated only among patients with distant metastasis in the liver ([dist_mets_liver] = 1) and if cancer diagnosis is the patient's first BPC Project Cancer
- Note that this variable cannot be used as a time-to-event endpoint as it does not account for death or censoring

Distant Metastasis: Lymph Nodes

[dist_mets_lymph_nodes]

Value (Numeric)

- 0 = No distant metastasis, lymph nodes
- 1 = Distant metastasis, lymph nodes

Description

- Indicates the presence of distant metastasis in the lymph nodes post-diagnosis according to a radiology or pathology report (Appendix Table 4)
- For stage I-III patients, indicates the presence of distant metastasis in the lymph nodes post-diagnosis
- For stage IV patients, indicates the presence of distant metastasis in the lymph nodes at the time of diagnosis or post-diagnosis
- Populated only if cancer diagnosis is the patient's first BPC Project Cancer

Time (Days, Months, Years) from Diagnosis to Distant Metastasis in Lymph Nodes

[dx_to_dist_mets_lymph_nodes_days], [dx_to_dist_mets_lymph_nodes_mos], [dx_to_dist_mets_lymph_nodes_yrs]

Value (Numeric)

Description

- Interval in days [dx_to_dist_mets_lymph_nodes_days]; months [dx_to_dist_mets_lymph_nodes_mos]; or years [dx_to_dist_lymph_nodes_yrs] from diagnosis to first distant metastasis in the lymph nodes
- For stage IV patients with distant metastasis in the lymph nodes at diagnosis, time in days, months, and years will be 0
- Populated only among patients with distant metastasis in the lymph_nodes ([dist_mets_lymph_nodes] = 1) and if cancer diagnosis is the patient's first BPC Project Cancer
- Note that this variable cannot be used as a time-to-event endpoint as it does not account for death or censoring

Distant Metastasis: Other

[dist_mets_other]

Value (Numeric)

- 0 = No distant metastasis, other
- 1 = Distant metastasis, other

Description

- Indicates the presence of other distant metastasis post-diagnosis according to a radiology or pathology report (Appendix Table 4)

- For stage I-III patients, indicates the presence of other distant metastasis post-diagnosis
- For stage IV patients, indicates the presence of other distant metastasis at the time of diagnosis or post-diagnosis
- Populated only if cancer diagnosis is the patient's first BPC Project Cancer

Time (Days, Months, Years) from Diagnosis to Other Distant Metastasis

[dx_to_dist_mets_other_days], [dx_to_dist_mets_other_mos], [dx_to_dist_mets_other_yrs]

Value (Numeric)

Description

- Interval in days [dx_to_dist_mets_other_days]; months [dx_to_dist_mets_other_mos]; or years [dx_to_dist_other_yrs] from diagnosis to first other distant metastasis
- For stage IV patients with other distant metastasis at diagnosis, time in days, months, and years will be 0
- Populated only among patients with other distant metastasis ([dist_mets_other] = 1) and if cancer diagnosis is the patient's first BPC Project Cancer
- Note that this variable cannot be used as a time-to-event endpoint as it does not account for death or censoring

Distant Metastasis: Pelvis

[dist_mets_pelvis]

Value (Numeric)

- 0 = No distant metastasis, pelvis
- 1 = Distant metastasis, pelvis

Description

- Indicates the presence of distant metastasis in the pelvis post-diagnosis according to a radiology or pathology report (Appendix Table 4)
- For stage I-III patients, indicates the presence of distant metastasis in the pelvis post-diagnosis
- For stage IV patients, indicates the presence of distant metastasis in the pelvis at the time of diagnosis or post-diagnosis
- Populated only if cancer diagnosis is the patient's first BPC Project Cancer

Time (Days, Months, Years) from Diagnosis to Distant Metastasis in Pelvis

[dx_to_dist_mets_pelvis_days], [dx_to_dist_mets_pelvis_mos], [dx_to_dist_mets_pelvis_yrs]

Value (Numeric)

Description

- Interval in days [dx_to_dist_mets_pelvis_days]; months [dx_to_dist_mets_pelvis_mos]; or years [dx_to_dist_pelvis_yrs] from diagnosis to first distant metastasis in the pelvis
- For stage IV patients with distant metastasis in the pelvis at diagnosis, time in days, months, and years will be 0
- Populated only among patients with distant metastasis in the pelvis ([dist_mets_pelvis] = 1) and if cancer diagnosis is the patient's first BPC Project Cancer
- Note that this variable cannot be used as a time-to-event endpoint as it does not account for death or censoring

Distant Metastasis: Pericardial and Malignant Pericardial Effusion

[dist_mets_pericardial_and_malignant_pericardial_effusion]

Value (Numeric)

- 0 = No distant metastasis, pericardium
- 1 = Distant metastasis, pericardium

Description

- Indicates the presence of distant metastasis in the pericardium post-diagnosis according to a radiology or pathology report (Appendix Table 4)
- For stage I-III patients, indicates the presence of distant metastasis in the pericardium post-diagnosis
- For stage IV patients, indicates the presence of distant metastasis in the pericardium at the time of diagnosis or post-diagnosis
- Populated only if cancer diagnosis is the patient's first BPC Project Cancer

Time (Days, Months, Years) from Diagnosis to Pericardial and Malignant Pericardial Effusion Distant Metastasis

[dx_to_dist_mets_pericardial_and_malignant_pericardial_effusion_days],
[dx_to_dist_mets_pericardial_and_malignant_pericardial_effusion_mos],
[dx_to_dist_mets_pericardial_and_malignant_pericardial_effusion_yrs]

Value (Numeric)

Description

- Interval in days [dx_to_dist_mets_pericardial_and_malignant_pericardial_effusion_days]; months [dx_to_dist_mets_pericardial_and_malignant_pericardial_effusion_mos]; or years [dx_to_dist_mets_pericardial_and_malignant_pericardial_effusion_yrs] from diagnosis to first distant metastasis in the pericardium
- For stage IV patients with distant metastasis in the pericardium at diagnosis, time in days,

months, and years will be 0
- Populated only among patients with distant metastasis in the pericardium ([dist_mets_pericardial_and_malignant_pericardial_effusion] = 1) and if cancer diagnosis is the patient's first BPC Project Cancer
- Note that this variable cannot be used as a time-to-event endpoint as it does not account for death or censoring

Distant Metastasis: Peritoneum and Malignant Peritoneal Effusion

[dist_mets_peritoneum_and_malignant_peritoneal_effusion]

Value (Numeric)

- 0 = No distant metastasis, peritoneum and malignant peritoneal effusion
- 1 = Distant metastasis, peritoneum and malignant peritoneal effusion

Description

- Indicates the presence of distant metastasis in the peritoneum post-diagnosis according to a radiology or pathology report (Appendix Table 4)
- For stage I-III patients, indicates the presence of distant metastasis in the peritoneum post-diagnosis
- For stage IV patients, indicates the presence of distant metastasis in the peritoneum at the time of diagnosis or post-diagnosis
- Populated only if cancer diagnosis is the patient's first BPC Project Cancer

Time (Days, Months, Years) from Diagnosis to Distant Metastasis in Peritoneum and Peritoneal Effusion

[dx_to_dist_mets_peritoneum_and_malignant_peritoneal_effusion_days],
[dx_to_dist_mets_peritoneum_and_malignant_peritoneal_effusion_mos],
[dx_to_dist_mets_peritoneum_and_malignant_peritoneal_effusion_yrs]

Value (Numeric)

Description

- Interval in days [dx_to_dist_mets_peritoneum_and_malignant_peritoneal_effusion_days]; months [dx_to_dist_mets_peritoneum_and_malignant_peritoneal_effusion_mos]; or years [dx_to_dist_mets_peritoneum_and_malignant_peritoneal_effusion_yrs] from diagnosis to first distant metastasis in the peritoneum
- For stage IV patients with distant metastasis in the peritoneum at diagnosis, time in days, months, and years will be 0
- Populated only among patients with distant metastasis in the peritoneum ([dist_mets_peritoneum_and_malignant_peritoneal_effusion] = 1) and if cancer diagnosis is the patient's first BPC Project Cancer

- Note that this variable cannot be used as a time-to-event endpoint as it does not account for death or censoring

Distant Metastasis: Pleura and Malignant Pleural Effusion

[dist_mets_pleura_and_malignant_pleural_effusion]

Value (Numeric)

- 0 = No distant metastasis, pleura and malignant pleural effusion
- 1 = Distant metastasis, pleura and malignant pleural effusion

Description

- Indicates the presence of distant metastasis in the pleura post-diagnosis according to a radiology or pathology report (Appendix Table 4)
- For stage I-III patients, indicates the presence of distantmetastasis in the pleura post-diagnosis
- For stage IV patients, indicates the presence of distant metastasis in the pleura at the time of diagnosis or post-diagnosis
- Populated only if cancer diagnosis is the patient's first BPC Project Cancer

Time (Days, Months, Years) from Diagnosis to Pleura and Malignant Pleural Effusion Distant Metastasis

[dx_to_dist_mets_pleura_and_malignant_pleural_effusion_days],
[dx_to_dist_mets_pleura_and_malignant_pleural_effusion_mos],
[dx_to_dist_mets_pleura_and_malignant_pleural_effusion_yrs]

Value (Numeric)

Description

- Interval in days [dx_to_dist_mets_pleura_and_malignant_pleural_effusion_days]; months [dx_to_dist_mets_pleura_and_malignant_pleural_effusion_mos]; or years [dx_to_dist_mets_pleura_and_malignant_pleural_effusion_yrs] from diagnosis to first distant metastasis in the pleura
- For stage IV patients with distant metastasis in the pleura at diagnosis, time in days, months, and years will be 0
- Populated only among patients with distant metastasis in the pleura ([dist_mets_pleura_and_malignant_pleural_effusion] = 1) and if cancer diagnosis is the patient's first BPC Project Cancer
- Note that this variable cannot be used as a time-to-event endpoint as it does not account for death or censoring

Distant Metastasis: Pulmonary

[dist_mets_pulmonary]

Value (Numeric)

- 0 = No distant metastasis, pulmonary
- 1 = Distant metastasis, pulmonary

Description

- Indicates the presence of pulmonary distant metastasis post-diagnosis according to a radiology or pathology report (Appendix Table 4)
- For stage I-III patients, indicates the presence of pulmonary distant metastasis post-diagnosis
- For stage IV patients, indicates the presence of pulmonary distant metastasis at the time of diagnosis or post-diagnosis
- Populated only if cancer diagnosis is the patient's first BPC Project Cancer

Time (Days, Months, Years) from Diagnosis to Pulmonary Distant Metastasis

[dx_to_dist_mets_pulmonary_days], [dx_to_dist_mets_pulmonary_mos], [dx_to_dist_mets_pulmonary_yrs]

Value (Numeric)

Description

- Interval in days [dx_to_dist_mets_pulmonary_days]; months [dx_to_dist_mets_pulmonary_mos]; or years [dx_to_dist_mets_pulmonary_yrs] from diagnosis to first distant pulmonary metastasis
- For stage IV patients with distant pulmonary metastasis at diagnosis, time in days, months, and years will be 0
- Populated only among patients with distant pulmonary metastasis ([dist_mets_pulmonary] = 1) and if cancer diagnosis is the patient's first BPC Project Cancer
- Note that this variable cannot be used as a time-to-event endpoint as it does not account for death or censoring

Distant Metastasis: Skin

[dist_mets_skin]

Value (Numeric)

- 0 = No distant metastasis, skin
- 1 = Distant metastasis, skin

Description

- Indicates the presence of distant metastasis on the skin post-diagnosis according to a radiology or pathology report (Appendix Table 4)
- For stage I-III patients, indicates the presence of distant metastasis on the skin post-diagnosis
- For stage IV patients, indicates the presence of distant metastasis on the skin at the time of diagnosis or post-diagnosis
- Populated only if cancer diagnosis is the patient's first BPC Project Cancer

Time (Days, Months, Years) from Diagnosis to Distant Metastasis in Skin

[dx_to_dist_mets_skin_days], [dx_to_dist_mets_skin_mos], [dx_to_dist_mets_skin_yrs]

Value (Numeric)

Description

- Interval in days [dx_to_dist_mets_skin_days]; months [dx_to_dist_mets_skin_mos]; or years [dx_to_dist_skin_yrs] from diagnosis to first distant metastasis on the skin
- For stage IV patients with distant metastasis on the skin at diagnosis, time in days, months, and years will be 0
- Populated only among patients with distant metastasis on the skin ([dist_mets_skin] = 1) and if cancer diagnosis is the patient's first BPC Project Cancer
- Note that this variable cannot be used as a time-to-event endpoint as it does not account for death or censoring

Distant Metastasis: Thorax

[dist_mets_thorax]

Value (Numeric)

- 0 = No distant metastasis, thorax
- 1 = Distant metastasis, thorax

Description

- Indicates the presence of distant thoracic metastasis post-diagnosis according to a radiology or pathology report (Appendix Table 4)
- For stage I-III patients, indicates the presence of distant thoracic metastasis post-diagnosis
- For stage IV patients, indicates the presence of distant thoracic metastasis at the time of diagnosis or post-diagnosis
- Populated only if cancer diagnosis is the patient's first BPC Project Cancer

Time (Days, Months, Years) from Diagnosis to Distant Metastasis in Thorax

[dx_to_dist_mets_thorax_days], [dx_to_dist_mets_thorax_mos], [dx_to_dist_mets_thorax_yrs]

Value (Numeric)

Description

- Interval in days [dx_to_dist_mets_thorax_days]; months [dx_to_dist_mets_thorax_mos]; or years [dx_to_dist_mets_thorax_yrs] from diagnosis to first distant thoracic metastasis
- For stage IV patients with distant thoracic metastasis at diagnosis, time in days, months, and years will be 0
- Populated only among patients with distant thoracic metastasis ([dist_mets_thorax] = 1) and if cancer diagnosis is the patient's first BPC Project Cancer
- Note that this variable cannot be used as a time-to-event endpoint as it does not account for death or censoring

Patient Received a Cancer-Directed Regimen Prior to Distant Metastasis

[reg_rcvd_before_distant_mets]

Value (Character)

- No
- Yes

Description

- Indicates whether the patient received a cancer-directed regimen prior to distant metastasis
- Populated only if diagnosed with distant metastasis ([dmets_post_dx] = 1)

Overall Survival from Diagnosis: Status Indicator

[os_dx_status]

Value (Numeric)

- 1 = Dead
- 0 = Censored

Description

- An event is defined by death
- Patients were censored if not known to be dead

Overall Survival from Diagnosis (Days, Months, Years)

[tt_os_dx_days], [tt_os_dx_mos], [tt_os_dx_yrs]

Value (Numeric)

Description

- Time from diagnosis to death or the date last known alive
- Interval in days [tt_os_dx_days]; months [tt_os_dx_mos]; or years [tt_os_dx_yrs] from cancer diagnosis to last known alive date or death

Overall Survival from Advanced Disease: Status Indicator

[os_adv_status]

Value (Numeric)

- 1 = Dead
- 0 = Censored

Description

- An event is defined by death
- Patients were censored if not known to be dead
- Populated only if:
  - BPC Project Cancer
  - Populated only for BPC Project Cancers with distant metastasis. This includes patients diagnosed with stage IV disease with distant metastasis present at diagnosis ([stage_dx] = “Stage IV” and [ca_dmets_yn] = “Yes”) and patients who developed distant metastasis post-diagnosis ([dmets_post_dx] = 1).

Overall Survival from Advanced Disease (Days, Months, Years)

[tt_os_adv_days], [tt_os_adv_mos], [tt_os_adv_yrs]

Value (Numeric)

Description

- Time from advanced disease to death or the date last known alive
- Interval in days [tt_os_adv_days]; months [tt_os_adv_mos]; or years [tt_os_adv_yrs] from date of advanced disease to last known alive date or death
- Populated only for BPC Project Cancers with distant metastasis. This includes patients diagnosed with stage IV disease with distant metastasis present at diagnosis ([stage_dx] = “Stage IV” and [ca_dmets_yn] = “Yes”) and patients who developed distant metastasis post-diagnosis ([dmets_post_dx] = 1).

Progression-Free Survival Cohort

[pfs_cohort]

Value (Character)

- Stage I-III with Distant Mets
- Stage IV

Description

- Indicates whether PFS was defined from first date of distant metastasis (stage I-III) or from diagnosis (stage IV)
- Populated only if BPC Project Cancer

Progression Free Survival-Imaging (PFS-I): Status Indicator

[pfs_i_adv_status]

Value (Numeric)

- 1 = Progression according to radiologist assessment; death
- 0 = Censored

Description

- An event is defined by:
    - Radiologist assessment of the change in the patient's cancer status as progressing/worsening/enlarging
    - Death
- Patients were censored in the absence of an event
- Populated only for BPC Project Cancers with distant metastasis. This includes patients diagnosed with stage IV disease with distant metastasis present at diagnosis ([stage_dx] = “Stage IV” and [ca_dmets_yn] = “Yes”) and patients who developed distant metastasis post-diagnosis ([dmets_post_dx] = 1).

Progression Free Survival-Imaging (PFS-I)

[tt_pfs_i_adv_days], [tt_pfs_i_adv_mos], [tt_pfs_i_adv_yrs]

Value (Numeric)

Description

- Interval in days [tt_pfs_i_adv_days]; months [tt_pfs_i_adv_mos]; or years [tt_pfs_i_adv_yrs] from stage IV diagnosis or date of distant metastasis (stage I-III at diagnosis) to:
    - Radiologist assessment of progressing/worsening/enlarging

  - Death
  - Last known alive date, if patient is censored according to PFS-I status indicator [pfs_i_adv_status]
- Populated only for BPC Project Cancers with distant metastasis. This includes patients diagnosed with stage IV disease with distant metastasis present at diagnosis ([stage_dx] = "Stage IV" and [ca_dmets_yn] = "Yes") and patients who developed distant metastasis post-diagnosis ([dmets_post_dx] = 1).

Progression Free Survival - Medical Oncologist Assessment (PFS-M): Status Indicator

[pfs_m_adv_status]

Value (Numeric)

- 1 = Progression according to medical oncologist assessment; death
- 0 = Censored

Description

- An event is defined by:
  - Medical oncologist assessment of the change in the patient's cancer status as progressing/worsening/enlarging
  - Death
- Patients were censored in the absence of an event
- Populated only for BPC Project Cancers with distant metastasis. This includes patients diagnosed with stage IV disease with distant metastasis present at diagnosis ([stage_dx] = "Stage IV" and [ca_dmets_yn] = "Yes") and patients who developed distant metastasis post-diagnosis ([dmets_post_dx] = 1).

Progression Free Survival-Medical Oncologist (PFS-M)

[tt_pfs_m_adv_days], [tt_pfs_m_adv_mos], [tt_pfs_m_adv_yrs]

Value (Numeric)

Description

- Interval in days [tt_pfs_m_adv_days]; months [tt_pfs_m_adv_mos]; or years [tt_pfs_m_adv_yrs] from stage IV diagnosis or date of distant metastasis (stage I-III at diagnosis) to:
  - Medical oncologist assessment of progressing/worsening/enlarging
  - Death
  - Last known alive date, if patient is censored according to PSF-M status indicator [pfs_m_adv _status]
- Populated only for BPC Project Cancers with distant metastasis. This includes patients

diagnosed with stage IV disease with distant metastasis present at diagnosis ([stage_dx] = "Stage IV" and [ca_dmets_yn] = "Yes") and patients who developed distant metastasis post-diagnosis ([dmets_post_dx] = 1).

Progression Free Survival-Imaging or Medical Oncologist Assessment (PFS-I-or-M): Status Indicator

[pfs_i_or_m_adv_status]

Value (Numeric)

- 1 = Progression according to radiologist assessment or medical oncologist assessment; death
- 0 = Censored

Description

- An event is defined by:
  - Radiologist assessment of the change in the patient's cancer status as progressing/worsening/enlarging
  - Medical oncologist assessment of the change in the patient's cancer status as progressing/worsening/enlarging
  - Death
- Patients were censored in the absence of an event
- Populated only for BPC Project Cancers with distant metastasis. This includes patients diagnosed with stage IV disease with distant metastasis present at diagnosis ([stage_dx] = "Stage IV" and [ca_dmets_yn] = "Yes") and patients who developed distant metastasis post-diagnosis ([dmets_post_dx] = 1).

Progression Free Survival-Imaging or Medical Oncologist Assessment (PFS-I-or-M)

[tt_pfs_i_or_m_adv_days], [tt_pfs_i_or_m_adv_mos], [tt_pfs_i_or_m_adv_yrs]

Value (Numeric)

Description

- Time from diagnosis to radiologist impression of progressing/worsening/enlarging cancer status, last known alive date, or death
- Interval in days [tt_pfs_i_or_m_adv_days]; months [tt_pfs_i_or_m_adv_mos]; or years [tt_pfs_i_or_m_adv_yrs] from stage IV diagnosis or date of distant metastasis (stage I-III at diagnosis) to:
  - The first of a radiologist assessment of progressing/worsening/enlarging, medical oncologist assessment of progressing/worsening/enlarging, or death
  - Last known alive date, if patient is censored according to PFS-I-or-M status indicator [pfs_i_or_m_adv_status]

- Populated only for BPC Project Cancers with distant metastasis. This includes patients diagnosed with stage IV disease with distant metastasis present at diagnosis ([stage_dx] = "Stage IV" and [ca_dmets_yn] = "Yes") and patients who developed distant metastasis post-diagnosis ([dmets_post_dx] = 1).

Progression Free Survival-Imaging and Medical Oncologist Assessment (PFS-I-and-M): Status Indicator

[pfs_i_and_m_adv_status]

Value (Numeric)

- 1 = Progression according to radiologist assessment and medical oncologist assessment; death
- 0 = Censored

Description

- An event is defined by:
  - Radiologist assessment and medical oncologist assessment of the change in the patient's cancer status as progressing/worsening/enlarging
  - Death
- Patients were censored in the absence of an event
- Populated only for BPC Project Cancers with distant metastasis. This includes patients diagnosed with stage IV disease with distant metastasis present at diagnosis ([stage_dx] = "Stage IV" and [ca_dmets_yn] = "Yes") and patients who developed distant metastasis post-diagnosis ([dmets_post_dx] = 1).

Progression Free Survival-Imaging and Medical Oncologist Assessment (PFS-I-and-M)

[tt_pfs_i_and_m_adv_days], [tt_pfs_i_and_m_adv_mos], [tt_pfs_i_and_m_adv_yrs]

Value (Numeric)

Description

- Time from diagnosis to radiologist assessment or medical oncologist assessment of progressing/worsening/enlarging cancer status, last known alive date or death
- Interval in days [tt_pfs_i_and_m_adv_days]; months [tt_pfs_i_and_m_adv_mos]; or years [tt_pfs_i_and_m_adv_yrs] from stage IV diagnosis or date of distant metastasis (stage I-III at diagnosis) to:
  - The latter of radiologist assessment or medical oncologist assessment documenting progressing disease, if the patient progressed according to both sources
  - Death
  - Last known alive date, if patient is censored according to PFS-I-and-M status indicator [pfs_i_and_m_adv _status]

- Populated only for BPC Project Cancers with distant metastasis. This includes patients diagnosed with stage IV disease with distant metastasis present at diagnosis ([stage_dx] = “Stage IV” and [ca_dmets_yn] = “Yes”) and patients who developed distant metastasis post-diagnosis ([dmets_post_dx] = 1).

Release Version Number

[release_version]

Value (Character)

- 1.0-public

Description

- Indicates the version number of the data release

CANCER-DIRECTED REGIMEN DATASET

The Cancer-Directed Regimen dataset is structured as one record per regimen-associated cancer diagnosis, per patient. For example, if a regimen is associated with a single cancer diagnosis, there will be one corresponding record in this dataset. If a regimen is associated with two cancer diagnoses, then there will be two corresponding records in this dataset: one for the first associated cancer diagnosis and another for the second associated cancer diagnosis. If it is unknown which cancer diagnosis the regimen is associated with, there will still be one record in this dataset.

Cancer-directed regimens were curated for all cancer diagnoses, including both BPC Project and non-BPC Project Cancers. A regimen can only consist of one drug or up to five drugs given together. Cancer-directed drugs include anti-neoplastic drugs, immunotherapies, targeted therapies, and hormone therapies. A break in treatment of ≥8 weeks was used to indicate the end of a regimen; even if all drugs in the regimen were re-initiated 8+ weeks later, this was considered a new regimen. All LHRH agonists ([regimen_drugs] containing "Goserlin Acetate", "Histrelin Acetate", "Leuprolide Acetate", "Triptorelin") were curated as separate regimens even if administered with additional drugs.

Compliance with data privacy requires redaction of the name and duration of investigational drugs. If the cancer-directed drug was part of an investigational drug trial, the drug name(s) [drugs_drug_1-drugs_drug_5] were set to "Investigational Drug" and the end date interval will match the start date interval. Identification of investigational drugs varies by institution depending on contractual obligations. Each institution followed their respective data privacy specifications for investigational drug masking.

This dataset can be linked to the following datasets:

- *BPC Project Cancer Diagnosis, Non-BPC Project Cancer Diagnosis, Cancer-Directed Radiation Therapy, and Cancer Panel Test datasets using the variables [cohort], [record_id], and [ca_seq].*
- *Patient Characteristics, PRISSMM Pathology, PRISSMM Imaging, PRISSMM Medical Oncologist Assessment, and PRISSMM Tumor Marker datasets using the variables [cohort] and [record_id].*

Field names shaded in gray indicate that an alternative variable is preferred for analysis. The recommended variable is noted in the description.

BPC Project Cohort

[cohort]

Value (Character)

- BrCa

Description

- Indicates the BPC Project Cancer type
- Primary key for merging across all datasets

Record ID

[record_id]

Value (Character)

- GENIE-[INSTITUTION]-XXXXXX

Description

- De-identified, unique patient ID
- Conforms to the following the convention: GENIE-[INSTITUTION]-XXXX. The first component is the string, “GENIE”; the second component is the institution’s abbreviation; the third component is a unique ID for the patient.
- Primary key for the AACR Project GENIE genomic datasets [PATIENT_ID].

Institution

[institution]

Value (Character)

- DFCI = Dana Farber Cancer Institute
- MSK = Memorial Sloan Kettering Cancer Center
- VICC = Vanderbilt Ingram Cancer Center

Description

- Indicates the patient’s internal institution of cancer care
- Corresponds to variable [center] in AACR Project GENIE data

Cancer Sequence Associated with Regimen

[ca_seq]

Value (Numeric)

- 0 = first and only cancer
- 1 = first of two or more primaries
- 2 = second of two or more primaries
- …10 = tenth of ten or more primaries

Description

- The cancer sequence number associated with this regimen
- When regimens are associated with multiple cancer diagnoses, each associated cancer diagnosis will be a separate row in the dataset
- Populated only if it is known which diagnosis is associated with this cancer-directed regimen
- Primary key for the Cancer Diagnosis, Cancer-Directed Regimen and Cancer Panel Test datasets

Regimen Number

[regimen_number]

Value (Numeric)

- 1 = first regimen
- 2 = second regimen
- …n = nth regimen

Description

- Order for cancer-directed regimens based on the start date of the first cancer-directed drug
- The cancer-directed regimen dataset is structured as one row per regimen-associated cancer diagnosis, per patient. For example, if a regimen is associated with a single cancer diagnosis, there will be one corresponding record in this dataset. If a regimen is associated with two cancer diagnoses, then there will be two corresponding records in this dataset: one for the first associated cancer diagnosis and another for the second associated cancer diagnosis. If it is unknown which diagnosis the regimen is associated with, there will be one row in the cancer-directed regimen dataset.
  - This implies that multiple records will have the same regimen number in this dataset when they are associated with multiple cancer diagnoses (e.g., if the first cancer-directed regimen that a patient received is associated with their first and second cancer diagnosis then there will be two rows in this dataset, one for the first cancer diagnosis and one for the second cancer diagnosis, and both will have [regimen_number] = 1).

Regimen Number Within Cancer Diagnosis

[regimen_number_within_cancer]

Value (Numeric)

- 1 = first regimen associated with specific cancer diagnosis
- 2 = second regimen associated with specific cancer diagnosis
- …n = nth regimen associated with specific cancer diagnosis

Description

- Order of cancer-directed regimens that are associated with a specific cancer diagnosis [ca_seq] based on the start date of the first cancer-directed drug

BPC Project (Index) Cancer Indicator

[redcap_ca_index]

Value (Character)

- Yes
- No

Description

- Indicates whether the regimen was associated with a BPC Project Cancer
- The BPC Project Cancer is defined as the cancer that met eligibility criteria, underwent genomic sequencing and was submitted to AACR Project GENIE.
    - Each patient has at least one BPC Project Cancer. Patients may have multiple BPC Project Cancers, though this is rare.
    - PRISSMM data elements are curated for BPC Project Cancers.
    - The terms “BPC Project Cancer” and “index cancer” are used interchangeably.
    - Further details regarding the definition of BPC Project and non-BPC Project Cancers can be found in Appendix 1. BPC Project and Non-BPC Project Cancers.

Number of Cancer-Directed Drugs in a Regimen

[drugs_num]

Value (Numeric)

- 1-5

Description

- Number of cancer-directed drugs in a regimen; up to 5 recorded

Institution That Administered/Ordered Regimen

[drugs_inst]

Value (Character)

- At the internal/native institution only
- Split across internal and external institution
- At external institution only

Description

- Location where regimen was administered/ordered
- Indicates whether regimen was administered/ordered at the internal (i.e. the same institution as the [institution] variable) or external institution.

Institution That First Ordered the Regimen

[drugs_firstinst]

Value (Character)

- Internal institution
- External institution

Description

- Location where regimen was first ordered
- Indicates whether regimen was administered/ordered at the internal (i.e. the same institution as the [institution] variable) or external institution.
- Populated only if the regimen was administered/ordered across internal and external institutions ([drugs_inst] = "Split across internal and external institution")

Regimen Was Part of a Clinical Trial

[drugs_ct_yn]

Value (Character)

- Yes
- No

Description

- Indicator for whether a cancer-directed drug was part of a clinical trial
- If the cancer-directed drug was part of an investigational drug trial, the drug name(s) [drugs_drug_1-drugs_drug_5] will be masked as "Investigational drug" and the end date interval will match the start date interval.
  - Identification of investigational drugs varies by institution depending on contractual obligations. For some institutions, all drugs that are part of an investigational trial are required to be masked, even if standard of care. For other institutions, only the investigational drug(s) are required to be masked.

Regimen Discontinuation Status

[drugs_dc_ynu]

Value (Character)

- Yes
- No
- Unknown, no documentation found

Description

- Indicator for whether the drug regimen was discontinued. If not discontinued, the patient was still receiving cancer-directed regimen at time of curation, or it is unknown whether the regimen had ended
- The response to this variable specifies whether there will be a cancer-directed regimen end date or a date of last administration specified in the variables [dx_drug_end_or_lastadm_int_1 – dx_drug_end_or_lastadm_int_5]

Names of Drugs in Regimen

[regimen_drugs]

Value (Character)

Description

- Names of cancer-directed drugs received together
- Concatenation of variables [drugs_drug_1]-[drugs_drug_5]

Name of Cancer-Directed Drug in Regimen, Drugs 1-5

[drugs_drug_1] – [drugs_drug_5]

Value (Character)

- Name of cancer-directed drug 1 through 5 in each regimen (Appendix 2)

Description

- The cancer drug label contains the generic/ingredient name with the synonyms in parentheses (e.g. Nivolumab (BMS936558, MDX1106, NIVO, ONO4538, Opdivo))
- If the cancer-directed drug was part of an investigational drug trial, the drug name(s) [drugs_drug_1-drugs_drug_5] will be masked as "Investigational Drug" and the end date interval will match the start date interval.
    - Identification of investigational drugs varies by institution depending on contractual obligations. For some institutions, all drugs that are part of an investigational trial are required to be masked, even if standard of care. For other institutions, only the investigational drug(s) are required to be masked.

Data Standard: National Cancer Institute Thesaurus: Antineoplastic Agents

Time (Days) from Date of Birth to Start of Cancer-Directed Drug in Regimen, Drugs 1-5

[drugs_startdt_int_1] – [drugs_startdt_int_5]

Value (Numeric)

Description

- Interval in days from date of birth to start of each cancer-directed drug 1 – drug 5

Time (Days, Months) from Associated Cancer Diagnosis to Start of Cancer-Directed Drug in Regimen, Drugs 1-5

[dx_drug_start_int_1] – [dx_drug_start_int_5], [dx_drug_start_int_mos_1] – [dx_drug_start_int_mos_5]

Value (Numeric)

Description

- Interval in days [dx_drug_start_int_1] – [dx_drug_start_int_5] or months [dx_drug_start_int_mos_1] – [dx_drug_start_int_mos_5] from cancer diagnosis to start of cancer-directed drug 1 – drug 5

Time (Days, Months) from Associated Cancer Diagnosis to End of Cancer-Directed Drug in Regimen, Drugs 1-5

[dx_drug_end_int_1] – [dx_drug_end_int_5], [dx_drug_end_int_mos_1] – [dx_drug_end_int_mos_5]

Value (Numeric)

Description

- Number of days ([dx_drug_end_int_1-dx_drug_end_int_5]) or months ([dx_drug_end_int_mos_1-dx_drug_end_int_mos5]) from cancer diagnosis to end of cancer-directed drug 1 – drug 5
- If the cancer-directed drug was part of an investigational drug trial, the drug name(s) [drugs_drug_1-drugs_drug_5] will be set to “Investigational drug” and the end date interval will match the start date interval.
    - Identification of investigational drugs varies by institution depending on contractual obligations. For some institutions, all drugs that are part of an investigational trial are required to be masked, even if standard of care. For other institutions, only the investigational drug(s) are required to be masked.
- Populated only if drug regimen was discontinued ([drugs_dc_ynu]="Yes")

Time (Days) from Associated Cancer Diagnosis to End Date or Last Known Administration Date of Drugs 1-5

[dx_drug_end_or_lastadm_int_1]-[dx_drug_end_or_lastadm_int_5]

Value (Numeric)

Description

- Number of days from associated cancer diagnosis to the end of cancer-directed drug 1 – drug 5 if regimen was discontinued ([drugs_dc_ynu] = "Yes") or to last known administration date of cancer-directed drug 1 – drug 5 if regimen was not discontinued ([drugs_dc_ynu] = "No" or "Unknown")

Time from Associated Cancer Diagnosis to Start of Cancer-Directed Regimen

[dx_reg_start_int], [dx_reg_start_int_mos], [dx_reg_start_int_yrs]

Value (Numeric)

Description

- Interval in days [dx_reg_start_int]; months [dx_reg_start_int_mos]; or years [dx_reg_start_int_yrs] from associated cancer diagnosis to the start of the first drug in cancer-directed regimen

Time (Days, Months, Years) from Associated Cancer Diagnosis to End of First Drug Discontinued in Cancer-Directed Regimen

[dx_reg_end_any_int], [dx_reg_end_any_int_mos], [dx_reg_end_any_int_yrs]

Value (Numeric)

Description

- Number of days [dx_reg_end_any_int]; months [dx_reg_end_any_int_mos]; or years [dx_reg_end_any_int_yrs] from associated cancer diagnosis to the end of first drug discontinued in cancer-directed regimen
- Populated only if regimen is known to be discontinued ([drugs_dc_ynu] = “Yes”)

Time (Days, Months, Years) from Associated Cancer Diagnosis to End of All Drugs in Cancer-Directed Regimen

[dx_reg_end_all_int], [dx_reg_end_all_int_mos], [dx_reg_end_all_int_yrs]

Value (Numeric)

Description

- Number of days [dx_reg_end_all_int]; months [dx_reg_end_all_int_mos]; or years [dx_reg_end_all_int_yrs] from associated cancer diagnosis to the end of all drugs in the cancer-directed regimen
- Populated only if regimen is known to be discontinued ([drugs_dc_ynu] = "Yes")

Overall Survival from Start of Cancer-Directed Drug: Status Indicator

[os_d_status]

Value (Numeric)

- 1 = Dead
- 0 = Censored

Description

- An event is defined by death
- Patients were censored if not known to be dead

Overall Survival from Start of Cancer-Directed Drug (Days, Months, Years)

[tt_os_d1_days] – [tt_os_d5_days], [tt_os_d1_mos] – [tt_os_d5_mos], [tt_os_d1_yrs] – [tt_os_d5_yrs]

Value (Numeric)

Description

- Time from start of cancer-directed drug to death/last known alive date
- Interval in days [tt_os_d1_days- tt_os_d5_days]; months [tt_os_d1_mos- tt_os_d5_mos]; or years [tt_os_d1_yrs- tt_os_d5_yrs] from start of cancer-directed drug [drugs_startdt_int_drug_1-drugs_startdt_int_drug_5] to death [hybrid_death_int] or last known alive date [dob_lastalive_int]

Overall Survival from Start of Cancer-Directed Regimen: Status Indicator

[os_g_status]

Value (Numeric)

- 1 = Dead
- 0 = Censored

Description

- An event is defined by death
- Patients were censored if not known to be dead

Overall Survival from Start of Cancer-Directed Regimen (Days, Months, Years)

[tt_os_g_days], [tt_os_g_mos], [tt_os_g_yrs]

Value (Numeric)

Description

- Time from start of cancer-directed regimen to death/last known alive date
- Interval in days [tt_os_g_days]; months [tt_os_g_mos]; or years [tt_os_g_yrs] from the start of the cancer-directed regimen to death [hybrid_death_int] or last known alive date [dob_lastalive_int]

Progression Free Survival – Imaging (PFS-I) from Start of Cancer-Directed Regimen: Status Indicator

[pfs_i_g_status]

Value (Numeric)

- 1 = Progression or death
- 0 = Censored

Description

- An event is defined by:
    - Radiologist assessment of the change in the patient's cancer status as progressing/worsening/enlarging
    - Death
- Patients were censored at the time of next cancer-directed regimen or end of follow up.
    - Since LHRH agonists were recorded as separate regimens, the next cancer-directed regimen may begin before the LHRH agonist treatment ends.
- Populated only for regimens associated with BPC Project Cancers with stage IV disease at diagnosis or stage I-III disease at diagnosis with subsequent distant metastasis [dmets_post_dx = 1].

Progression Free Survival – Imaging (PFS-I) from Start of Cancer-Directed Regimen

[tt_pfs_i_g_days], [tt_pfs_i_g_mos], [tt_pfs_i_g_yrs]

Value (Numeric)

Description

- Interval in days [tt_pfs_i_g_days]; months [tt_pfs_i_g_mos]; or years [tt_pfs_i_g_yrs] from start of cancer-directed regimen to:
    - Radiologist assessment of progressing/worsening/enlarging
    - Death
    - Start of next regimen (if applicable) or last known alive date, if patient is censored according to PFS-I status indicator [pfs_i_g_status]
- Populated only for regimens associated with BPC Project Cancers with stage IV disease at diagnosis or stage I-III disease at diagnosis with subsequent distant metastasis ([dmets_post_dx] = 1). For patients with stage I-III disease at diagnosis with subsequent distant metastasis, PFS is only defined for regimens that began after date of distant metastasis

Progression Free Survival – Medical Oncologist Assessment (PFS-M) from Start of Cancer-Directed Regimen: Status Indicator

[pfs_m_g_status]

Value (Numeric)

- 1 = Progression according to medical oncologist assessment; death
- 0 = Censored

Description

- An event is defined by:
    - Medical oncologist assessment of the change in the patient's cancer status as progressing/worsening/enlarging
    - Death
- Patients were censored at the time of next cancer-directed regimen or end of follow up.
    - Since LHRH agonists were recorded as separate regimens, the next cancer-directed regimen may begin before the LHRH agonist treatment ends.
- Populated only for regimens associated with BPC Project Cancers with stage IV disease at diagnosis or stage I-III disease at diagnosis with subsequent distant metastasis ([dmets_post_dx] = 1). For patients with stage I-III disease at diagnosis with subsequent distant metastasis, PFS is only defined for regimens that began after date of distant metastasis

Progression Free Survival – Medical Oncologist Assessment (PFS-M) from Start of Cancer-Directed Regimen

[tt_pfs_m_g_days], [tt_pfs_m_g_mos], [tt_pfs_m_g_yrs]

Value (Numeric)

Description

- Interval in days [tt_pfs_m_g_days]; months [tt_pfs_m_g_mos]; or years [tt_pfs_m_g_yrs] from start of cancer-directed regimen to:
    - Medical oncologist assessment of mixed or progressing disease
    - Death
    - Start of next regimen (if applicable) or last known alive date, if patient is censored according to PFS-M status indicator [pfs_m_g_status]
- Populated only for regimens associated with BPC Project Cancers with stage IV disease at diagnosis or stage I-III disease at diagnosis with subsequent distant metastasis ([dmets_post_dx] = 1). For patients with stage I-III disease at diagnosis with subsequent distant metastasis, PFS is only defined for regimens that began after date of distant metastasis.

Progression Free Survival – Imaging or Medical Oncologist Assessment (PFS-I-or-M) from Start of Cancer-Directed Regimen: Status Indicator

[pfs_i_or_m_g_status]

Value (Numeric)

- 1 = Progression according to radiologist assessment or medical oncologist assessment; death
- 0 = Censored

Description

- An event is defined by:
    - Radiologist assessment of the change in the patient's cancer status as progressing/worsening/enlarging
    - Medical oncologist assessment of the change in the patient's cancer status as progressing/worsening/enlarging
    - Death
- Patients were censored at the time of next cancer-directed regimen or end of follow up.
    - Since LHRH agonists were recorded as separate regimens, the next cancer-directed regimen may begin before the LHRH agonist treatment ends.
- Populated only for cancer-directed regimens associated with BPC Project Cancers with stage IV disease at diagnosis or stage I-III disease at diagnosis with subsequent distant metastasis [dmets_post_dx = 1].

Progression Free Survival-Imaging or Medical Oncologist Assessment (PFS-I-or-M)

[tt_pfs_i_or_m_g_days], [tt_pfs_i_or_m_g_mos], [tt_pfs_i_or_m_g_yrs]

Value (Numeric)

Description

- Interval in days [tt_pfs_i_or_m_g_days]; months [tt_pfs_i_or_m_g_mos]; or years [tt_pfs_i_or_m_g_yrs] from start of cancer-directed regimen to:
    - The first of a radiologist assessment of progressing/worsening/enlarging, medical oncologist assessment of progressing/worsening/enlarging, or death
    - Start of next regimen (if applicable) or last known alive date, if patient is censored according to PFS-I-or-M status indicator [pfs_i_or_m_g_status]
- Populated only for regimens associated with BPC Project Cancers with stage IV disease at diagnosis or stage I-III disease at diagnosis with subsequent distant metastasis ([dmets_post_dx] = 1). For patients with stage I-III disease at diagnosis with subsequent distant metastasis, PFS is only defined for regimens that began after date of distant metastasis.

Progression Free Survival – Imaging and Medical Oncologist Assessment (PFS-I-and-M) from Start of Cancer-Directed Regimen: Status Indicator

[pfs_i_and_m_g_status]

Value (Numeric)

- 1 = Progression according to radiologist assessment and medical oncologist assessment; death
- 0 = Censored

Description

- An event is defined by:
    - Radiologist assessment and medical oncologist assessment of the change in the patient's cancer status as progressing/worsening/enlarging
    - Death
- Patients were censored at the time of next cancer-directed regimen or end of follow up.
    - Since LHRH agonists were recorded as separate regimens, the next cancer-directed regimen may begin before the LHRH agonist treatment ends.
- Populated only for cancer-directed regimens associated with BPC Project Cancers with stage IV disease at diagnosis or stage I-III disease at diagnosis with subsequent distant metastasis [dmets_post_dx = 1].

Progression Free Survival-Imaging and Medical Oncologist Assessment (PFS-I-and-M)

[tt_pfs_i_and_m_g_days], [tt_pfs_i_and_m_g_mos], [tt_pfs_i_and_m_g_yrs]

Value (Numeric)

Description

- Interval in days [tt_pfs_i_and_m_g_days]; months [tt_pfs_i_and_m_g_mos]; or years [tt_pfs_i_and_m_g_yrs] from start of cancer-directed regimen to:
    - The latter of a radiologist assessment of progressing/worsening/enlarging, medical oncologist assessment of progressing/worsening/enlarging, or death
    - Start of next regimen (if applicable) or last known alive date, if patient is censored according to PFS-I-and-M status indicator [pfs_i_and_m_g_status]
- Populated only for regimens associated with BPC Project Cancers with stage IV disease at diagnosis or stage I-III disease at diagnosis with subsequent distant metastasis ([dmets_post_dx] = 1). For patients with stage I-III disease at diagnosis with subsequent distant metastasis, PFS is only defined for regimens that began after date of distant metastasis.

Time (Days, Months, Years) to Next Treatment For Any Cancer: Status Indicator

[ttnt_any_ca_status]

Value (Numeric)

- 1 = Initiation of a subsequent cancer-directed regimen for any cancer; death
- 0 = Censored

Description

- An event is defined by:
    - Initiation of a subsequent cancer-directed regimen for any cancer diagnosis
    - Death
- Patients were censored in the absence of an event

Time (Days, Months, Years) to Next Treatment for Any Cancer

[ttnt_any_ca_days], [ttnt_any_ca_mos], [ttnt_any_ca_yrs]

Value (Numeric)

Description

- Time from start of cancer-directed regimen to initiation of subsequent cancer-directed regimen for any cancer or death
- Interval in days [ttnt_any_ca_days]; months [ttnt_any_ca_mos]; or years [ttnt_any_ca_yrs] from start of cancer-directed regimen to:
    - Initiation of subsequent cancer-directed regimen for any cancer

  - Death
  - Last known alive date, if patient is censored according to time to next treatment (any cancer) status indicator [ttnt_any_ca_days]
- Note: TTNT is not defined for LHRH agonists ([regimen_drugs] containing “Goserlin Acetate”, “Histrelin Acetate”, “Leuprolide Acetate”, “Triptorelin”), and they are excluded from TTNT calculations for other regimens

Time to Next Treatment For This Cancer: Status Indicator

[ttnt_ca_seq_status]

Value (Numeric)

- 1 = Initiation of a subsequent cancer-directed regimen for the same cancer; death
- 0 = Censored

Description

- An event is defined by:
  - Initiation of a subsequent cancer-directed regimen for the same cancer diagnosis
  - Death
- Patients were censored in the absence of an event

Time (Days, Months, Years) to Next Treatment for this Cancer

[ttnt_ca_seq_days], [ttnt_ca_seq_mos], [ttnt_ca_seq_yrs]

Value (Numeric)

Description

- Time from start of cancer-directed regimen to initiation of subsequent cancer-directed regimen for the same cancer or death
- Interval in days [ttnt_ca_seq_days]; months [ttnt_ca_seq_mos]; or years [ttnt_ca_seq_yrs] from start of cancer-directed regimen to:
  - Initiation of subsequent cancer-directed regimen for this cancer
  - Death
  - Last known alive date, if patient is censored according to time to next treatment (the same cancer) status indicator [ttnt_ca_seq_days]
- Note: TTNT is not defined for LHRH agonists ([regimen_drugs] containing “Goserlin Acetate”, “Histrelin Acetate”, “Leuprolide Acetate”, “Triptorelin”), and they are excluded from TTNT calculations for other regimens

Release Version Number

[release_version]

Value (Character)

- 1.0-public

Description

- Indicates the version number of the data release

PRISSMM PATHOLOGY DATASET

The Pathology dataset is structured as one record per pathology report, per patient. All pathology reports beginning with the month and year of the first BPC Project Cancer diagnosis are curated. All subsequent pathology reports were recorded (including pathology reports corresponding to non-BPC Project Cancer and subsequent BPC Project Cancer diagnoses; Appendix 1). Additionally, all non-BPC Project BrCa pathology reports are curated.

The PRISSMM Pathology dataset can be linked to the following datasets:

- *Cancer Panel Test dataset using the variables [cohort], [record_id], [path_proc_number], and [path_report_number].*
-
    - *Patient Characteristics, BPC Project and non-BPC Project Cancer Diagnosis, Cancer-Directed Regimen, Cancer-Directed Radiation Therapy, PRISSMM Imaging, PRISSMM Medical Oncologist Assessment, and PRISSMM Tumor Marker Assessment datasets using the variables [cohort] and [record_id].*

Field names shaded in gray indicate that an alternative variable is preferred for analysis. The preferred variable is noted in the description.

BPC Project Cohort

[cohort]

Value (Character)

- BrCa

Description

- Indicates the BPC Project Cancer type
- Primary key for merging across all datasets

Record ID

[record_id]

Value (Character)

- GENIE-[INSTITUTION]-XXXXXX

Description

- De-identified, unique patient ID
- Conforms to the following the convention: GENIE-[INSTITUTION]-XXXX. The first component is the string, “GENIE”; the second component is the institution’s abbreviation; the third component

is a unique ID for the patient.
- Primary key for the AACR Project GENIE genomic datasets [PATIENT_ID].

Institution

[institution]

Value (Character)

- DFCI = Dana Farber Cancer Institute
- MSK = Memorial Sloan Kettering Cancer Center
- VICC = Vanderbilt Ingram Cancer Center

Description

- Indicates the patient's internal institution of cancer care
- Corresponds to variable [center] in AACR Project GENIE data

Pathology Procedure Number

[path_proc_number]

Value (Numeric)

Description

- Order of pathology procedures based on the date of the procedure [path_proc_int]
- Pathology procedures occurring on the same date have the same procedure number
- Primary key for PRISSMM pathology and cancer panel test datasets.

Pathology Report Number

[path_rep_number]

Value (Numeric)

Description

- Order of pathology reports from the same pathology procedure [path_proc_number]
- For example, a pathology procedure with three associated reports will have pathology report numbers 1-3.
- Primary key for PRISSMM pathology and cancer panel test datasets.

Institution Where Procedure Was Performed

[path_proc_inst]

Value (Character)

- Internal institution
- External institution

Description

- Indicates whether the pathology procedure was performed at the internal (i.e. the same institution as the [institution] variable) or external institution.

Institution Where Pathology was Reviewed

[path_rep_inst]

Value (Character)

- Internal institution
- External institution

Description

- Indicates whether the pathology procedure was reviewed at the internal (i.e. the same institution as the [institution] variable) or external institution.

Time (Days) from Date of Birth to Pathology Procedure Date

[path_proc_int]

Value (Numeric)

Description

- Interval in days from date of birth to pathology procedure date

Time (Days, Months, Years) from First BPC Project Cancer to Pathology Procedure Date

[dx_path_proc_days], [dx_path_proc_mos], [dx_path_proc_yrs]

Value (Numeric)

Description

- Interval in days [dx_path_proc_days]; months [dx_path_proc_mos]; or years [dx_path_proc_yrs] from first BPC Project Cancer diagnosis to pathology procedure date

Pathology Type

[path_proc_type]

Value (Character)

- Cytology
- Surgical pathology
- Other

Description

- Type of pathology

Number of Specimens Included in the Pathology Report

[path_num_spec]

Value (Numeric)

- 1-45

Description

- The number of distinct specimens from one procedure that are included in the report
- Up to 45 specimens can be curated from each pathology report

Anatomic Site for Each Specimen 1-45

[path_site1] – [path_site45]

Value (Character)

- ICD-O-3 topography code

Description

- The anatomic site of the specimen is often different than the type of invasive cancer

In Situ Cancer Found in at Least One Specimen in the Pathology Report

[path_insitu_any]

Value (Character)

- Yes
- No

Description

- Indicates whether in situ cancer was found in any of the 45 specimens on the pathology report
- Based on [path_insitu1] – [path_insitu45]

Number of Specimens with In Situ Cancer in Pathology Report

[n_specimen_insitu]

Value (Numeric)

- 0-45

Description

- Number of specimens with in situ cancer on the pathology report
- Based on [path_insitu1] – [path_insitu45]

In Situ Cancer Identified in Specimen 1-45

[path_insitu1] – [path_insitu45]

Value (Character)

- Yes
- No

Description

- Indicates whether any in situ cancer (ICD-O-3 Behavior Code = 2) is present in the specimen

Non-invasive histology in Specimen 1-45

[path_ca_ishist1] – [path_ca_ishist45]

Value (Character)

- 8010 Carcinoma in situ
- 8050 Papillary carcinoma in situ

- 8070 Squamous cell carcinoma in situ NOS
- 8120 Urothelial carcinoma in situ
- 8122 Non-papillary non-invasive high-grade (WHO-grade III)
- 8130 Papillary trans. cell carcinoma non-invasive
- 8201 Cribriform carcinoma in situ
- 8500 Ductal carcinoma in situ
- 8501 Comedocarcinoma, noninfiltrating
- 8503 Intraductal papilloma with ductal carcinoma in situ
- 8504 Encapsulated papillary carcinoma
- 8507 Intraductal micropapillary carcinoma
- 8520 Lobular carcinoma in situ
- 8522 Intraductal and lobular in situ carcinoma
- 8523 Intraductal mixed with oth types carcinoma in situ
- 8543 Paget disease and intraductal carcinoma
- Other noninvasive histology

Description

- Cancer histology type associated with specimen with in situ cancer
- Populated only if noninvasive or in situ cancer is present ([path_insitu1] –[path_insitu45] = "Yes")

Invasive Cancer Found in at Least One Specimen in the Pathology Report

[path_ca_inv_any]

Value (Character)

- Yes
- No

Description

- Indicates whether invasive cancer was found in any of the 45 specimens on the pathology report
- Based on [path_ca1] – [path_ca45]

Number of Specimens with Invasive Cancer in Pathology Report

[n_specimen_inv]

Value (Numeric)

- 0-45

Description

- Number of specimens with invasive cancer in pathology report
- Based on [path_ca1] – [path_ca45]

Invasive Cancer Identified in Specimen 1-45

[path_ca1] – [path_ca45]

Value (Character)

- Yes
- No

Description

- Indicates whether any invasive cancer (ICD-O-3 Behavior Code = 3) is present in the specimen

Invasive Cancer Type for Each Specimen 1-45

[path_ca_type1] – [path_ca_type45]

Value (Character)

- Adrenocortical Carcinoma
- Anal Cancer
- Appendix Cancer
- Bile Duct Cancer
- Bladder Cancer
- Brain Cancer
- Breast Cancer
- Breast Sarcoma
- NET or Carcinoid
- Cervical Cancer
- Colon Cancer
- Colon/Rectum Cancer
- Esophagus Cancer
- Ewing Sarcoma
- Fallopian Tube Cancer
- Gallbladder Cancer
- Germ Cell Tumor
- GIST
- Head and Neck Cancer
- Mesothelioma
- Ill Defined/Cancer of Unknown Primary
- Liver Cancer
- Lung Cancer, NOS
- Melanoma

- Merkel Cell
- Neuroblastoma
- Non Small Cell Lung Cancer
- Osteosarcoma
- Ovarian Cancer
- Pancreatic Cancer
- Parathyroid Cancer
- Penis Cancer
- Peritoneum Cancer
- Placenta Cancer
- Prostate Cancer
- Rectum and Rectosigmoid Cancer
- Renal Kidney Cancer
- Renal Pelvis Cancer
- Retinoblastoma
- Rhabdomyosarcoma
- Scrotum Cancer
- Small Cell Lung Cancer
- Small Intestine Cancer
- Soft tissue sarcoma
- Stomach Cancer
- Testis Cancer
- Thymus Cancer
- Thyroid Cancer
- Uterus Cancer NOS
- Vagina Cancer
- Vulva Cancer
- Wilms Tumor
- Leukemia
- Non Hodgkin's Lymphoma
- Hodgkins Lymphoma
- Multiple Myeloma
- Other hematopoietic or lymphoid neoplasm
- Other
- Not stated
- Unknown

Description

- Indicates cancer type identified in the specimen with invasive cancer
- Populated only if invasive cancer is present ([path_ca1] – [path_ca45] = “Yes”)

Cancer Histology Type for Specimen 1-45

[path_ca_hist1] – [path_ca_hist45]

Value (Character)

- ICD-O-3 morphology code

Description

- Cancer histology type associated with specimen with invasive cancer
- Populated only if invasive cancer is present ([path_ca1] –[path_ca45] = “Yes”)

Biomarkers

Up to three tests for PD-L1 can be associated with a pathology report. The biomarker information corresponding to each of the three tests is consistent across variables, i.e. information corresponding to the first PD-L1 test is stored in variables [pdl1_yn], [pdl1_prepaint], [pdl1_test], etc. and the information corresponding to the second PD-L1 test is stored in variables [pdl1_yn_2], [pdl1_prepaint_2], [pdl1_test_2], etc. Summary variables indicating any testing and any positive result across all three instances are also provided (i.e. [pdl1_testing] and [pdl1_positive_any]).

Any PD-L1 Testing Reported on Pathology Report

[pdl1_testing]

Value (Character)

- Yes
- No

Description

- Indicates whether any PD-L1 testing was reported on a given pathology report
- Combines variables [pdl1_yn], [pdl1_yn_2], [pdl1_yn_3]

PD-L1 Testing Reported

[pdl1_yn], [pdl1_yn_2], [pdl1_yn_3]

Value (Character)

- Yes
- No

Description

- Indicates whether PD-L1 testing was reported on a given pathology report
- Up to three PD-L1 tests on a single pathology report are curated
- Populated only if any in situ or invasive cancer identified in specimens 1-45 ([path_insitu1]-[path_insitu45] = “Yes” or [path_ca1]-[path_ca45] = “Yes”)
- Preferred derived variable [pdl1_testing] incorporates all three PD-L1 tests.

Time (Days) from Date of Birth to PD-L1 Report Date

[pdl1_prepaint], [pdl1_prepaint_2], [pdl1_prepaint_3]

Value (Numeric)

Description

- Interval in days from date of birth to pathology report PD-L1 test result
- Populated only if PD-L1 testing reported ([pdl1_yn], [pdl1_yn_2], [pdl1_yn_3] = “Yes”)

PD-L1 Antibody Test Type

[pdl1_test], [pdl1_test_2], [pdl1_test_3]

Value (Character)

- 22C3
- 28-8
- E1L3N
- SP142
- SP263
- Other
- Unknown

Description

- PD-L1 test identified in pathology report
- Up to three tests can be recorded per pathology report
- Populated only if PD-L1 testing reported ([pdl1_yn], [pdl1_yn_2], [pdl1_yn_3] = “Yes”)

PD-L1 Results Presented as a Percentage or Percentage Range of Tumor Cells

[pdl1_type___1], [pdl1_type_2___1], [pdl1_type_3___1]

Value (Character)

- Percentage or Percentage Range of Tumor Cells

Description

- Indicates that PD-L1 test results are represented as percentage or percentage range of tumor cells
- Populated only if PD-L1 results are represented as the percentage or percentage range of tumor cells

PD-L1 Results Presented as a Percentage or Percentage Range of Infiltrating Immune Cells

[pdl1_type___2], [pdl1_type_2___2], [pdl1_type_3___2]

Value (Character)

- Percentage or Percentage Range of Infiltrating Immune Cells

Description

- Indicates that PD-L1 test results are represented as the percentage or percentage range of infiltrating immune cells
- Populated only if PD-L1 results are represented as the percentage or percentage range of infiltrating immune cells

PD-L1 Results Presented as a Numeric Combined Positive Score

[pdl1_type___3], [pdl1_type_2___3], [pdl1_type_3___3]

Value (Character)

- Numeric (Combined Positive Score)

Description

- Indicates that PD-L1 test results are represented as a numeric combined positive score
- Populated only if PD-L1 results are represented as a numeric combined positive score

PD-L1 Results Presented as Summary Assessment

[pdl1_type___4], [pdl1_type_2___4], [pdl1_type_3___4]

Value (Character)

- Summary Assessment

Description

- Indicates that PD-L1 test results are represented as a summary assessment
- Populated only if PD-L1 results are represented as a summary assessment

Any Positive PD-L1 Result Reported on Pathology Report

[pdl1_positive_any]

Value (Character)

- Yes
- No

Description

- Indicates whether any PD-L1 testing on a given pathology report returned a positive result
- Based on any of the following criteria being met:
  - A low, high, or positive summary score
  - A percentage or percentage range tumor cells greater than 0
  - A percentage or percentage range of infiltrating immune cells greater than 0
- Populated only if PD-L1 testing is indicated on pathology report (i.e. [pdl1_testing] = “Yes”)

Percentage of Tumor Cells Positive for PD-L1

[pdl1_perc], [pdl1_perc_2], [pdl1_perc_3]

Value (Numeric)

- 0-100

Description

- The percentage of tumor cells positive for PD-L1
- Populated only if PD-L1 testing reported as percentage or percentage range of tumor cells ([pdl1_type___1], [pdl1_type_2___1], [pdl1_type_3___1] = “Percentage or Percentage Range of Tumor Cells”)

Lower Range (%) of Tumor Cells Positive for PD-L1

[pdl1_tclrange], [pdl1_tclrange_2], [pdl1_tclrange_3]

Value (Character)

- < 1
- 0-100
- Other
- Not Applicable

Description

- Lower range (%) of tumor cells positive for PD-L1
- Populated only if PD-L1 testing reported as percentage or percentage range of tumor cells ([pdl1_type___1], [pdl1_type_2___1], [pdl1_type_3___1] = “Percentage or Percentage Range of Tumor Cells”)

Upper Range (%) of Tumor Cells Positive for PD-L1

[pdl1_tcurange], [pdl1_tcurange_2], [pdl1_tcurange_3]

Value (Character)

- < 1
- 0-100
- Other
- Not Applicable

Description

- Upper range (%) of tumor cells positive for PD-L1
- Populated only if PD-L1 testing reported as percentage or percentage range of tumor cells ([pdl1_type___1], [pdl1_type_2___1], [pdl1_type_3___1] = “Percentage or Percentage Range of Tumor Cells”)

Percentage of Infiltrating Immune Cells Positive for PD-L1

[pdl1_icperc], [pdl1_icperc_2], [pdl1_icperc_3]

Value (Numeric)

- 0-100

Description

- The percentage of infiltrating immune cells positive for PD-L1
- Populated only if PD-L1 testing reported as percentage or percentage range of infiltrating immune cells ([pdl1_type___2], [pdl1_type_2___2], [pdl1_type_3___2] = “Percentage or Percentage Range of Infiltrating Immune Cells”)

Lower Range (%) of Infiltrating Immune Cells Positive for PD-L1

[pdl1_iclrange], [pdl1_iclrange_2], [pdl1_iclrange_3]

Value (Character)

- < 1
- 0-100
- Other
- Not Applicable

Description

- Lower range (%) of infiltrating immune cells positive for PD-L1
- Populated only if PD-L1 testing reported as percentage or percentage range of infiltrating immune cells ([pdl1_type___2], [pdl1_type_2___2], [pdl1_type_3___2] = “Percentage or Percentage Range of Infiltrating Immune Cells”)

Upper Range (%) of Infiltrating Immune Cells Positive for PD-L1

[pdl1_icurange], [pdl1_icurange_2], [pdl1_icurange_3]

Value (Character)

- < 1
- 0-100
- Other
- Not Applicable

Description

- Upper range (%) of infiltrating immune cells positive for PD-L1
- Populated only if PD-L1 testing reported as percentage or percentage range of infiltrating immune cells ([pdl1_type___2], [pdl1_type_2___2], [pdl1_type_3___2] = “Percentage or Percentage Range of Infiltrating Immune Cells”)

Numeric Combined Positive Score (CPS) for PD-L1

[pdl1_num], [pdl1_num_2], [pdl1_num_3]

Value (Numeric)

Description

- The Combined Positive Score (CPS) value
- The minimum CPS is 0
- Populated only if PD-L1 testing reported as a numeric combined positive score ([pdl1_type___3], [pdl1_type_2___3], [pdl1_type_3___3] = “Numeric (Combined Positive Score)”)

Lower Range of the Combined Positive Score (CPS) for PD-L1

[pdl1_lcpsrange], [pdl1_lcpsrange_2], [pdl1_lcpsrange_3]

Value (Character)

- < 1
- 0-100
- Other
- Not Applicable

Description

- The lower range of the Combined Positive Score (CPS)
- Populated only if PD-L1 testing reported as a numeric combined positive score

([pdl1_type___3], [pdl1_type_2___3], [pdl1_type_3___3] = “Numeric (Combined Positive Score)”)

Upper Range of the Combined Positive Score (CPS) for PD-L1

[pdl1_ucpsrange], [pdl1_ucpsrange_2], [pdl1_ucpsrange_3]

Value (Character)

- < 1
- 0-100
- Other
- Not Applicable

Description

- The upper range of the Combined Positive Score (CPS)
- Populated only if PD-L1 testing reported as a numeric combined positive score ([pdl1_type___3], [pdl1_type_2___3], [pdl1_type_3___3] = “Numeric (Combined Positive Score)”)

Summary Assessment of PD-L1

[pdl1_sum], [pdl1_ sum _2], [pdl1_ sum _3]

Value (Character)

- High
- Low
- Positive
- Negative
- Indeterminate/Not stated

Description

- Overall summary assessment value of PD-L1 as stated in pathology report; not specific to immune cells or tumor cells
- Populated only if PD-L1 testing reported as a summary score ([pdl1_type___4], [pdl1_type_2___4], [pdl1_type_3___4] = “Summary Assessment”)

Any Microsatellite Instability (MSI) Testing Reported on Pathology Report

[msi_testing]

Value (Character)

- Yes
- No

Description

- Indicates whether any microsatellite instability (MSI) testing was reported on pathology report
- Combines variables [msi_yn], [msi_yn_2], [msi_yn_3]

Microsatellite Instability (MSI) Testing Using Polymerase Chain Reaction (PCR) Reported

[msi_yn], [msi_yn_2], [msi_yn_3]

Value (Character)

- Yes
- No

Description

- Indicates whether microsatellite instability (MSI) testing using PCR was reported for a specimen containing cancer

Time (Days) from Date of Birth to Microsatellite Instability (MSI) Testing Report Date

[msi_prepaint], [msi_prepaint_2], [msi_prepaint_3]

Value (Numeric)

Description

- Interval in days from date of birth to report date for pathology report Microsatellite Instability (MSI) test result
- Populated only if MSI testing reported ([msi_yn], [msi_yn_2], [msi_yn_3] = “Yes”)

Microsatellite Instability (MSI) Result

[msi_result], [msi_result_2], [msi_result_3]

Value (Character)

- MSS: STABLE
- MSI-H: HIGH

- MSI-L: LOW or instability in < 30% of microsatellite markers
- Indeterminate/Not stated

Description

- Microsatellite Instability (MSI) result from pathology report
- Populated only if MSI testing reported ([msi_testing] = “Yes”)

Any MSI-H test result

[msi_high_any]

Value (Character)

- Yes
- No

Description

- Indicates whether any Microsatellite Instability (MSI) result is high ([msi_result], [msi_result_2], [msi_result_3] = “MSI-H: High)

Microsatellite Instability (MSI) – H: High Details

[msi_high], [msi_high_2], [msi_high_3]

Value (Character)

- >=30% of the markers exhibit instability
- 2 or more of the 5 markers exhibit instability
- Other

Description

- Details of result if Microsatellite Instability (MSI) testing is MSI-H: HIGH
- Populated only if MSI result is high ([msi_result], [msi_result_2], [msi_result_3] = “MSI-H: HIGH”)

Microsatellite Instability (MSI) – L: Low Details

[msi_low], [msi_low_2], [msi_low_3]

Value (Character)

- 1 - 29% of the markers exhibit instability
- 1 of the 5 markers exhibit instability
- Other

Description

- Details of result if Microsatellite Instability (MSI) testing is MSI-L: Low
- Populated only if MSI result is low ([msi_result], [msi_result_2], [msi_result_3] = “MSI-L: Low”)

Any Mismatch Repair (MMR) Testing Reported on Pathology Report

[mmr_testing]

Value (Character)

- Yes
- No

Description

- Indicates whether any mismatch repair (MMR) testing was reported on pathology report
- Combines variables [mmr_yn], [mmr_yn_2], [mmr_yn_3]

Mismatch Repair (MMR) or Expression/Presence of MH1, or MSH2 or MSH6 or PMS2 Testing

[mmr_yn], [mmr_yn_2], [mmr_yn_3]

Value (Character)

- Yes
- No

Description

- Indicates whether testing was done for Mismatch Repair or expression/presence of MLH1, MSH2 or MSH6 or PMS2 for a specimen containing cancer

Mismatch Repair (MMR): Time (Days) from Date of Birth to MMR Testing Report Date

[mmr_prepaint], [mmr_prepaint_2], [mmr_prepaint_3]

Value (Numeric)

Description

- Interval in days from date of birth to report date of pathology report with Mismatch Repair (MMR) testing
- Populated only if MMR testing reported ([mmr_yn], [mmr_yn_2], [mmr_yn_3] = “Yes”)

Mismatch Repair (MMR): MLH1 Expression

[mmr_mlh1], [mmr_mlh1_2], [mmr_mlh1_3]

Value (Character)

- Intact expression/present
- Loss of expression/absent
- Cannot be determined/Not mentioned

Description

- Specifies MLH1 expression from Mismatch Repair (MMR) testing
- Populated only if MMR testing reported ([mmr_testing] = “Yes”)

Mismatch Repair (MMR): MSH2 Expression

[mmr_msh2], [mmr_msh2_2], [mmr_msh2_3]

Value (Character)

- Intact expression/present
- Loss of expression/absent
- Cannot be determined/Not mentioned

Description

- Specifies MSH2 expression from Mismatch Repair (MMR) testing
- Populated only if MMR testing reported ([mmr_testing] = “Yes”)

Mismatch Repair (MMR): MSH6 Expression

[mmr_msh6], [mmr_msh6_2], [mmr_msh6_3]

Value (Character)

- Intact expression/present
- Loss of expression/absent
- Cannot be determined/Not mentioned

Description

- Specifies MSH6 expression from Mismatch Repair (MMR) testing
- Populated only if MMR testing reported ([mmr_testing] = “Yes”)

Mismatch Repair (MMR): PMS2 Expression

[mmr_pms2], [mmr_pms2_2], [mmr_pms2_3]

Value (Character)

- Intact expression/present
- Loss of expression/absent
- Cannot be determined/Not mentioned

Description

- Specifies PMS2 expression from Mismatch Repair (MMR) testing
- Populated only if MMR testing reported ([mmr_testing] = “Yes”)

Mismatch Repair (MMR): Overall Interpretation

[mmr_result], [mmr_result_2], [mmr_result_3]

Value (Character)

- No loss of nuclear expression of MMR proteins: No evidence of deficient mismatch repair (low probability of MSI-H): proficient
- Loss of nuclear expression of one or more MMR proteins; deficient mismatch repair
- Indeterminate/Not stated

Description

- Overall Mismatch Repair (MMR) interpretation
- Populated only if MMR testing reported ([mmr_testing] = “Yes”)

Mismatch Repair Proficient (MMR-P): Report Details

[mmrp_det], [mmrp_det_2], [mmrp_det _3]

Value (Character)

- No loss of nuclear expression of MLH1, MSH2, MSH6 or PMS2 proteins
- Described as MMR intact, preserved, normal

Description

- Details of MMR status that were recorded in pathology report if interpretation of MMR testing was MMR-P: Proficient
- Populated only if MMR result is proficient ([mmr_result], [mmr_result_2], [mmr_result_3] = “No loss of nuclear expression of MMR proteins: No evidence of deficient mismatch repair (low probability of MSI-H): proficient”)

Mismatch Repair Deficient (MMR-D): Report Details

[mmrd_det], [mmrd_det _2], [mmrd_det _3]

Value (Character)

- Loss of nuclear expression (or absence of) one or more of MLH1, MSH2, MSH6 or PMS2 proteins
- Described as abnormal

Description

- Details of MMR status that were recorded in pathology report if interpretation of MMR testing was MMR Deficient (MMR-D)
- Populated only if MMR result is deficient ([mmr_result], [mmr_result_2], [mmr_result_3] = “Loss of nuclear expression of one or more MMR proteins; deficient mismatch repair”)

Hormone Receptors (HR) or HER2 Testing Reported

[path_erprher_yn_1], [path_erprher_yn_2], [path_erprher_yn_3], [path_erprher_yn_4], [path_erprher_yn_5]

Value (Character)

- Yes
- No

Description

- Indicates whether hormone receptor (ER, PR) or HER2 testing was reported within pathology sample

ER Result

[path_er_1], [path_er_2], [path_er_3], [path_er_4], [path_er_5]

Value (Character)

- Negative
- Equivocal

- Positive
- Test not done

Description

- ER test result value
- Populated only if hormone receptor (ER, PR) or HER2 testing was reported ([path_erprher_yn_1], [path_erprher_yn_2], [path_erprher_yn_3], [path_erprher_yn_4], [path_erprher_yn_5] = “Yes”)

PR Result

[path_pr_1], [path_pr_2], [path_pr_3], [path_pr_4], [path_pr_5]

Value (Character)

- Negative
- Equivocal
- Positive
- Test not done

Description

- PR test result value
- Populated only if hormone receptor (ER, PR) or HER2 testing was reported ([path_erprher_yn_1], [path_erprher_yn_2], [path_erprher_yn_3], [path_erprher_yn_4], [path_erprher_yn_5] = “Yes”)

HER2 IHC Result

[path_herihc_1], [path_herihc_2], [path_herihc_3], [path_herihc_4], [path_herihc_5]

Value (Character)

- Negative
- Negative (1+) IHC only
- Equivocal
- Positive
- Test not done

Description

- HER2 immunohistochemistry test result value
- Populated only if hormone receptor (ER, PR) or HER2 testing was reported ([path_erprher_yn_1], [path_erprher_yn_2], [path_erprher_yn_3], [path_erprher_yn_4], [path_erprher_yn_5] = “Yes”)

HER2 ISH Result

[path_herish_1], [path_herish_2], [path_herish_3], [path_herish_4], [path_herish_5]

Value (Character)

- Negative
- Negative (1+) IHC only
- Equivocal
- Positive
- Test not done

Description

- HER2 in situ hybridization test result value
- Populated only if hormone receptor (ER, PR) or HER2 testing was reported ([path_erprher_yn_1], [path_erprher_yn_2], [path_erprher_yn_3], [path_erprher_yn_4], [path_erprher_yn_5] = “Yes”)

Time (Days) from Date of Birth to Date of HR/HER2 Addendum Path Report

[path_erprher_add1_int], [path_erprher_add2_int], [path_erprher_add3_int], [path_erprher_add4_int], [path_erprher_add5_int]

Value (Numeric)

Description

- Time (days) from date of birth to date of HR/HER2 pathology report
- Populated only if:
  - Hormone receptors (ER, PR) or HER2 testing was reported ([path_erprher_yn_1], [path_erprher_yn_2], [path_erprher_yn_3], [path_erprher_yn_4], [path_erprher_yn_5] = “Yes”)
  - The HR and/or HER2 results were provided in an addendum report that may have a different date from the primary pathology report

Release Version Number

[release_version]

Value (Character)

- 1.0-public

Description

- Indicates the version number of the data release

PRISSMM IMAGING DATASET

The PRISSMM Imaging dataset is structured as one record per imaging report, per patient. Imaging reports were curated beginning within 30 days of the first BPC Project cancer diagnosis. All subsequent imaging reports were recorded (including imaging reports corresponding to non-BPC Project Cancers and subsequent BPC Project Cancer diagnoses; Appendix 1).

The PRISSMM Imaging dataset can be linked to all datasets using the variables [cohort] and [record_id].

Field names shaded in gray indicate that an alternative variable is preferred for analysis. The preferred variable is noted in the description.

A small discrepancy was identified in the approach to curation of multi-focal metastatic sites of disease that affects the imaging module only. Specifically, when a patient experienced complete resolution of a metastatic locus of disease, the curators at the MSK site interpreted this as "evidence of disease" and "responding"; however, curators at other sites interpreted this as "no evidence of disease" (and thus no response option was recorded). Because complete resolution of metastatic disease foci is quite rare, this is expected to have minimal effects on ascertaining response. This is anticipated to have no effect on estimation of time to progression. The consensus approach is to consider these sites as "no evidence of cancer".

BPC Project Cohort

[cohort]

Value (Character)

- BrCa

Description

- Indicates the BPC Project Cancer type
- Primary key for merging across all datasets

Record ID

[record_id]

Value (Character)

- GENIE-[INSTITUTION]-XXXXXX

Description

- De-identified, unique patient ID
- Conforms to the following the convention: GENIE-[INSTITUTION]-XXXX. The first component is the string, "GENIE"; the second component is the institution's abbreviation; the third component is a unique ID for the patient.
- Primary key for the AACR Project GENIE genomic datasets [PATIENT_ID].

Institution

[institution]

Value (Character)

- DFCI = Dana Farber Cancer Institute
- MSK = Memorial Sloan Kettering Cancer Center
- VICC = Vanderbilt Ingram Cancer Center

Description

- Indicates the patient's internal institution of cancer care
- Corresponds to variable [center] in AACR Project GENIE data

Imaging Report Number

[scan_number]

Value (Numeric)

Description

- Unique identifier for imaging reports based on the date of the scan [image_scan_int]
- Different scans occurring on the same date have distinct scan numbers (e.g. a CT and MRI occurring on the same date may be numbered as scan 1 and scan 2, though the ordering is arbitrary)

Time (Days) from Date of Birth to Imaging Date

[image_scan_int]

Value (Numeric)

Description

- Interval in days from date of birth to scan date

Time (Days, Months, Years) from First BPC Project Cancer to Imaging Date

[dx_scan_days], [dx_scan_mos], [dx_scan_yrs]

Value (Numeric)

Description

- Interval in days [dx_scan_days]; months [dx_scan_mos]; or years [dx_scan_yrs] from first BPC Project Cancer diagnosis to scan date

Time (Days) from Date of Birth to Reference Imaging Date

[image_ref_scan_int]

Value (Numeric)

Description

- Interval days in from date of birth to reference imaging date
- Populated only if there is evidence of cancer on the imaging report ([image_ca] = "Yes, the Impression states or implies there is evidence of cancer")

Time (Days, Months, Years) from First BPC Project Cancer to Reference Imaging Date

[dx_ref_scan_days], [dx_ref_scan_mos], [dx_ref_scan_yrs]

Value (Numeric)

Description

- Interval in days [dx_ref_scan_days]; months [dx_ref_scan_mos]; or years [dx_ref_scan_yrs] from first BPC Project Cancer diagnosis to reference imaging date

Institution Where Scan was Performed

[image_inst_perf]

Value (Character)

- Internal institution
- External institution

Description

- Indicates whether scan was performed at the internal (i.e. the same institution as the [institution] variable) or external institution.

Institution Where Image was Interpreted

[image_inst_inter]

Value (Character)

- Internal institution
- External institution

Description

- Indicates whether image was interpreted at the internal (i.e. the same institution as the [institution] variable) or external institution.
- Populated only if imaging was performed at an external institution ([image_inst_perf] = "External institution")

Imaging Scan Type

[image_scan_type]

Value (Character)

- CT
- MRI
- PET or PET-CT
- Bone Scan
- Other Nuclear Medicine Scan
- Mammogram - Use for Breast Cancer only
- Other CA-specific scan

Description

- Type of imaging scan

Imaging Site: Brain/Head

[image_scansite___1]

Value (Character)

- Brain/Head

Description

- Indicates a scan of the brain/head
- Populated only if scan is of brain/head

Imaging Site: Spine

[image_scansite___2]

Value (Character)

- Spine

Description

- Indicates a scan of the spine
- Populated only if scan is of the spine

Imaging Site: Neck

[image_scansite___3]

Value (Character)

- Neck

Description

- Indicates a scan of the neck
- Populated only if scan is of the neck

Imaging Site: Chest

[image_scansite___4]

Value (Character)

- Chest

Description

- Indicates a scan of the chest
- Populated only if scan is of the chest

Imaging Site: Abdomen

[image_scansite___5]

Value (Character)

- Abdomen

Description

- Indicates a scan of the abdomen
- Populated only if scan is of the abdomen

Imaging Site: Pelvis

[image_scansite___6]

Value (Character)

- Pelvis

Description

- Indicates a scan of the pelvis
- Populated only if scan is of the pelvis

Imaging Site: Extremity

[image_scansite___7]

Value (Character)

- Extremity

Description

- Indicates a scan of an extremity
- Populated only if scan is of an extremity

Imaging Site: Full body

[image_scansite___8]

Value (Character)

- Full body

Description

- Indicates a full body scan
- Populated only if a full body scan

Imaging Sites

[scan_sites]

Value (Character)

Description

- List of sites scanned on this imaging report
- Concatenation of sites scanned ([image_scansite___1] – [image_scansite___8])

Radiologist Assessment of any Evidence of Cancer on this Imaging Report

[image_ca]

Value (Character)

- Yes, the Impression states or implies there is evidence of cancer
- No, the Impression states or implies there is no evidence of cancer
- The Impression is uncertain, indeterminate, or equivocal
- The Impression does not mention cancer

Description

- Indicates whether the radiologist assessment indicates any evidence of cancer

Radiologist Assessment of Change in Cancer Status

[image_overall]

Value (Character)

- Improving/Responding
- Stable/No change
- Mixed
- Progressing/Worsening/Enlarging
- Not stated/Indeterminate

Description

- Radiologist’s assessment of the change in the patient’s cancer status
- Populated only if there is evidence of cancer on the imaging report ([image_ca] = “Yes, the Impression states or implies there is evidence of cancer”)

Location of Cancer Based on Imaging Report

[image_casite1]-[image_casite15]

Value (Character)

- ICD-O-3 topography code

Description

- Location of cancer on imaging report based on ICD-O-3 topography code
- Populated only if there is evidence of cancer on the imaging report ([image_ca] = "Yes, the Impression states or implies there is evidence of cancer")

Release Version Number

[release_version]

Value (Character)

- 1.0-public

Description

- Indicates the version number of the data release

PRISSMM MEDICAL ONCOLOGIST ASSESSMENT DATASET

The PRISSMM Medical Oncologist Assessment dataset is structured as one record per curated medical oncologist assessment, per patient. Medical oncologist assessments were curated beginning with the month and year of the first BPC Project Cancer diagnosis (Appendix 1). One medical oncologist assessment per month was curated; curation instructions regarding the selection of the assessment to curate are provided in Appendix 3.

The PRISSMM Medical Oncologist Assessment dataset can be linked to all datasets using the variables [cohort] and [record_id].

Field names shaded in gray indicate that an alternative variable is preferred for analysis. The preferred variable is noted in the description.

BPC Project Cohort

[cohort]

Value (Character)

- BrCa

Description

- Indicates the BPC Project Cancer type
- Primary key for merging across all datasets

Record ID

[record_id]

Value (Character)

- GENIE-[INSTITUTION]-XXXXXX

Description

- De-identified, unique patient ID
- Conforms to the following the convention: GENIE-[INSTITUTION]-XXXX. The first component is the string, “GENIE”; the second component is the institution’s abbreviation; the third component is a unique ID for the patient.
- Primary key for the AACR Project GENIE genomic datasets [PATIENT_ID].

Institution

[institution]

Value (Character)

- DFCI = Dana Farber Cancer Institute
- MSK = Memorial Sloan Kettering Cancer Center
- VICC = Vanderbilt Ingram Cancer Center

Description

- Indicates the patient's internal institution of cancer care
- Corresponds to variable [center] in AACR Project GENIE data

Medical Oncologist Visit Identifier

[md_visit_number]

Value (Numeric)

Description

- Unique identifier for curated medical oncologist assessments based on the visit date [md_onc_visit_int]
- Only one medical oncologist assessment per month was curated (Appendix 3)

Time (Days) from Date of Birth to Medical Oncologist Visit

[md_onc_visit_int]

Value (Numeric)

Description

- Interval in days from date of birth to date of medical oncologist visit
- Based on the date the visit occurred, not the date the assessment was signed or uploaded.

Time (Days, Months, Years) from First BPC Project Cancer to Medical Oncologist Visit

[dx_md_visit_days], [dx_md_visit_mos], [dx_md_visit_yrs]

Value (Numeric)

Description

- Interval in days [dx_md_visit_days]; months [dx_md_visit_mos]; or years [dx_md_visit_yrs] from first BPC Project Cancer diagnosis to medical oncologist visit.
    - Based on the date the visit occurred, not the date the assessment was signed or uploaded.

Institution Where Medical Oncology Visit Occurred

[md_inst]

Value (Character)

- Internal institution
- External institution

Description

- Indicates whether the medical oncology visit was at the internal (i.e. the same institution as the [institution] variable) or external institution.

Cancer Diagnosis Assessed by Medical Oncologist

[md_type_ca_cur]

Value (Character)

- Adrenocortical Carcinoma
- Anal Cancer
- Appendix Cancer
- Bile Duct Cancer
- Bladder Cancer
- Brain Cancer
- Breast Cancer
- Breast Sarcoma
- Cervical Cancer
- Colon Cancer
- Colon/Rectum Cancer
- Esophagus Cancer
- Ewing Sarcoma
- Fallopian Tube Cancer
- Gallbladder Cancer
- Germ Cell Tumor
- GIST
- Head and Neck Cancer
- Mesothelioma
- Ill Defined/Cancer of Unknown Primary
- Liver Cancer
- Lung Cancer, NOS

- Melanoma
- Merkel Cell
- Neuroblastoma
- NET or Carcinoid
- Non Small Cell Lung Cancer
- Osteosarcoma
- Ovarian Cancer
- Pancreatic Cancer
- Parathyroid Cancer
- Penis Cancer
- Peritoneum Cancer
- Placenta Cancer
- Prostate Cancer
- Rectum and Rectosigmoid Cancer
- Renal Kidney Cancer
- Renal Pelvis Cancer
- Retinoblastoma
- Rhabdomyosarcoma
- Scrotum Cancer
- Small Cell Lung Cancer
- Small Intestine Cancer
- Stomach Cancer
- Testis Cancer
- Thymus Cancer
- Thyroid Cancer
- Uterus Cancer
- Vagina Cancer
- Vulva Cancer
- Wilms Tumor
- Other
- Not stated

Description

- Indicates the cancer diagnosis associated with the given medical oncologist assessment

Medical Oncologist Assessment of Evidence of Cancer

[md_ca]

Value (Character)

- Yes, the Impression/Plan states or implies there is evidence of cancer
- No, the Impression/Plan states or implies there is no evidence of cancer
- Impression/Plan is uncertain, indeterminate, or equivocal
- Impression/Plan does not mention cancer

Description

- Medical oncologist’s assessment of whether there is evidence of cancer

Medical Oncologist Assessment of Change in Cancer Status

[md_ca_status]

Value (Character)

- Improving/Responding
- Stable/No change
- Mixed
- Progressing/Worsening/Enlarging
- Not stated/Indeterminate

Description

- Medical oncologist’s assessment of the change in the patient’s cancer status
- Populated only if there is evidence of cancer on the medical oncologist assessment ([md_ca] = "Yes, the Impression/Plan states or implies there is evidence of cancer")

Release Version Number

[release_version]

Value (Character)

- 1.0-public

Description

- Indicates the version number of the data release

PRISSMM TUMOR MARKER DATASET

The PRISSMM Tumor Marker dataset is structured as one record per curated tumor marker result, per patient. All serum-based tumor markers that are related to the diagnosis/prognosis of cancer were curated.

Note: Variables pertaining to PD-L1, MSI and MMR are recorded in the pathology dataset.

The PRISSMM Tumor Marker dataset can be linked to all datasets using the variables [cohort] and [record_id].

Field names shaded in gray indicate that an alternative variable is preferred for analysis. The preferred variable is noted in the description.

BPC Project Cohort

[cohort]

Value (Character)

- BrCa

Description

- Indicates the BPC Project Cancer type
- Primary key for merging across all datasets

Record ID

[record_id]

Value (Character)

- GENIE-[INSTITUTION]-XXXXXX

Description

- De-identified, unique patient ID
- Conforms to the following the convention: GENIE-[INSTITUTION]-XXXX. The first component is the string, “GENIE”; the second component is the institution’s abbreviation; the third component is a unique ID for the patient.
- Primary key for the AACR Project GENIE genomic datasets [PATIENT_ID].

Institution

[institution]

Value (Character)

- DFCI = Dana Farber Cancer Institute
- MSK = Memorial Sloan Kettering Cancer Center
- VICC = Vanderbilt Ingram Cancer Center

Description

- Indicates the patient's internal institution of cancer care
- Corresponds to variable [center] in AACR Project GENIE data

Tumor Marker Assessment Number

[tm_number]

Value (Numeric)

Description

- Unique identifier for tumor marker assessment based on the specimen collection date [tm_spec_collect_int]

Time (Days) from Date of Birth to Date of Tumor Marker Specimen Collection

[tm_spec_collect_int]

Value (Numeric)

Description

- Interval in days from date of birth to tumor marker specimen collection

Time (Days, Months, Years) from First BPC Project Cancer to Tumor Marker Collection

[dx_tm_days], [dx_tm_mos], [dx_tm_yrs]

Value (Numeric)

Description

- Interval in days [dx_tm_days]; months [dx_tm_mos]; or years [dx_tm_yrs] from first BPC Project Cancer diagnosis to tumor marker specimen collection

Tumor Marker Lab Test

[tm_type]

Value (Character)

- CA15-3
- CA19-9
- CA2729
- CEA
- PSA
- Testosterone

Description

- The type of tumor marker test
- There may be multiple tumor markers obtained on the same date

Tumor Marker Result

[tm_num_result]

Value (Numeric)

Description

- Numeric value of tumor marker result
- For CEA tests ([tm_type] = “CEA”), tumor marker results of 0.4 indicate non-zero and non-elevated values

Tumor Marker Result Unit

[tm_result_units]

Value (Character)

- kU/L
- mIU/mL
- ng/mL
- ng/dL
- nmol/L
- pg/mL
- U/mL
- ug/L

Description

- The units corresponding to the tumor marker results [tm_num_results]

Tumor Marker Lower Limit of Normal

[tm_normal_range_lower]

Value (Numeric)

Description

- Lower limit of normal for tumor marker results

Tumor Marker Upper Limit of Normal

[tm_normal_range_upper]

Value (Numeric)

Description

- Upper limit of normal for tumor marker results

Release Version Number

[release_version]

Value (Character)

- 1.0-public

Description

- Indicates the version number of the data release

CANCER PANEL TEST (NEXT GENERATION SEQUENCING) DATASET

The Cancer Panel Test refers to high-throughput next generation sequencing (NGS) that has been performed through multi-gene panels. The Cancer Panel Test dataset is structured as one record per cancer panel test and associated cancer diagnosis, per patient. For example, if a cancer panel test was definitively associated with one cancer diagnosis, there will be one corresponding record in this dataset. If the curator was unsure of which diagnosis out of two cancer diagnoses corresponded to the cancer panel test, there will be two corresponding records in this dataset: one for the first potentially associated cancer diagnosis and another for the second potentially associated cancer diagnosis. All BPC Project Cancers have an associated cancer panel test (Appendix 1).

The terms "cancer panel test (CPT)" and "next generation sequencing (NGS)" are used interchangeably.

Note that the NGS dataset serves as the link between the clinical and genomic data, where the NGS dataset includes one record per NGS report per patient, including the NGS sample ID that is used to link to the genomic data files. These data can be accessed through either Synapse or cBioPortal.

- *Some data sets (e.g. CNA files) may appear in wide format in Synapse data versus long format in cBioPortal data, or column attributes and names may appear sightly different (e.g. fusions files).*
- *By default, cBioPortal filters out Silent, Intron, IGR, 3'UTR, 5'UTR, 3'Flank and 5'Flank, except for the promoter mutations of the TERT gene. See cBioPortal documentation for more details. These mutations are retained in Synapse processing pipelines.*
- *Some genes have more than one accepted Hugo Symbol and may be referred to differently between data sources (e.g. NSD3 is an alias for WHSC1L1).*

The Cancer Panel Test dataset can be linked to the following datasets:

- *BPC Project Cancer Diagnosis and Cancer-Directed Regimen datasets using the variables [cohort], [record_id], and [ca_seq].*
- *PRISSMM Pathology dataset using the variables [cohort], [record_id], [path_proc_number], and [path_report_number].*
- *Patient Characteristics, PRISSMM Imaging, and PRISSMM Medical Oncologist Assessment, and PRISSMM Tumor Marker datasets using the variables [cohort] and [record_id].*
- *Cannot be linked to the Non-BPC Project Cancer Diagnosis dataset because non-BPC Project Cancer diagnoses were not genomically sequenced (Appendix 1)*

Field names shaded in gray indicate that an alternative variable is preferred for analysis. The preferred variable is noted in the description.

BPC Project Cohort

[cohort]

Value (Character)

- BrCa

Description

- Indicates the BPC Project Cancer type
- Primary key for merging across all datasets

Record ID

[record_id]

Value (Character)

- GENIE-[INSTITUTION]-XXXXXX

Description

- De-identified, unique patient ID
- Conforms to the following the convention: GENIE-[INSTITUTION]-XXXX. The first component is the string, “GENIE”; the second component is the institution’s abbreviation; the third component is a unique ID for the patient.
- Primary key for the AACR Project GENIE genomic datasets [PATIENT_ID].

Institution

[institution]

Value (Character)

- DFCI = Dana Farber Cancer Institute
- MSK = Memorial Sloan Kettering Cancer Center
- VICC = Vanderbilt Ingram Cancer Center

Description

- Indicates the patient’s internal institution of cancer care
- Corresponds to variable [center] in AACR Project GENIE data

Cancer Panel Test (Next Generation Sequencing) Number

[cpt_number]

Value (Numeric)

- 1 = First curated next generation sequencing (NGS) test for this patient
- 2 = Second curated NGS test for this patient
- … n = nth NGS test for this patient

Description

- Order for the curated next generation sequencing (NGS) test based on the report date

Cancer Diagnosis Number Associated with NGS Test

[ca_seq]

Value (Numeric)

- 0 = first and only cancer
- 1 = first of two or more primaries
- 2 = second of two or more primaries
- …10 = tenth of ten or more primaries

Description

- The cancer sequence number associated with this NGS
- If the NGS test cannot be definitively associated with a single cancer diagnosis, this variable represents the cancer sequence of each potentially associated cancer diagnosis. If that is the case, each potentially associated cancer diagnosis will be a separate row in the dataset.
- Populated only if the diagnosis associated with the NGS test is known
- Primary key for the Cancer Diagnosis, Cancer-Directed Regimen and Cancer Panel Test datasets

Number of Cancer Diagnoses Associated with NGS Test

[cpt_n_ca_seq]

Value (Numeric)

Description

- The number of associated cancer diagnoses for this NGS test
- If >1 then the NGS test cannot be definitively associated with a single cancer diagnosis

Time (Days) from Date of Birth to NGS Order Date

[cpt_order_int]

Value (Numeric)

Description

- Interval in days from date of birth to the date that the cancer panel test was ordered
- This variable is not available for all institutions

Year of Next Generation Sequencing

[cpt_seq_date]

Value (Numeric)

Description

- Year of NGS

Derived Time (Days, Months, Years) from Date of Birth to NGS Report

[dob_cpt_report_days], [dob_cpt_report_mos], [dob_cpt_report_yrs]

Value (Numeric)

Description

- Interval in days [dob_cpt_report_days], months [dob_cpt_report_mos] and years [dob_cpt_report_yrs] from date of birth to date of sequencing report

NGS Test Report Returned On or After Date of Death

[cpt_report_post_death]

Value (Numeric)

- 1 = Yes, cancer panel test was returned on or after patient's date of death
- 0 = No, cancer panel test was not returned on or after patient's date of death

Description

- Indicates whether cancer panel test report was returned on or after patient's date of death

NGS Test Report Returned On or After Date Patient Last Known Alive

[cpt_report_post_last_alive]

Value (Numeric)

- 1 = Yes, cancer panel test was returned on or after patient's last known alive date
- 0 = No, cancer panel test was not returned on or after patient's last known alive date

Description

- Indicates whether cancer panel test report was returned on or after patient’s last known alive date
- Populated only if patient was not known to be dead at the time of curation. For an indicator for whether the cancer panel test report was returned after death, see variable [cpt_report_post_death].

Time (Days, Months, Years) from Diagnosis to NGS Report

[dx_cpt_rep_days], [dx_cpt_rep_mos], [dx_cpt_rep_yrs]

Value (Numeric)

Description

- Interval in days [dx_cpt_rep_days]; months [dx_cpt_rep_mos]; or years [dx_cpt_rep_yrs] from cancer diagnosis to sequencing report

Time (Days, Months, Years) from Diagnosis to Pathology Procedure Corresponding to the NGS Report

[dx_path_proc_cpt_days], [dx_path_proc_cpt_mos], [dx_path_proc_cpt_yrs]

Value (Numeric)

Description

- Interval in days [dx_path_proc_cpt_days]; months [dx_path_proc_cpt_mos]; or years [dx_path_proc_cpt_yrs] from cancer diagnosis to date of pathology procedure corresponding to cancer panel test

Time (Days, Months, Years) from Pathology Procedure to the NGS Report Date

[path_proc_cpt_rep_days], [path_proc_cpt_rep_mos], [path_proc_cpt_rep_yrs]

Value (Numeric)

Description

- Interval in days [path_proc_cpt_rep_days]; months [path_proc_cpt_rep_mos]; or years [path_proc_cpt_rep_yrs] from date of pathology procedure corresponding to cancer panel test to date of cancer panel test report

Pathology Procedure Number of NGS Specimen

[path_proc_number]

Value (Numeric)

Description

- Pathology procedure in which the specimen is described
- Primary key for PRISSMM Pathology and Cancer Panel Test datasets

Pathology Report Number of NGS Specimen

[path_rep_number]

Value (Numeric)

Description

- Pathology report in which the specimen is described
- Primary key for PRISSMM Pathology and Cancer Panel Test datasets

GENIE Sample ID

[cpt_genie_sample_id]

Value (Character)

Description

- GENIE sample ID corresponding to specimen
- Corresponds to variable [sample_id] in AACR Tier 1 data

NGS Specimen OncoTree Diagnosis Code

[cpt_oncotree_code]

Value (Character)

Description

- The primary cancer diagnosis code based on the OncoTree ontology
- Corresponds to variable [oncotree_code] in AACR Tier 1 data

Data Standard: OncoTree Ontology

Specimen Sample Type

[sample_type]

Value (Character)

- Primary tumor
- Lymph node metastasis
- Distant organ metastasis
- Metastasis site unspecified
- Local recurrence
- Not otherwise specified
- Not applicable or hematologic malignancy

Description

- Sample type associated with specimen on which NGS was performed
- Corresponds to variable [sample_type] in AACR Tier 1 data

Sequencing Assay ID

[cpt_seq_assay_id]

Value (Character)

Description

- The institutional assay identifier for the NGS genomic testing platform.
- Components are separated by hyphens, with the first component corresponding to the institution's abbreviation.
- All specimens tested by the same platform should have the same identifier.
- Corresponds to variable [seq_assay_id] in AACR Tier 1 data

Release Version Number

[release_version]

Value (Character)

- 1.0-public

Description

- Indicates the version number of the data release

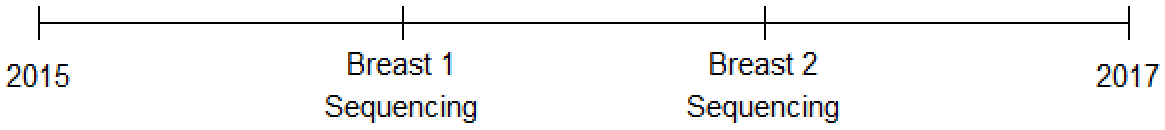
Appendix 1. BPC Project and Non-BPC Project Cancers

Definition of BPC Project Cancer: A BPC Project Cancer is the cancer that met the eligibility criteria for the project (i.e. genomic sequencing reported). The terms "BPC Project Cancer" and "index cancer" are used interchangeably. In Scenario 1, the patient had a single eligible cancer with associated genomic sequencing.

Scenario 1: Single BPC Project Cancer, No Second Primary Breast

Some patients may have more than one BPC Project Cancer because they have multiple sequenced cancers that met the eligibility criteria. For example, in Scenario 2, a patient with a diagnosis of Breast that was sequenced in 2015 and a second primary of BrCa that was sequenced in 2017 will have each diagnosis classified as a BPC Project Cancer in the BrCa BPC Project.

Scenario 2: Multiple BPC Project Cancers

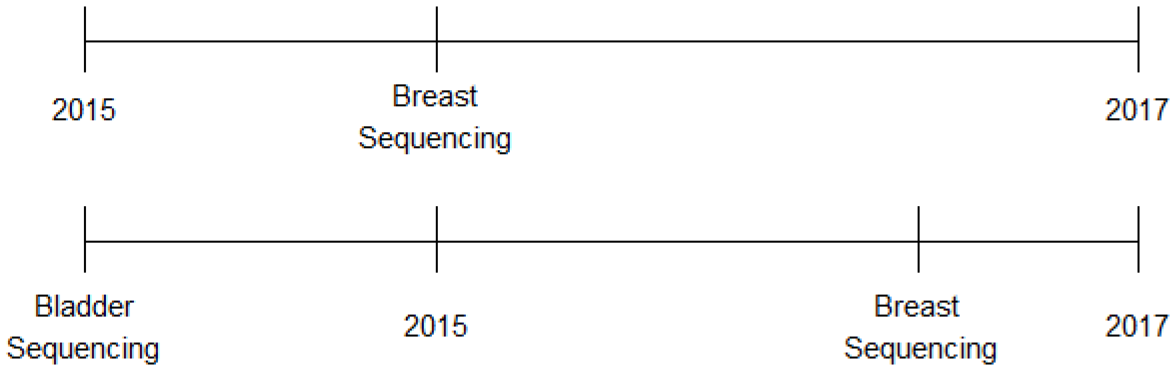


Definition of Non-BPC Project Cancer: A non-BPC Project Cancer diagnosis can be a diagnosis of the same or different cancer type as the BPC Project Cancer that occurs prior to, simultaneous with, or after the BPC Project Cancer. Non-BPC Project Cancers curated included: 1) non-BPC Project invasive cancer and in situ/non-invasive cancer diagnoses, and 2) other benign tumors. In a very small number of instances, some benign tumors (behavior codes 0 and 1), such as a benign brain tumor or a hemangioma, are obtained from the institution's tumor registry. These tumors were ineligible for curation, in which case only the tumor registry data is available.

The non-BPC Project Cancers do not have associated genomic sequencing.

A non-BPC Project cancer diagnosis could be a different cancer type altogether, such as a bladder cancer diagnosis in a patient included in the Breast BPC Project. The Breast cancer is the cancer that made the patient eligible for the BPC Project, but some information regarding the bladder cancer diagnosis was curated. In Scenario 3, the bladder cancer that was sequenced is classified as a non-BPC Project cancer for the BrCa BPC Project.

Scenario 3: Single BPC Project Cancer, Single Non-BPC Project Cancer



Appendix 2. Cancer-Directed Drugs

- Abemaciclib(LY2835219,Verzenio)
- Alpelisib(BYL719, Piqray)
- Anastrozole(Arimidex,ICID1033,ZD1033)
- Apatinib(YN968D1)
- Atezolizumab(MPDL3280A,RG7446,RO5541267,Tecentriq)
- BCG Solution(Bacillus Calmette Guerin Solution, TICE BCG Solution)
- Bendamustine (Bendeka, L0501,Treanda)
- Bevacizumab(AntiVEGF rhuMAb, BEVZ92,Bevacizumab-awwb,Bevacizumab BI 695502,Bevacizumab CBT 124,Bevacizumab FKB238,Bevacizumab HD204,Bevacizumab HLX04,Bevacizumab MIL60,Bevacizumab QL 1101,HD204,Avastin,Alymsys,Mvasi,Zirabev)
- Bicalutamide(Casodex,Cosudex,ICI176334)
- Capecitabine(Ro091978 000,Xeloda)
- Carboplatin(Blastocarb,CBDCA, Paraplat,Paraplatin,Paraplatine,Platinwas,Ribocarbo)
- Cetuximab(Erbitux,IMCC225)
- Cisplatin(Briplatin,Cisplatina,Cisplatinum,Cisplatyl,Citoplatino,Citosin,Cysplatyna,CDDP, DDP,Lederplatin,Metaplatin,Neoplatin, Peyrone's Salt,Placis,Plastistil,Platamine,Platiblastin,PlatiblastinS,Platinex,Platinol,Platinoxan,Platinum Diamminodichloride,Platiran,Platistin,Platosin)
- Cyclophosphamide(CTX,CYCLOcell, ,Cytoxan,Endoxan,Fosfaseron,Genoxal,Genuxal,Ledoxina,Mitoxan,Neosar, WR138719)
- Dacarbazine(Asercit,Biocarbazine,DIC,DTIC,DTICDome,Dacarbazina,Dacarbazine DTIC,Dacatic,Dakarbazin,Deticene,Detimedac, Fauldetic,Imidazole Carboxamide,WR139007)
- Dasatinib(BMS354825,Sprycel)
- Docetaxel (Taxotere,RP56976)
- Doxorubicin HCL(ADM,Adriamycin, DOX,DOXOCELL,Doxorubin,FI 106,FI106,Rubex)
- Epirubicin HCL(Ellence,IMI28,Pharmorubicin PFS)
- Eribulin Mesylate(B1939 Mesylate,E7389,ER086526,Halaven)
- Erlotinib HCL(CP358,774,OSI774,Tarceva)
- Etoposide(EPEG,Lastet,Toposar,VP 16213,VP16,VP16213,Vepesid)
- Everolimus(Afinitor,Afinitor Disperz,Certican,RAD001,Votubia,Zortress)
- Exemestane(Aromasin,FCE24304)
- Floxuridine(FDUR,FUDR,WR138720)
- Fluorouracil(5FU, AccuSite,Adrucil,Carac, Fluracedyl,Fluracil, Ribofluor,Ro29757)
- Fulvestrant(Faslodex,ICI182780,ZD9238)
- Gemcitabine HCL(FF10832,Gemzar,LY188011, dFdCyd)
- Goserlin Acetate(Zoladex,ZDX)
- Ifosfamide(Cyfos,Holoxan,Holoxane,IFO,IFOCell,IFX,Ifex,Ifolem,Ifomida,Ifomide,Ifosfamidum,Ifoxan,MJF9325,Mitoxana,Naxamide,Seromida,Tronoxal,Z4942)
- Investigational Drug
- Ipilimumab(BMS734016,MDX010,MDXCTLA4, Yervoy)
- Irinotecan HCL(CPT11,Campto,Camptosar,U101440E)
- Irinotecan liposome(Onivyde, MM398,PEP02)
- Ixabepilone(mRNA2416,Ixempra,BMS247550)
- Lapatinib Ditosylate(Tykerb)
- Larotrectinib(Vitrakvi,LOXO 101, ARRY470)
- Lenvatinib Mesylate(E7080,Lenvima)
- Letrozole(CGS 20267,Femara)
- Leucovorin(Calfolex,Calinat,Cehafolin,Citofolin,Citrec,Citrovorum Factor,Cromatonbic Folinico,Dalisol,Disintox,Divical,Ecofol,Emovis, FOLIcell, Wellcovorin)
- Leuprolide(A43818, Carcinil,DepoEligard,Enanton,Enantone,EnantoneGyn,Ginecrin,LEUP,Leuplin,Lucrin Depot,Lupron,Trenantone,UnoEnantone,Viadur)
- Lurbinectedin(PM01183,Zepzelca)
- Megestrol Acetate(Megace,Megestat,Megestil,Niagestin,Ovaban,Pallace,SC10363)
- Methotrexate(Abitrexate,AlphaMethopterin, Amethopterin,Brimexate,CL14377,Emtexate,Farmitrexat,Fauldexato,Folex,Lantarel,Ledertrexate,Lumexon,MTX, WR19039)

- Nabpaclitaxel(ABI 007,ABI007,Abraxane,Albumin bound PaclitaxeL)
- Neratinib(HKI272,PB272,Nerlynx)
- Nivolumab(BMS936558,MDX1106,NIVO,ONO 4538,Opdivo)
- Olaparib(AZD2281,KU0059436,Lynparza)
- Olaratumab(IMC3G3,Lartruvo)
- Other NOS
- Other antineoplastic
- Other hormone
- Oxaliplatin(1OHP,Ai Heng,DACPLAT,Dacotin,ELOXATIN,Eloxatine,JM83)
- Paclitaxel Loaded Polymeric Micelle(Cynviloq TM,Genexol PM,IG001)
- Paclitaxel(Anzatax,Asotax,Bristaxol,Praxel,Taxol)
- Palbociclib(Ibrance,PD0332991)
- Pazopanib HCL(GW786034B,Votrient)
- Pegylated Liposomal Doxorubicin(ATI0918,Caelyx,DOXSL,DOXIL ,Evacet,LipoDox,LipoDox,Lipodox 50,Liposomal Adriamycin, D99)
- Pembrolizumab(Keytruda,Lambrolizumab,MK 3475,SCH900475)
- Pemetrexed Disodium(Alimta,LY231514)
- Pertuzumab(2C4,Omnitarg, Perjeta,RO4368451, rhuMAb2C4)
- Pertuzumab-Trastuzumab-Hyaluronidase-ZZXF(Phesgo)
- Ponatinib HCL(Iclusig)
- Ramucirumab(Cyramza,IMC1121B,LY3009806)
- Ribociclib(LEE011, Kisqali)
- Rituximab(ABP798,BI 695500,C2B8 Monoclonal Antibody,CTP10,IDEC102,IDECC2B8,MabThera,PF05280586,RTXM83,Rituxan,Rituximab ABP 798,Rituximab BI 695500,Rituximab CTP10,Rituximab GB241,Rituximab IBI301,Rituximab PF05280586,Rituximab RTXM83,Rituximab SAIT101,Truxima,rituximab abbs)
- Sacituzumab Govitecan(IMMU-132,RS7-SN38,Trodelvy)
- Sorafenib Tosylate(BAY439006 Tosylate,BAY549085,Nexavar)
- Talazoparib(BMN673, Talzenna)
- Tamoxifen (ApoTamox,Clonoxifen,Dignotamoxi,Ebefen,Emblon,Estroxyn,Fentamox,GenTamoxifen,Genox,Jenoxifen,Kessar,Ledertam,Lesporene,Nolgen,Noltam,Nolvadex,NolvadexD,Nourytam, NovoTamoxifen,Novofen,Noxitem,Oestrifen, Oncotam,,Soltamox,TAM)
- Taselisib(GDC0032,RO5537381)
- Tegafurgimeraciloteracil Potassium(BMS247616,S1,STS1,Teysuno)
- Temozolomide(Methazolastone,RP46161,SCH 52365,Temcad,Temodal,Temodar,Temomedac)
- Topotecan HCL(Hycamptamine, Hycamtin,SKF S104864A)
- Toremifene(Farestone)
- Trametinib(GSK1120212, JTP74057, Mekinist)
- Trastuzumab Deruxtecan(DS8201a,Enhertu,WHO10516)
- Trastuzumab Emtansine(Kadcyla, PRO132365, RO5304020, TDM1 ,Trastuzumab DM1,Trastuzumab MCCDM1)
- Trastuzumab(ABP980,ALT02,Antip185HER2, ,Herceptin,PF05280014,Herzuma,Kanjinti,Ogivri, Ontruzant,Trazimera,PF05280014, RO0452317,Trastuzumab ABP 980,Trastuzumab ALT02,Trastuzumab HLX02, Trastuzumab EG12014, Trastuzumab PF05280014, Trastuzumab-anns, Trastuzumab-dkst, Trastuzumab-dttb, Trastuzumab-qyyp, cerb2 Monoclonal Antibody, rhuMAb HER2,Trastuzumab SIBP-01, Trastuzumab SB3, Trastuzumab-pkrb)
- Trastuzumab/Hyaluronidase-oysk(Herceptin Hylecta)
- Tucatinib(ARRY380, Irbinitinib, ONT380,Tukysa)
- Vinorelbine Tartrate(Biovelbin,Eunades,KW2307,NVB,Navelbine)
- Yttrium Y90 Ibritumomab Tiuxetan(IDECY2B8,Y 90 Zevalin)

Appendix 3. Curation Instructions for Medical Oncologist Assessments

- Find the date of diagnosis of the cancer of interest.
- Some patients have more than one cancer diagnosis. Do review and curate only notes for the cancer of interest.
- Do review and curate one clinical assessment per month, beginning at time of diagnosis (+/- 30 days).
    - Do choose first Medical Oncology note of the month that is authored by an MD actively following the patient for the cancer of interest.
    - If there is no note by a medical oncologist (MD) that month. Do use the first note by a nurse practitioner or physician assistant (NP/PA) from a medical oncology practice.
        - Rarely look at more than one note per month. If the patient has imaging scans after the first note in the month, look on or up to 7 days following the imaging scans.
    - Do give priority to internal visit notes. Only curate notes from an outside institution if there is no DFCI/Partners oncology note in that month. Prioritize as follows:
        1. Internal MD
        2. Internal NP/PA
        3. External MD
        4. External NP/PA
    - If a fellow has a note and an attending physician adds an addendum, review any information in the Summary/impression/Evaluation/Plan, including the addendum.
    - Do not use notes from Radiation Oncology. Surgery/Surgical Oncology, inpatient care, primary care, or other specialists not related to cancer (e.g. dermatology, cardiology).
        - Patients with early stage cancers are sometimes only followed by a surgical oncologist. This may mean that they are many months without notes that qualify for curation. That is ok!
    - Rarely curate notes by an oncologist from a different specialty than the cancer of interest. Occasionally a patient will transfer care to a different type of oncologist due to the particulars of their disease. For example. a patient with brain metastases from their primary cancer may be followed by Neuro Oncology. Do curate these notes if notes from the primary oncologist are not available. Do not curate notes from a different oncology specialist that are pertaining to a different cancer diagnosis.
- Do review only the Impression/Plan section at the bottom of the note as well as the reason for visit.
    - Rarely a medical oncology provider will summarize the cancer status directly above the Impression/Plan section, which can be reviewed for curation.
- Do not review any of the other sections. including the physical exam, interval history, lab results. etc.
    - The Impression/Plan section may have a different name in medical oncologist progress notes; other section headers could include Assessment, Summary, Conclusion, Problem List Items Addressed this Visit.
    - When there are no section headings in a provider’s note. Do review everything located beneath the physical exam.

Appendix 4. Breast Distant Metastasis Classification

ICD-O-3 Topography Code	Metastasis Classification
C00.0 External upper lip	Distant (Head and neck)
C00.1 External lower lip	Distant (Head and neck)
C00.2 External lip NOS	Distant (Head and neck)
C00.3 Mucosa of upper lip	Distant (Head and neck)
C00.4 Mucosa of lower lip	Distant (Head and neck)
C00.5 Mucosa of lip NOS	Distant (Head and neck)
C00.6 Commissure of lip	Distant (Head and neck)
C00.8 Overlapping lesion of lip	Distant (Head and neck)
C00.9 Lip NOS	Distant (Head and neck)
C01.9 Base of tongue NOS	Distant (Head and neck)
C02.0 Dorsal surface of tongue NOS	Distant (Head and neck)
C02.1 Border of tongue	Distant (Head and neck)
C02.2 Ventral surface of tongue NOS	Distant (Head and neck)
C02.3 Anterior 2/3 of tongue NOS	Distant (Head and neck)
C02.4 Lingual tonsil	Distant (Head and neck)
C02.8 Overlapping lesion of tongue	Distant (Head and neck)
C02.9 Tongue NOS	Distant (Head and neck)
C03.0 Upper Gum	Distant (Head and neck)
C03.1 Lower gum	Distant (Head and neck)
C03.9 Gum NOS	Distant (Head and neck)
C04.0 Anterior floor of mouth	Distant (Head and neck)
C04.1 Lateral floor of mouth	Distant (Head and neck)
C04.8 Overlapping lesion of floor of mouth	Distant (Head and neck)
C04.9 Floor of mouth NOS	Distant (Head and neck)
C05.0 Hard palate	Distant (Head and neck)
C05.1 Soft palate NOS	Distant (Head and neck)
C05.2 Uvula	Distant (Head and neck)
C05.8 Overlapping lesion of palate	Distant (Head and neck)
C05.9 Palate NOS	Distant (Head and neck)
C06.0 Cheek mucosa	Distant (Head and neck)
C06.1 Vestibule of mouth	Distant (Head and neck)
C06.2 Retromolar area	Distant (Head and neck)
C06.8 Overlapping lesion of other and unspecified parts of mouth	Distant (Head and neck)
C06.9 Mouth NOS	Distant (Head and neck)
C07.9 Parotid gland	Distant (Head and neck)
C08.0 Submandibular gland	Distant (Head and neck)
C08.1 Sublingual gland	Distant (Head and neck)
C08.8 Overlapping lesion of major salivary glands	Distant (Head and neck)
C08.9 Major salivary gland NOS	Distant (Head and neck)
C09.0 Tonsillar fossa	Distant (Head and neck)
C09.1 Tonsillar pillar	Distant (Head and neck)
C09.8 Overlapping lesion of tonsil	Distant (Head and neck)
C09.9 Tonsil NOS	Distant (Head and neck)
C10.0 Vallecula	Distant (Head and neck)
C10.1 Anterior surface of epiglottis	Distant (Head and neck)
C10.2 Lateral wall of oropharynx	Distant (Head and neck)
C10.3 Posterior wall of oropharynx	Distant (Head and neck)
C10.4 Branchial cleft	Distant (Head and neck)
C10.8 Overlapping lesions of oropharynx	Distant (Head and neck)
C10.9 Oropharynx NOS	Distant (Head and neck)
C11.0 Superior wall of nasopharynx	Distant (Head and neck)
C11.1 Posterior wall of nasopharynx	Distant (Head and neck)
C11.2 Lateral wall of nasopharynx	Distant (Head and neck)
C11.3 Anterior wall of nasopharynx	Distant (Head and neck)
C11.8 Overlapping lesion of nasopharynx	Distant (Head and neck)
C11.9 Nasopharynx NOS	Distant (Head and neck)
C12.9 Pyriform sinus	Distant (Head and neck)
C13.0 Postcricoid region	Distant (Head and neck)
C13.1 Hypopharyngeal aspect of aryepiglottic fold	Distant (Head and neck)
C13.2 Posterior wall of hypopharynx	Distant (Head and neck)
C13.8 Overlapping lesion of hypopharynx	Distant (Head and neck)
C13.9 Hypopharynx NOS	Distant (Head and neck)
C14.0 Pharynx NOS	Distant (Head and neck)
C14.2 Waldeyer ring	Distant (Head and neck)
C14.8 Overlapping lesion of lip oral cavity and pharynx	Distant (Head and neck)
C15.0 Cervical esophagus	Distant (Head and neck)
C15.1 Thoracic esophagus	Distant (Thorax)
C15.2 Abdominal esophagus	Distant (Abdomen)
C15.3 Upper third of esophagus	Distant (Head and neck)
C15.4 Middle third of esophagus	Distant (Thorax)
C15.5 Lower third of esophagus	Distant (Abdomen)
C15.8 Overlapping lesion of esophagus	Distant (Other)
C15.9 Esophagus NOS	Distant (Other)
C16.0 Cardia NOS	Distant (Abdomen)
C16.1 Fundus of stomach	Distant (Abdomen)
C16.2 Body of stomach	Distant (Abdomen)
C16.3 Gastric antrum	Distant (Abdomen)
C16.4 Pylorus	Distant (Abdomen)
C16.5 Lesser curvature of stomach NOS	Distant (Abdomen)
C16.6 Greater curvature of stomach NOS	Distant (Abdomen)
C16.8 Overlapping lesion of stomach	Distant (Abdomen)
C16.9 Stomach NOS	Distant (Abdomen)
C17.0 Duodenum	Distant (Abdomen)
C17.1 Jejunum	Distant (Abdomen)
C17.2 Ileum	Distant (Abdomen)
C17.3 Meckel diverticulum	Distant (Abdomen)
C17.8 Overlapping lesion of small intestine	Distant (Abdomen)
C17.9 Small intestine NOS	Distant (Abdomen)
C18.0 Cecum	Distant (Abdomen)
C18.1 Appendix	Distant (Abdomen)
C18.2 Ascending colon	Distant (Abdomen)
C18.3 Hepatic flexure of colon	Distant (Abdomen)
C18.4 Transverse colon	Distant (Abdomen)
C18.5 Splenic flexure of colon	Distant (Abdomen)
C18.6 Descending colon	Distant (Abdomen)
C18.7 Sigmoid colon	Distant (Pelvis)
C18.8 Overlapping lesion of colon	Distant (Abdomen)
C18.9 Colon NOS	Distant (Abdomen)
C19.9 Rectosigmoid junction	Distant (Pelvis)
C20.9 Rectum NOS	Distant (Pelvis)

ICD-O-3 Topography Code	Metastasis Classification
C21.0 Anus NOS	Distant (Pelvis)
C21.1 Anal canal	Distant (Pelvis)
C21.2 Cloacogenic zone	Distant (Pelvis)
C21.8 Overlapping lesion of rectum anus and anal canal	Distant (Pelvis)
C22.0 Liver	Distant (Liver)
C22.1 intrahepatic bile duct	Distant (Abdomen)
C23.9 Gallbladder	Distant (Abdomen)
C24.0 Extrahepatic bile duct	Distant (Abdomen)
C24.1 Ampulla of Vater	Distant (Abdomen)
C24.8 Overlapping lesion of biliary tract	Distant (Abdomen)
C24.9 Biliary tract NOS	Distant (Abdomen)
C25.0 Head of pancreas	Distant (Abdomen)
C25.1 Body of pancreas	Distant (Abdomen)
C25.2 Tail of pancreas	Distant (Abdomen)
C25.3 Pancreatic duct	Distant (Abdomen)
C25.4 Islets of Langerhans	Distant (Abdomen)
C25.7 Other specified parts of pancreas	Distant (Abdomen)
C25.8 Overlapping lesion of pancreas	Distant (Abdomen)
C25.9 Pancreas NOS	Distant (Abdomen)
C26.0 Intestinal tract NOS	Distant (Abdomen)
C26.8 Overlapping lesion of digestive system	Distant (Abdomen)
C26.9 Gastrointestinal tract NOS	Distant (Abdomen)
C30.0 Nasal cavity	Distant (Head and neck)
C30.1 Middle ear	Distant (Head and neck)
C31.0 Maxillary sinus	Distant (Head and neck)
C31.1 Ethmoid sinus	Distant (Head and neck)
C31.2 Frontal sinus	Distant (Head and neck)
C31.3 Sphenoid sinus	Distant (Head and neck)
C31.8 Overlapping lesion of accessory sinuses	Distant (Head and neck)
C31.9 Accessory sinus NOS	Distant (Head and neck)
C32.0 Glottis	Distant (Head and neck)
C32.1 Supraglottis	Distant (Head and neck)
C32.2 Subglottis	Distant (Head and neck)
C32.3 Laryngeal cartilage	Distant (Head and neck)
C32.8 Overlapping lesion of larynx	Distant (Head and neck)
C32.9 Larynx NOS	Distant (Head and neck)
C33.9 Trachea	Distant (Thorax)
C34.0 Main bronchus	Distant (Pulmonary)
C34.1 Upper lobe lung	Distant (Pulmonary)
C34.2 Middle lobe lung	Distant (Pulmonary)
C34.3 Lower lobe lung	Distant (Pulmonary)
C34.8 Overlapping lesion of lung	Distant (Pulmonary)
C34.9 Lung NOS	Distant (Pulmonary)
C37.9 Thymus	Distant (Thorax)
C38.0 Heart	Distant (Pericardial and malignant pericardial effusion)
C38.1 Anterior mediastinum	Distant (Thorax)
C38.2 Posterior mediastinum	Distant (Thorax)
C38.3 Mediastinum NOS	Distant (Thorax)
C38.4 Pleura NOS	Distant (Pleura and malignant pleural effusion)
C38.8 Overlapping lesion of heart mediastinum and pleura	Distant (Pleura and malignant pleural effusion)
C39.0 Upper respiratory tract NOS	Distant (Thorax)
C39.8 Overlapping lesion of respiratory system and intrathoracic organs	Distant (Thorax)
C39.9 Ill-defined sites within respiratory system	Distant (Thorax)
C40.0 Long bones of upper limb scapula and associated joints	Distant (Bone)
C40.1 Short bones of upper limb and associated joints	Distant (Bone)
C40.2 Long bones of lower limb and associated joints	Distant (Bone)
C40.3 Short bones of lower limb and associated joints	Distant (Bone)
C40.8 Overlapping lesion of bones joints and articular cartilage of limbs	Distant (Bone)
C40.9 Bone of limb NOS	Distant (Bone)
C41.0 Bones of skull and face and associated joints	Distant (Bone)
C41.1 Mandible	Distant (Bone)
C41.2 Vertebral column	Distant (Bone)
C41.3 Rib sternum clavicle and associated joints	Distant (Bone)
C41.4 Pelvic bones sacrum coccyx and associated joints	Distant (Bone)
C41.8 Overlapping lesion of bones joints and articular cartilage	Distant (Bone)
C41.9 Bone NOS	Distant (Bone)
C42.0 Blood	Distant (Other)
C42.1 Bone marrow	Distant (Bone marrow)
C42.2 Spleen	Distant (Abdomen)
C42.3 Reticuloendothelial system NOS	Distant (Other)
C42.4 Hematopoietic system NOS	Distant (Other)
C44.0 Skin of lip NOS	Distant (Skin)
C44.1 Eyelid	Distant (Skin)
C44.2 External ear	Distant (Skin)
C44.3 Skin of other and unspecified parts of face	Distant (Skin)
C44.4 Skin of scalp and neck	Distant (Skin)
C44.5 Skin of trunk	
C44.6 Skin of upper limb and shoulder	Distant (Skin)
C44.7 Skin of lower limb and hip	Distant (Skin)
C44.8 Overlapping lesion of skin	Distant (Skin)
C44.9 Skin NOS	Distant (Skin)
C47.0 Peripheral nerves and autonomic nervous system of head face and neck	Distant (Head and neck)
C47.1 Peripheral nerves and autonomic nervous system of upper limb and shoulder	Distant (Other)

ICD-O-3 Topography Code	Metastasis Classification
C47.2 Peripheral nerves and autonomic nervous system of lower limb and hip	Distant (Other)
C47.3 Peripheral nerves and autonomic nervous system of thorax	Local/Regional (Thorax)
C47.4 Peripheral nerves and autonomic nervous system of abdomen	Distant (Abdomen)
C47.5 Peripheral nerves and autonomic nervous system of pelvis	Distant (Pelvis)
C47.6 Peripheral nerves and autonomic nervous system of trunk NOS	Distant (Abdomen)
C47.8 Overlapping lesion of peripheral nerves and autonomic nervous system	Distant (Other)
C47.9 Autonomic nervous system NOS	Distant (Other)
C48.0 Retroperitoneum	Distant (Abdomen)
C48.1 Specified parts of peritoneum	Distant (Peritoneum and malignant peritoneal effusion)
C48.2 Peritoneum NOS	Distant (Peritoneum and malignant peritoneal effusion)
C48.8 Overlapping lesion of retroperitoneum and peritoneum	Distant (Peritoneum and malignant peritoneal effusion)
C49.0 Connective Subcutaneous and other soft tissues of head face and neck	Distant (Head and neck)
C49.1 Connective Subcutaneous and other soft tissues of upper limb and shoulder	Distant (Other)
C49.2 Connective Subcutaneous and other soft tissues of lower limb and hip	Distant (Other)
C49.3 Connective Subcutaneous and other soft tissues of thorax	Local/Regional (Thorax)
C49.4 Connective Subcutaneous and other soft tissues of abdomen	Distant (Abdomen)
C49.5 Connective Subcutaneous and other soft tissues of pelvis	Distant (Pelvis)
C49.6 Connective Subcutaneous and other soft tissues of trunk NOS	Distant (Abdomen)
C49.8 Overlapping lesion of connective subcutaneous and other soft tissues	Distant (Other)
C49.9 Connective Subcutaneous and other soft tissues NOS	Distant (Other)
C50.0 Nipple	Local/Regional (Breast)
C50.1 Central portion of breast	Local/Regional (Breast)
C50.2 Upper-inner quadrant of breast	Local/Regional (Breast)
C50.3 Lower-inner quadrant of breast	Local/Regional (Breast)
C50.4 Upper-outer quadrant of breast	Local/Regional (Breast)
C50.5 Lower-outer quadrant of breast	Local/Regional (Breast)

ICD-O-3 Topography Code	Metastasis Classification
C50.6 Axillary tail of breast	Local/Regional (Breast)
C50.8 Overlapping lesion of breast	Local/Regional (Breast)
C50.9 Breast NOS	Local/Regional (Breast)
C51.0 Labium majus	Distant (Pelvis)
C51.1 Labium minus	Distant (Pelvis)
C51.2 Clitoris	Distant (Pelvis)
C51.8 Overlapping lesion of vulva	Distant (Pelvis)
C51.9 Vulva NOS	Distant (Pelvis)
C52.9 Vagina NOS	Distant (Pelvis)
C53.0 Endocervix	Distant (Pelvis)
C53.1 Exocervix	Distant (Pelvis)
C53.8 Overlapping lesion of cervix uteri	Distant (Pelvis)
C53.9 Cervix uteri	Distant (Pelvis)
C54.0 Isthmus uteri	Distant (Pelvis)
C54.1 Endometrium	Distant (Pelvis)
C54.2 Myometrium	Distant (Pelvis)
C54.3 Fundus uteri	Distant (Pelvis)
C54.8 Overlapping lesion of corpus uteri	Distant (Pelvis)
C54.9 Corpus uteri	Distant (Pelvis)
C55.9 Uterus NOS	Distant (Pelvis)
C56.9 Ovary	Distant (Pelvis)
C57.0 Fallopian tube	Distant (Pelvis)
C57.1 Broad ligament	Distant (Pelvis)
C57.2 Round ligament	Distant (Pelvis)
C57.3 Parametrium	Distant (Pelvis)
C57.4 Uterine adnexa	Distant (Pelvis)
C57.7 Other specified parts of female genital organs	Distant (Pelvis)
C57.8 Overlapping lesion of female genital organs	Distant (Pelvis)
C57.9 Female genital tract NOS	Distant (Pelvis)
C58.9 Placenta	Distant (Pelvis)
C60.0 Prepuce	Distant (Pelvis)
C60.1 Glans penis	Distant (Pelvis)
C60.2 Body of penis	Distant (Pelvis)
C60.8 Overlapping lesion of penis	Distant (Pelvis)
C60.9 Penis NOS	Distant (Pelvis)
C61.9 Prostate gland	Distant (Pelvis)
C62.0 Undescended testis	Distant (Pelvis)
C62.1 Descended testis	Distant (Pelvis)
C62.9 Testis NOS	Distant (Pelvis)
C63.0 Epididymis	Distant (Pelvis)
C63.1 Spermatic cord	Distant (Pelvis)
C63.2 Scrotum NOS	Distant (Pelvis)
C63.7 Other specified parts of male genital organs	Distant (Pelvis)
C63.8 Overlapping lesion of male genital organs	Distant (Pelvis)
C63.9 Male genital organs NOS	Distant (Pelvis)
C64.9 Kidney NOS	Distant (Abdomen)
C65.9 Renal pelvis	Distant (Abdomen)
C66.9 Ureter	Distant (Pelvis)
C67.0 Trigone of bladder	Distant (Pelvis)
C67.1 Dome of bladder	Distant (Pelvis)
C67.2 Lateral wall of bladder	Distant (Pelvis)
C67.3 Anterior wall of bladder	Distant (Pelvis)

ICD-O-3 Topography Code	Metastasis Classification
C67.4 Posterior wall of bladder	Distant (Pelvis)
C67.5 Bladder neck	Distant (Pelvis)
C67.6 Ureteric orifice	Distant (Pelvis)
C67.7 Urachus	Distant (Pelvis)
C67.8 Overlapping lesion of bladder	Distant (Pelvis)
C67.9 Bladder NOS	Distant (Pelvis)
C68.0 Urethra	Distant (Pelvis)
C68.1 Paraurethral gland	Distant (Pelvis)
C68.8 Overlapping lesion of urinary organs	Distant (Pelvis)
C68.9 Urinary system NOS	Distant (Pelvis)
C69.0 Conjunctiva	Distant (Head and neck)
C69.1 Cornea NOS	Distant (Head and neck)
C69.2 Retina	Distant (Brain/CNS)
C69.3 Choroid	Distant (Head and neck)
C69.4 Ciliary body	Distant (Head and neck)
C69.5 Lacrimal gland	Distant (Head and neck)
C69.6 Orbit NOS	Distant (Head and neck)
C69.8 Overlapping lesion of eye and adnexa	Distant (Head and neck)
C69.9 Eye NOS	Distant (Head and neck)
C70.0 Cerebral meninges	Distant (Brain/CNS)
C70.1 Spinal meninges	Distant (Brain/CNS)
C70.9 Meninges NOS	Distant (Brain/CNS)
C71.0 Cerebrum	Distant (Brain/CNS)
C71.1 Frontal lobe	Distant (Brain/CNS)
C71.2 Temporal lobe	Distant (Brain/CNS)
C71.3 Parietal lobe	Distant (Brain/CNS)
C71.4 Occipital lobe	Distant (Brain/CNS)
C71.5 Ventricle NOS	Distant (Brain/CNS)
C71.6 Cerebellum NOS	Distant (Brain/CNS)
C71.7 Brain stem	Distant (Brain/CNS)
C71.8 Overlapping lesion of brain	Distant (Brain/CNS)
C71.9 Brain NOS	Distant (Brain/CNS)
C72.0 Spinal cord	Distant (Brain/CNS)
C72.1 Cauda equina	Distant (Brain/CNS)
C72.2 Olfactory nerve	Distant (Brain/CNS)
C72.3 Optic nerve	Distant (Brain/CNS)
C72.4 Acoustic nerve	Distant (Brain/CNS)
C72.5 Cranial nerve NOS	Distant (Brain/CNS)
C72.8 Overlapping lesion of brain and central nervous system	Distant (Brain/CNS)
C72.9 Nervous system NOS	Distant (Brain/CNS)
C73.9 Thyroid gland	Distant (Head and neck)
C74.0 Cortex of adrenal gland	Distant (Adrenal)
C74.1 Medulla of adrenal gland	Distant (Adrenal)
C74.9 Adrenal gland NOS	Distant (Adrenal)
C75.0 Parathyroid gland	Distant (Head and neck)
C75.1 Pituitary gland	Distant (Head and neck)
C75.2 Craniopharyngeal duct	Distant (Head and neck)
C75.3 Pineal gland	Distant (Head and neck)
C75.4 Carotid body	Distant (Head and neck)
C75.5 Aortic body and other paraganglia	Distant (Abdomen)
C75.8 Overlapping lesion of endocrine glands and related structures	Distant (Other)
C75.9 Endocrine gland NOS	Distant (Other)
C76.0 Head face or neck NOS	Distant (Head and neck)
C76.1 Thorax NOS	Distant (Thorax)
C76.2 Abdomen NOS	Distant (Abdomen)
C76.3 Pelvis NOS	Distant (Pelvis)
C76.4 Upper limb NOS	Distant (Other)
C76.5 Lower limb NOS	Distant (Other)
C76.7 Other ill-defined sites	Distant (Other)
C76.8 Overlapping lesion of ill-defined sites	Distant (Other)
C77.1 Intrathoracic lymph nodes	Distant (Lymph nodes)
C77.2 Intra-abdominal lymph nodes	Distant (Lymph nodes)
C77.3 Lymph nodes of axilla or arm	Local/Regional (Lymph nodes)
C77.4 Lymph nodes of inguinal region or leg	Distant (Lymph nodes)
C77.5 Pelvic lymph nodes	Distant (Lymph nodes)
C77.8 Lymph nodes of multiple regions	Distant (Lymph nodes)
F10 Cerebral Spinal Fluid	Distant (Brain/CNS)
F50 Pericardial Fluid	Distant (Pericardial and malignant pericardial effusion)
F20 Peritoneal Fluid/Ascites	Distant (Peritoneum and malignant peritoneal effusion)
F30 Pleural Fluid	Distant (Pleura and malignant pleural effusion)
F40 Urine	Distant (Pelvis)